Use of ecological niche modelling to predict distributions of freshwater fish species in Kansas

McNyset KM. Use of ecological niche modelling to predict distributions of freshwater fish species in Kansas. Ecology of Freshwater Fish 2005: 14: 243–255. © Blackwell Munksgaard, 2005

Abstract - An essential innovation in aquatic biodiversity research would be a robust approach to accurately predict species' potential distributions. In this paper, I conduct an analysis to test the efficacy of ecological niche modelling for predicting fish species' potential distributions using an artificial-intelligence algorithm, the Genetic Algorithm for Rule-Set Prediction (GARP). Models of species' ecological niches are developed using GARP, and projected onto geography to predict species distributions. To test the validity of this approach, I used freshwater fish distribution data for twelve fish species occurring in Kansas. These taxa were chosen to represent phylogenetic, distribution, and habit diversity. I subset these data by omitting half of the counties from model building, and test models using the omitted counties. Models were tested using Receiver Operating Characteristic (ROC) analyses. Of the species tested, all were statistically significant with the models showing excellent predictive ability. Omission errors across taxa ranged from 0 to 17%. This inferential capacity opens doors to many synthetic analyses based on primary pointoccurrence data.

K. M. McNyset

Natural History Museum and Biodiversity Research Center, The University of Kansas, Lawrence, KS, USA

Key words: ecological niche modelling; GARP; ROC

K. M. McNyset, Natural History Museum and Biodiversity Research Center, The University of Kansas, 1735 Jayhawk Blvd., Lawrence, KS 66045-7561, USA; e-mail: mcnyset@ku.edu

Accepted for publication April 20, 2005

Un resumen en español se incluye detrás del texto principal de este artículo.

Introduction

Fisheries biologists, ichthyologists, and policy-makers have an ever-increasing need to understand distributions of aquatic species on local, regional and global scales and to better understand the parameters that influence those distributions. Response of species to human-altered landscapes, climate change and conservation management strategies are just a few of the immediate concerns (Tonn 1990; Fausch et al. 2002). An essential innovation is a robust approach to accurately predict potential distributions under a variety of scenarios of present and future conditions. One primary need, access to information regarding the actual localities in specimen records, is being addressed through distributed database technology systems that tie data providers into a cooperative information network such as The Species Analyst and FishNet (Vieglais et al. 2000). These networks provide access to millions of specimen records stored remotely at partner institutions and return data in a timely manner (seconds to a few minutes). Records can be used in a geographic information system (GIS) environment to generate species distribution maps. However, these maps are based entirely on the often limited and incomplete data available as point occurrences (Krohn 1996).

Several approaches have been used to predict potential distributions based on models of a species' ecological niche. For example, BIOCLIM (Nix 1986) tallies frequencies of occurrences of a taxon in environmental categories, trims the extreme 5% of the distribution, and then finds the conjunction of appropriate areas across all environmental categories. BIOCLIM, however, only predicts where a species does not occur, not where it does occur, resulting in numerous false predictions of occurrence (Stockwell & Peters 1999). Logistic regression has been shown to be a more effective approach to modelling ecological niches (Austin et al. 1990), although generating useful results with categorical data (e.g. land cover type) is problematic (Peterson et al. 2002), and it is sensitive

to numbers of occurrences relative to total area considered (Cumming 2000). In addition, areas of absence must be known although that data is most often unavailable for a given taxon (Peterson & Cohoon 1999; Stockwell & Peterson 2002). Recently, a variety of artificial intelligence approaches have been applied to the problem of predicting aquatic species distributions, i.e. artificial neural networks (Olden 2000; Olden & Jackson 2001, 2002; Vander Zanden et al. 2004) and genetic algorithms. For example, The Genetic Algorithm for Rule-Set Prediction (GARP) has been used successfully to predict species distributions in marine systems (Wiley et al. 2003). GARP uses a super-set of rule-building methods (logistic regression, range rules, negated range rules and atomic rules) in a machine-learning (artificial intelligence) environment to build heterogeneous rule-sets that describe ecological niches (Stockwell & Noble 1992; Stockwell & Peters 1999). The heterogeneous nature of the rule-set is critical because it allows niche parameters to vary over the known range of the species (Dunham et al. 2002). Input into GARP consists of point-occurrence data (latitude-longitude pairs) and environmental data in the form of raster grids. GARP uses an iterative learning process to develop a rule-set defining a species' niche relative to the environmental data sets. This niche model is then projected back onto the landscape, generating a prediction of potential distribution for that taxon. The predictions are robust and accurate generally across many terrestrial taxa (Peterson 2001; Peterson et al. 2002; Stockwell & Peterson 2002). GARP models should be understood as partial niche models, as it is unlikely that all the relevant niche parameters are included in any given analysis (Wiley et al. 2003).

There is a long history of analyses of habitat associations and classifications in freshwater fishes, especially on local or regional scales (Grossman et al. 1999). Research related to correlations between ecological factors and community structure or abundance are particularly prevalent (e.g. Hughes et al. 1987; Lyons 1989; Douglas & Matthews 1992; Nelson et al. 1992; Taylor et al. 1992; Lamouroux et al. 1999; Jackson et al. 2001). Analyses attempting to predict individual fish species distributions based on ecological niche models are few (Olden & Jackson 2001; Wiley et al. 2003; Iguchi et al. 2004; Vander Zanden et al. 2004). Most researchers in the field of predictive modelling work on terrestrial taxa and the nascent U.S. Aquatic GAP programs (http://www.gap.uidaho.edu/ Projects/Aquatic/default.htm) have yet to produce published results of this nature. D'Angelo et al. (1995) successfully used a genetic algorithm to predict trout distributions within stream reaches, and Wiley et al. (2003) demonstrated the capability of ecological niche modelling using GARP to predict marine fish distributions. Kluza & McNyset (2005) used GARP to predict potential invasions of a freshwater mussel and an estuarine crab. However, there have been no tests of GARP for predicting landscape-scale distributions of freshwater fishes.

Several concepts of ecological niches have been developed. For the purposes of this analysis, an ecological niche is conceptualised as the combination of ecological parameters within which a species can maintain populations without immigration (Grinnell 1917). This niche can be visualised as an *N*-dimensional hypervolume, the dimensions of which are ecological parameters (Hutchinson 1957). What qualifies as a relevant niche parameter is dependent on the spatial scale being discussed, as different parameters are relevant only at the appropriate spatial scale (e.g. local vs. regional vs. worldwide; see Chapters 9–20 in Scott et al. 2002).

Herein, I conduct an analysis to test the efficacy of the ecological niche modelling using GARP for predicting potential distributions of 12 freshwater fish species in Kansas.

Materials and methods

Twelve fish species occurring in Kansas were selected for analysis (Table 1) from across the phylogenetic spectrum and that represent a range of distributions and a variety of habits. Collection information for each taxon was acquired from the University of Kansas Natural History Museum Fish Collection database using FishNet and The Species Analyst (http:// speciesanalyst.net/fishnet). Textual locality information was assigned latitude and longitude in decimal degrees using the Geographic Names Information System of the U.S. Geological Survey (USGS) (http://geonames.usgs.gov); township-section-range data were converted to decimal degrees using a conversion engine developed by the Montana State University Environmental Statistics Group (http:// www.esg.montana.edu/gl/trs-data.html). Records with ambiguous or incomplete locality information were excluded from analysis. As many unique collection localities as were available per taxon were included. The number of unique records per taxon ranged from 38 for Etheostoma flabellare to 141 for Gambusia affinis; varying approximately proportionate to the size of the actual geographic distribution. Collection dates ranged from 1946 to 2001.

A subsetting protocol was chosen to mimic a common pattern of data availability for fisheries researchers. That is, collection data is often available only in a patchy mosaic across a given region; some areas have been well collected while no collection information exists for others. To that end, pointoccurrence data for each species were subset by

Table	1.	List	of	Kansas	fish	taxa	included	in	this	analysis.
-------	----	------	----	--------	------	------	----------	----	------	-----------

Taxon	<i>N</i> to build	N to test	AUC	Omission (%)	Total % commission	Average % commission
Lepisosteus osseus	32	30	0.8502*	6	39.9	31.1
Cyprinella camura	18	29	0.8474*	17	27.9	22.2
Notropis topeka	53	27	0.9167*	4	68.7	42.2
Luxilus cornutus	23	17	0.8236*	0	75.2	45.1
Phoxinus erythrogaster	15	23	0.6985*	8	67.5	45.4
Noturus exilis	22	27	0.8734*	12	49	37.0
Noturus flavus	37	40	0.7923*	9	76.1	49.5
Fundulus zebrinus	34	24	0.7588*	6	86.3	72.8
Gambusia affinis	72	69	0.7571*	5	82.9	70.5
Etheostoma cragini	44	46	0.876*	7	60.5	39.8
Etheostoma flabellare	30	10	0.7652*	7	57.9	43.3
Morone chrysops	24	27	0.776*	8	53.4	46.0

Numbers of individuals for each taxon used to build and test the model sets are indicated. Area under the curve (AUC) derived from the receiver operating characteristic analysis are shown.

*Significance in a z-test at P < 0.01.

county and grouped in two datasets. County inclusion into either dataset was random. The data from half of the counties were used in GARP to build niche models. Data from the other half of the counties were withheld from analysis and used later as an independent validation data set to evaluate the resulting models (Copas & Corbett 2002).

An initial environmental coverage set characterising climate, physiographic and hydrologic parameters included 29 variables in the form of raster grids with a 0.01×0.01 pixel resolution available from USGS (http://www.usgs.gov), the Hydro1k dataset (http://edcdaac.usgs.gov/gtopo30/hydro/) and the Intergovernmental Panel on Climate Change worldwide climate data (http://ipcc-ddc.cru.uea.ac. uk/index.html) were used in the analysis (Table 2). Strahler stream order was derived from a stream map for Kansas available from the Kansas Geospatial Community Commons (http://gisdasc.kgs.ku.edu). This coverage set was subjected to a jackknife procedure for all taxa: an analysis designed to maximise predictive accuracy and cull variables prone to spurious over-fitting (Peterson & Cohoon 1999; Guisan & Zimmermann 2000). I optimised environmental variable inputs for each species's niche model by minimising model omission error (error of failing to predict known presence; Bowden et al. 2003) in the following manner: 20 niche models were generated for each taxon iteratively using N-1 variables (i.e. excluding a single variable from the coverage set in each model set iteration). Omission error (discussed below) was calculated across each 20 model set and a correlation matrix relating omission error and individual variable inclusion and exclusion was generated. Variables whose inclusion increased omission error for more than 50% of the taxa were excluded from the final coverage set. This process resulted in a coverage set including 13 of the original 29 variables. The

Table 2.	Description	of	environmental	variables	in	the	coverage	set.
----------	-------------	----	---------------	-----------	----	-----	----------	------

Description	Included
Annual average high in precipitation	
Annual average high temperature	Х
Annual average radiant energy	
Annual average runoff	
Annual average snowfall	
Aquifer	
Aspect	Х
Average annual frost free days	
Average surface slope	Х
Bulk soil density	
Depth to seasonally high water table	Х
Elevation	Х
Flow accumulation	Х
Flow direction	
Geologic unit	
Groundwater region	
High value in range for surface slope	
January average high temperature	Х
January average precipitation	
July average high temperature	Х
July average precipitation	Х
Low values in range for surface slope	
Potential vegetation coverage	Х
Soil organic matter content	Х
Soil permeability	
Soil thickness	
Soil wind erodibility	
Strahler stream order	Х
Wetness index {=In[flow accumulation/tan(slope)]}	Х

Variables included in the final analysis are indicated with an X.

13 variables included in the final coverage set are indicated in Table 2.

Distributional predictions were generated using DesktopGARP, a beta-version program developed at the University of Kansas (R. Scachetti-Pereira www.lifemapper.org/desktopgarp/). DesktopGARP is based on the GARP algorithm originally developed by David Stockwell (Stockwell & Noble 1992; Stockwell & Peters 1999). GARP is a genetic algorithm, a class of algorithms in which a rule-set is 'evolved'

iteratively following a molecular model of evolution (Holland 1975; see Wiley et al. 2003; or Kluza & McNyset 2005 for a more detailed explanation of GARP). This process is repeated until either a user-defined maximum of iterations is reached (=1000) or until a user-designated convergence criterion for predictive accuracy is reached (=0.01). The resulting model is a rule set in the form of logic strings describing a species' niche as a multi-dimensional hypervolume in ecological space (Hutchinson 1957; Stockwell & Peters 1999).

The model is then projected onto geography, highlighting areas in which the modelled niche parameters are realized, resulting in a GIS grid (ArcView v3.2, Environmental Systems Research Inc, 2003). At the beginning of each model building process, the locality data are randomly subset into a training and 'intrinsic' testing dataset in operator determined proportions (80% training and 20% testing, herein). Each model is tested internally using the intrinsic testing data set, to give an estimate of predictive accuracy in the form of a 'confusion matrix', a 2×2 square matrix in which *a* is the number of points correctly predicted present (true positives), b is the number of points incorrectly predicted as present (false positives), c is the number of points incorrectly predicted as absent (false negatives), and d is the number of points correctly predicted absent (true negatives).

Model evaluation and choice

As there is a stochastic element inherent in GARP model generation, models can differ markedly in their characteristics of omission and commission. Metrics based on the confusion matrix were used to choose a 'best subset' of the models generated. In this study, 200 models were generated for each species from which a 10-model best subset was chosen (limited experimentation with greater initial numbers of models did not indicate marked improvement in the best-subset quality), as described below. Ten best subsets models were developed as no change in accuracy metrics was observed when more models were included.

For each model, omission error [c/(a+c)] was calculated for the intrinsic test data and all models with omission error >0 were eliminated. For this analysis, we have no reason to accept models which fail to predict areas where we know the taxon occurs. For remaining (zero omission) models, a commission index (=pixels predicted present/all pixels) was calculated. As these metrics are calculated using pixels-predicted-present, and a pixel corresponds to actual area on the landscape (approximately 1 km²), the commission index is a measure of the proportional geographic area predicted present by the model (Anderson et al. 2002a). It includes both true commission error (over-prediction into unsuitable habitat), and correct prediction of areas not *known* to be inhabited (for reasons of under-sampling, historical constraint, competitive exclusion, etc.). The median commission index across all low-omission models was calculated; this median has been shown to be the best estimate of the actual geographic area occupied by a given taxon (Anderson et al. 2003). The 10 zero-omission models with commission indices closest to the median were chosen as the best subset, and resulting maps summed.

Statistical analysis

The 10 best model set for each taxon was evaluated using the area under the curve (AUC) in a Receiver Operating Characteristic (ROC) analysis following Wiley et al. (2003). A ROC analysis evaluates the specificity (absence of commission error) and sensitivity (absence of omission error) of a model set (DeLong et al. 1988; Zweig & Campbell 1993; Fielding & Bell 1997). A line for the model set is graphed on a sensitivity (true positive rate) versus 1-specificity (true negative rate) plot, and the AUC is calculated. This AUC is compared with the AUC of a random prediction using a z-test. No difference between the two AUCs indicates that the model set is predicting presence no better than random (Centor 1991). The higher the test AUC, the better the model set; a perfect prediction would have an AUC = 1.0(Hanley & McNeil 1982). A model set achieves a higher AUC as more of the validation data points fall in pixels where more models predict presence. For example, if all of the validation data points occur in pixels where all 10 of the best models predict presence, the maximum AUC for those data is achieved.

Omission error and commission indices were also calculated using the validation data. Omission error was taken as the average percentage across all 10 models of the independent validation points falling in pixels predicted as absent. Commission indices were calculated two ways. First, the percentage of the total area predicted present by the intersection of all 10 of the best models was calculated ('Total Commission'), then the average percentage of each individual model in the best set was calculated ('Average Commission'; Table 1).

Results

Model generation across all taxa reached accuracy convergence (0.01) before reaching the designated

Ecological niche modelling to predict distributions of freshwater fish



Fig. 1. Ten model best-subsets results for *Lepisosteus osseus* and *Morone chrysops* intersected with fourth order and higher streams. Red indicates five or more of the 10 best-subset models predicting presence. Yellow circles indicate training data used to build models; green circles indicate independent validation data. Black-and-white dashed lines denote known extent of ranges (Cross 1967; Cross & Collins 1995). Ks = Kansas River.

maximum iterations (=1000). Across all 12 species, AUCs in ROC analyses ranged from 0.69 for *Phoxinus erythrogaster* to 0.92 for *Notropis topeka* (Table 1). All species' AUCs were significantly different from a line of no information (P < 0.01). Omission errors ranged from 0% for *Luxilus cornutus* to 17% for *Cyprinella camura* (Table 1). These results were general: model sets were accurate regardless of range size, habit, or taxon. Geographic predictions for *Lepisosteus osseus*, *C. camura*, *Fundulus zebrinus* and *Morone chrysops* are shown (Figs 1 and 2). The rule set for one of the 10 best models for *F. zebrinus* is shown in Appendix A.

Discussion

These results indicate that ecological niche modelling using GARP generates accurate models and predictive distributional hypotheses for the included freshwater fishes. These results compliment others obtained with terrestrial and marine taxa (Peterson 2001; Anderson et al. 2002b; Papes & Peterson 2003; Peterson et al. 2003; Wiley et al. 2003; Anderson & Martinez-Meyer 2004). This is significant, particularly given that the environmental variables used in this analysis are publicly available: no specialised data sets or *in situ* data collection is necessary. Many of the included



Fig. 2. Ten model best-subsets results for *Fundulus zebrinus* and *Cyprinella camura* intersected with fourth order and higher streams. Red indicates five or more of the 10 best-subset models predicting presence. Yellow circles indicate training data used to build models; green circles indicate test data. Black-and-white dashed lines denote known extent of ranges (Cross 1967; Cross & Collins 1995). Ak = Arkansas River; Wa = Walnut River.

grids extend worldwide. The ability to generate accurate predictions for freshwater stream fishes using such coverages broadens the scope of possible future analyses to include almost any stream system on the planet. This is not to suggest that better environmental coverage sets are not possible. I endeavoured to include a broadly applicable data set to be of immediate use to the fish research community. There are undoubtedly taxon-specific differences in how important each individual variable is in the final analysis. It is likely that models can be improved on a taxon-by-taxon basis by including or excluding various combinations of variables (Gonzalez-Rebelez et al. 2002). That remains a potential area of inquiry. Also, this analysis includes a single level of spatial and temporal resolution and uses a data set only appropriate for larger-scale analyses (Jackson et al. 2001; Maurer 2002). If a researcher wants to investigate distributional phenomena at a more local scale, a different coverage set must be used (Ricklefs 1987; see Chapters 9–20 in Scott et al. 2002). Ecological niche modelling is flexible in this regard; there is

nothing inherent in either the GARP algorithm or DesktopGARP restricting the spatial resolution of an analysis. Research is limited only by data availability. Fish researchers face unique challenges in this regard. Instream, aquatic data are frequently only available at limited points along stream systems (e.g. at monitoring stations), and data collected often vary in content and completeness even from watershed to watershed. In this analysis. I included a set of derived aquatic variables (e.g. flow accumulation) along with what can be referred to as landscape-level variables (e.g. climate and physiographic variables). Landscape-level variables are most appropriate at regional and continental scales, although a number of studies addressing factors limiting fish species' distributions at different spatial scales have found landscape-scale variables to be as important in limiting local fish distributions as instream variables (Matthews 1987; Matthews & Robison 1988, 1998; Schlosser 1995; Marsh-Matthews & Matthews 2000). The continuous nature of landscape-level coverages may, in fact, be better suited than local scale variables when modelling distributions of aquatic organisms on larger scales. However, the use of these variables makes extracting directly interpretable ecological inference at the local level difficult and may prove dissatisfying to ecologists interested in local-level niche characterisation. Development of coverage sets at multiple scales is a longterm goal, nonetheless the ability to produce robust predictions based solely on the included variables means that many analytical avenues are open immediately. Future inclusion of more and higher-resolution variables should only increase model accuracy and predictivity.

Statistical tests of model-set sensitivity and specificity also remains an area in need of further development. ROC analysis has some properties that make it particularly useful for model-set evaluation. It is a nonparametric and threshold independent test (Hanley & McNeil 1982). So, the data are not expected to conform to a normal distribution, for example, and designation of a decision threshold is unnecessary, as all thresholds are evaluated. It is informative about not only whether a given model set is better than random at predicting presence but also how much better, for a given taxon. However, AUC values are subject to an area effect (Wiley et al. 2003). Although this area effect will not change whether or not a given model set is found to be significant, comparisons of model quality across taxa are not possible. If a taxon is broadly distributed across a landscape, AUCs will be lower for a model set that is as sensitive and specific for that taxon as a model set for a more narrowly distributed taxon. This occurs for two reasons. First, it is simply harder to distinguish a good model set from random when more of the entire area of a given

landscape is occupied by a given species. This effect can be mitigated by including larger potential areas in the analysis that extend beyond the putative range of a taxon, however AUCs for narrowly distributed taxa in the same region will increase as well. With more area to potentially predict as present, a small but correctly predicted area is recognised as being more specific. The second reason has to do with the nature of point occurrence data relative to perceived commission (specificity). Because we do not have a data point in every pixel where that the taxon may actually occur, perceived commission error increases and the AUC decreases. This effect is amplified for broadly distributed taxa where the proportion of data points to actually occupied pixels is usually lower than for more narrowly distributed taxa. This effect can be mitigated by increasing the number of data points included in the analysis, however an AUC of 1.0 will never be achieved even by a highly sensitive and specific model set. It should be noted that these are not failures of the model sets to accurately predict, but limitations of the statistics we use to evaluate them. It is important to understand these limitations when interpreting the results of multi-taxa analyses.

Although commonly used as a test of various modelling systems, measures of commission are problematic by their very nature. First, a direct measure of commission error assumes that the true range of a taxon is known for every pixel in the entire range. Even for well-understood taxa occurring in well-sampled areas, this is not true. Indeed, if it were true the need for analyses of this nature would be more limited. Because of this, commission error is overestimated. For example, the geographic prediction for C. camura (Fig. 2) includes portions of the Walnut River drainage. Cyprinella camura is known to occur there, although this occurrence data set does not happen to include any data points from that drainage. The same is true for F. zebrinus in northern tributaries of the Arkansas River (Fig. 2). In addition, commission as it is measured is actually a combination of true error (that is, an incorrect characterisation of the actual niche of the species) and a number of other historical and taxon-specific factors. Peterson et al. (1999) made the point that what are commonly called 'distributional models' in the literature are correctly interpreted as ecological niche models, and that as such model results are not necessarily a prediction of the actual distribution of the species (although distributions are predicted accurately). In these models, areas may exist in which niche requirements of a species are fulfilled, but the species is absent. These absences from suitable areas may be caused by a variety of factors: exclusion by competition or other interspecific interactions, and absence for historical reasons - either the species never occupied the region because of dispersal

constraints, a speciation event left a sister species in the area, an extinction event occurred, or favourable conditions in that area are ephemeral (Matthews & Robison 1988). Some may see this as limiting the usefulness of GARP analyses, while others will see that this opens the door to a number of analytical avenues. The phenomenon of historical causation can indicate areas of potential range expansion, potential areas for the successful establishment of exotic species, and/or the effects of interactions between species. So, the usefulness of commission to quantify model quality is limited. I include two measures of commission that must be compared with be truly informative, 'total' and 'average' commission (Table 1). Comparing these two numbers for each taxon provides an understanding of the amount of overlap between models within a given model set. It is possible to have a best-model set in which the size of the area predicted within the individual models is similar, but the models have very little geographic overlap. This would be evident in a large difference between total and average commission. A model-set of very low overlap, that is, one with bigger differences between individual models, should be considered suspect relative to a model-set of high overlap. Presumably, there is a truth, and a model-set that comes close to that truth, completely and accurately modelling the niche, is one in which all the models are in complete agreement.

While model agreement across best-model sets is not strictly interpretable as highlighting 'core' habitat across a region, it can be interpreted as areas of higher probability of presence of some portion of the fundamental niche. Because the training data is subject to a random subsetting and resampling procedure during each model-building process, the data points used to build a given model may represent a population occupying a specific, realized portion of the fundamental niche for that taxon (Hutchinson 1957; Malanson et al. 1992). This is not to say that areas where only five of the ten best models agree, for example, are somehow false or outside of the fundamental niche of a taxon. Rather, areas of higher model agreement are areas of commonality across variation in subpopulations and/or subsampled data. Including data points distributed across the known range of a taxon will increase the likelihood of capturing more of the realized portions of a niche so that a better understanding of the fundamental niche can be achieved. Using a best model set rather than a single best model allows for the identification of core areas of agreement across all the realized portions of the niche. This information can be used to identify suitable areas for re-introduction, conservation priority or threat assessment (Peterson et al. 2003; Anderson & Martinez-Meyer 2004).

All of the metrics and statistics reported for each model set have to be considered simultaneously when assessing model set quality. For example, in Fig. 1 the best-model sets for *M. chrvsops* and *L. osseus* are visualised along with the data points used to build and test the model sets. The L. osseus model set has a high AUC (0.85) and a low percentage of omission (6%)with high model agreement (an 8.8% difference). There are large portions of the central Kansas River drainage predicted present for which we happen to have no testing or training data points, however L. osseus is know to have occurred there, at least historically (Cross 1967). The M. chrysops model set has a relatively low, although significant, AUC (0.78), but also a low percentage of omission (8%) and one of the highest model agreements across all the included taxa (a 7.4% difference). I consider both of these to be excellent model sets although the individual metrics vary.

I subjected the data to a severe subsetting process (randomly by county), in part to mimic a common pattern of data availability and to test the modelling system with less-than-ideal data. I also limited the area to within political rather than zoogeographic boundaries (as will often be necessary for regional managers). In limiting the information available to GARP, I am most likely impinging on the quality of the results. This is necessary for testing, however, models should only become more accurate with the inclusion of data from across the entire range of the taxon. In future analyses, I would recommend including more data in the model-building process, while reserving a smaller portion for statistical tests.

Herein, I sought only a test of ecological niche modelling using GARP, however there are a number of research avenues that can arise from this sort of analysis. In terrestrial environments, ecological niche modelling has been used in research into how landscape-level phenomena affect species distributions (Walker 1990; Scott et al. 1996; Anderson et al. 2002b). Other advances include identifying potential distributional areas for poorly-known taxa (Peterson et al. 2002), proposing areas of conservation priority (Bojórquez-Tapia et al. 1995; Godown & Peterson 2000; Peterson et al. 2000; Loiselle et al. 2003), predicting potential distributions of exotic invasive taxa (Peterson & Vieglais 2001; Vander Zanden et al. 2004) and investigating evolutionary conservatism of niche requirements (Peterson et al. 1999). Fisheries researchers are often interested in what specific factors are limiting species distributions at any given spatial and temporal resolution. GARP niche models are a potentially source of insight into that question for the included variables. Rule-sets generated by GARP are usually large and complex. It is common for a single model to be made up of up to 50 individual rules with

each rule including multiple variables relevant to varying portions of a landscape (for an example, see Appendix A). This indicates that the relationship between distributions and the included variables is complex and often interactions between variables are more important than a value of a single variable in limiting distributions. What information can be teased out of these rule-sets remains unexplored.

Another area of future research should investigate ways of manipulating GARP-derived potential distributions down to actual distributions. This poses many more challenges than the ones illustrated in the current analysis. It would be necessary to include all of those factors of historical causation and species interactions discussed earlier. It may be appropriate to subject ecological niche models to an a posteriori process to capture the effects of these phenomena, such as, a decision-tree approach incorporating life-history characteristics (Kolar & Lodge 2002). Ecological niche models can be used as is to predict potential distributions of invasive species by modelling the niche on its native range, and projecting the niche model onto the invaded landscape (Peterson & Vieglais 2001; Iguchi et al. 2004; Kluza & McNyset 2005). They can also be used to predict impacts of climate change on species' distributions by projecting niche models onto climatechanged landscapes (Peterson et al. 2001). More generally, this method may be able to illustrate the effects of diverse factors on geographic distributions of freshwater fishes across landscapes.

Resumen

1. Una innovación esencial en la investigación sobre biodiversidad acuática consistiría en una aproximación robusta para predecir con precisión la distribución potencial de las especies. En este trabajo, analicé la eficacia de modelos de nicho ecológico para predecir las distribuciones potenciales de especies de peces utilizando un algoritmo de inteligencia artificial, el algoritmo genético de 'Rule-Set Prediction, GARP'. Modelos de nicho ecológico de especies se desarrollaron utilizando GARP y fueron proyectados sobre geografía para predecir distribuciones de las especies.

2. Para analizar la validez de esta aproximación utilicé datos de distribuciones de peces de agua dulce de 12 especies que ocurren en Kansas (USA). Se eligieron taxones que representan una diversidad filo-genética, de distribución y de hábitat. Sub-seleccioné estos datos omitiendo la mitad de los condados en la construcción del modelo y analicé modelos utilizando los condados omitidos. Los modelos fueron analizados utilizando análisis 'Receiver Operating Characteristics, ROC'.

3. De las especies analizadas todas fueron estadísticamente significativas con los modelos mostrando una habilidad predictiva excelente. La omisión de errores entre los taxones varío entre 0 y 17%. Esta capacidad de inferencia abre puertas a muchos análisis sintéticos basados en datos de ocurrencia puntual.

Acknowledgements

I am grateful to Ricardo Scachetti-Pereira for his work adapting GARP as a desktop application. I also thank Aimee Stewart at the University of Kansas, Division of Bioinformatics for help with GIS. Thanks to Geff Luttrell, Town Peterson, Dick Robins and Ed Wiley for their support, comments and criticisms. The comments of a number of other reviewers were insightful and helped craft a better paper. Funding for this study was provided by the U.S. National Science Foundation (DEB-9985737) and the Office of Naval Research (N0014-00-1-0887) through the National Oceanographic Partnership Program, the Census of Marine Life, and the Ocean Biogeographic Information System initiative.

References

- Anderson, R.P. & Martinez-Meyer, E. 2004. Modeling species' geographic distributions for preliminary conservation assessments: an implementation with the spiny pocket mice (*Heteromys*) of Ecuador. Biological Conservation 116: 167–179.
- Anderson, R.P., Gomez-Laverde, M. & Peterson, A.T. 2002a. Geographical distributions of spiny pocket mice in South America: insights from predictive models. Global Ecology 11: 131–141.
- Anderson, R.P., Peterson, A.T. & Gomez-Laverde, M. 2002b. Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. Oikos 98: 3–16.
- Anderson, R.P., Lew, D. & Peterson, A.T. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. Ecological Modelling 162: 211–232.
- Environmental Systems Research Inc. 2003. ArcView v3.2. Redlands, CA.
- Austin, M.P., Nicholls, A.O. & Margules, C. R. 1990. Measurement of the realized qualitative niche: environmental niches of five *Eucalyptus* species. Ecological Monographs 60: 161–177.
- Bojórquez-Tapia, L.A., Azuara, I., Ezcurra, E. & Floresvillela, O. 1995. Identifying Conservation Priorities in Mexico through Geographic Information-Systems and Modeling. Ecological Applications 5: 215–231.
- Bowden, G.J., Maier, H.R. & Dandy, G.C. 2003. An evaluation of methods for the selection of inputs for an artificial neural network based river model. In: Recknagel, F., ed. Ecological informatics: understanding ecology by biologically-inspired computation. Berlin, New York: Springer Verlag, pp. 215–232.
- Centor, R. M. 1991. Signal detectability: The use of ROC curves and their analyses. Medical Decision Making 11: 102–106.
- Copas, J.B. & Corbett, P. 2002. Overestimation of the receiver operating characteristic curve for logistic regression. Biometrika 89: 315–331.
- Cross, F.B. 1967. Handbook of fishes of Kansas. Lawrence, Kansas: University of Kansas. 357 pp.
- Cross, F.B. & Collins, J.T. 1995. Fishes in Kansas. Lawrence, Kansas: University of Kansas. 316 pp.
- Cumming, G.S. 2000. Using between-model comparisons to fine-tune linear models of species ranges. Journal of Biogeography 27: 441–455.

- D'Angelo, D.J., Howard, L.M., Meyer, J.L., Gregory, S.V. & Ashkenas, L.R. 1995. Ecological uses for genetic algorithms: predicting fish distributions in complex physical habitats. Canadian Journal of Fisheries and Aquatic Science 52: 1893–1908.
- DeLong, E.R., DeLong, D.M. & Clarke-Peterson, D.L. 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44: 837–845.
- Douglas, M.E. & Matthews, W.J. 1992. Does morphology predict ecology? Hypothesis testing within a freshwater stream fish assemblage. Oikos 65: 121–224.
- Dunham, J.B., Rieman, B.E. & Peterson, J.T. 2002. Patch-based models to predict species occurrence: lessons from salmonid fishes in streams. In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A. & Samson, F.B. eds. Predicting species occurrences: issues of accuracy and scale. Washington, D.C.: Island Press, pp. 327–334.
- Fausch, K.D., Torgersen, C.E., Baxter, C.V. & Li, H.W. 2002. Landscapes to riverscapes: bridging the gap between research and conservation of stream fishes. Bioscience 52: 483–498.
- 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.
- Godown, M.E. & Peterson, A.T. 2000. Preliminary distributional analysis of US endangered bird species. Biodiversity and Conservation 9: 1313–1322.
- Gonzalez-Rebelez, C., Thompson, B.C. & Bryant, F.D. 2002. Influence of selected environmental variables on GIS-habitat models used for gap analysis. In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A. & Samson, F.B. eds. Predicting species occurrences: issues of accuracy and scale. Washington, D.C.: Island Press, pp. 639– 652.
- Grinnell, J. 1917. Field tests of theories concerning distributional control. American Naturalist 51: 115–128.
- Grossman, D.H., Bourgeron, P., Busch, W.N., Cleland, D., Platts, W., Ray, G.C., Robins, C.R. & Roloff, G. 1999. Principles for ecological classification. In: Johnson, N.C., Malk, A.J., Sexton, W.T. & Szaro, R., eds. Ecological stewadrship: a common reference for ecosystem management. Oxford: Elsevier Science Ltd, pp. 353–393.
- Guisan, A. & Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135: 147–186.
- Hanley, J.A. & McNeil, B.J. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143: 29–36.
- Holland, J.H. 1975. Adaptation in natural and artificial systems. Ann Arbor, MI: The University of Michigan Press.
- Hughes, R.M., Rexstad, E. & Bond, C.E. 1987. The relationship of aquatic ecoregions, river basins and physiographic provinces to the ichthyogeographic regions of Oregon. Copeia 2: 423–432.
- Hutchinson, G. E. 1957. Concluding remarks. Populations studies: animal ecology and demography. Cold Springs Harbor Symposium on Quantitative Biology 22: 415–427
- Iguchi, K., Matsuura, K., McNyset, K.M., Peterson, A.T., Scachetti-Pereira, R., Powers, K., Vieglais, D., Wiley, E.O. & Yodo. T. 2004. Predicting invasions of basses in Japan.

Transactions of the American Fisheries Society 133: 845-854.

- Jackson, D.A., Peres-Neto, P.R. & Olden, J.D. 2001. What controls who is where in freshwater fish communities the roles of biotic, abiotic, and spatial factors. Canadian Journal of Fisheries and Aquatic Science 58: 157–170.
- Kluza, D. & McNyset, K.M. 2005. Ecological niche modeling of aquatic invasive species. Aquatic Invaders 16: 1–7.
- Kolar, C.S. & Lodge, D.M. 2002. Ecological predictions and risk assessment for alien fishes in North America. Science 298: 1233–1236.
- Krohn, W. B. 1996. Predicted vertebrate distributions from Gap analysis: considerations in the designs of statewide accuracy assessments. In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A. & Samson, F.B. eds. Predicting species occurrences: issues of accuracy and scale. Bethesda, Maryland: American Society for Photogrammetry and Remote Sensing, pp. 147–169.
- Lamouroux, N., Olivier, J.-M., Persat, H., Pouilly, M., Souchon, Y. & Statzner, B. 1999. Predicting community characteristics from habitat conditions: fluvial fish and hydraulics. Freshwater Biology 42: 275–299.
- Loiselle, B.A., Howell, C.A., Graham, C.H., Goerck, J.M., Brooks, T.M., Smith, K.G. & Williams, P.H. 2003. Avoiding pitfalls of using species distribution models in conservation planning. Conservation Biology 17: 1591–1600.
- Lyons, J. 1989. Correspondence between the distribution of fish assemblages in Wisconsin streams and Omernik's ecoregions. American Midland Naturalist 122: 163–182.
- Malanson, G.P., Westman, W.E. & Yan, Y.-L. 1992. Realized versus fundamental niche functions in a model of chaparral response to climatic change. Ecological Modelling 62: 261–277.
- Marsh-Matthews, E. & Matthews, W.J. 2000. Geographic, terrestrial and aquatic factors: which most influence the structure of stream fish assemblages in the Midwestern United States? Ecology of Freshwater Fish 9: 9–21.
- Matthews, W.J. 1987. Physicochemical tolerance and selectivity of stream fishes as related to their geographic ranges and local distributions. In: Matthews, W.J. & Heins, D.C., eds. Community and evolutionary ecology of North American stream fishes. Norman, OK: University of Oklahoma Press, pp. 111–120.
- Matthews, W.J. & Robison, H.W. 1988. The distribution of the fishes of Arkansas: A multivariate analysis. Copeia 2: 358–374.
- Matthews, W.J. & Robison, H.W. 1998. Influence of drainage connectivity, drainage area and regional species richness on fishes of the interior highlands in Arkansas. The American Midland Naturalist 139: 1–19.
- Maurer, B. A. 2002. Predicting distribution and abundance: Thinking within and between scales. In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A. & Samson, F.B. eds. Predicting species occurrences: issues of accuracy and scale. Washington, D.C.: Island Press, pp. 125–132.
- Nelson, R.L., Platts, W.S., Larsen, D.P. & Jensen, S.E. 1992. Trout distribution and habitat in relation to geology and geomorphology in the North Fork Humboldt River drainage, northeastern Nevada. Transactions of the American Fisheries Society 121: 405–426.

Ecological niche modelling to predict distributions of freshwater fish

- Nix, H.A. 1986. A biogeographic analysis of Australian elapid snakes. In: Longmore, R., ed. Atlas of elapid snakes of Australia. Canberra: Australian Government Publishing Service, pp. 4–15.
- Olden, J.D. 2000. An artificial neural network approach for studying phytoplankton succession. Hydrobiologia 436: 131–143.
- Olden, J.D. & Jackson, D.A. 2001. Fish-habitat relationships in lakes: gaining predictive and explanatory insight by using artificial neural networks. Transactions of the American Fisheries Society 130: 878–897.
- Olden, J.D. & Jackson, D.A. 2002. A comparison of statistical approaches for modelling fish species distributions. Freshwater Biology 47: 1976–1995.
- Papes, M. & Peterson, A.T. 2003. Predicting the potential invasive distribution for *Eupatorium adenophorum* Spreng. in China. Journal of Wuhan Botanical Research 21: 137–142.
- Peterson, A.T. 2001. Predicting species' geographic distributions based on ecological niche modeling. The Condor 103: 599–605.
- Peterson, A.T. & Cohoon, K.P. 1999. Sensitivity of distributional prediction algorithms to geographic data completeness. Ecological Modelling 117: 159–164.
- Peterson, A.T. & Vieglais, D.A. 2001. Predicting species invasions using ecological niche modeling: new approaches from bioinformatics attack a pressing problem. BioScience 51: 363–371.
- Peterson, A.T., Soberón, J. & Sánchez-Cordero, V. 1999. Conservatism of ecological niches in evolutionary time. Science 285: 1265–1267.
- Peterson, A.T., Egbert, S.L., Sánchez-Cordero, V. & Price, K.P. 2000. Geographic analysis of conservation priority: endemic birds and mammals in Veracruz, Mexico. Biological Conservation 93: 85–94.
- Peterson, A. T., Sánchez-Cordero, V., Soberón, J., Bartley, J., Buddemeier, R. W. & Navarro-Sigüenza, A.G. 2001. Effects of global climate change on geographic distributions of Mexican Cracidae. Ecological Modelling 144: 21–30.
- Peterson, A.T., Stockwell, D.R.B. & Kluza, D.A. 2002. Distributional prediction based on ecological niche modeling of primary occurrence data. In: Scott, J.M. et al. ed. Predicting species occurrences: issues of accuracy and scale. Washington, D.C.: Island Press, pp. 617–623.
- Peterson, A.T., Papes, M. & Kluza, D.A. 2003. Predicting the potential invasive distributions of four alien plant species in North America. Weed Science 51: 863–868.
- Ricklefs, R.E. 1987. Community diversity: relative roles of local and regional processes. Science 235: 167–171.

- Schlosser, I.J. 1995. Critical landscape attributes that influence fish population dynamics in headwater streams. Hydrobiologia 303: 71–81.
- Scott, J. M., Tear, T. H. & Davis, F. W. ed. (1996). Gap analysis: a landscape approach to biodiversity planning. Bethesda, Maryland: American Society for Photogrammetry and Remote Sensing, pp. ix+320.
- Scott, J. M., Heglund, P. J. and M. L. Morrison et al. eds (2002). Predicting species occurrences: issues of accuracy and scale. Washington, D.C.: Island Press. 868 pp.
- Stockwell, D.R.B. & Noble, I.R. 1992. Induction of sets of rules from animal distribution data: a robust and informative method of data analysis. Mathematics and Computers in Simulation 33: 385–390.
- Stockwell, D.R.B. & Peters, D. 1999. The GARP modeling system: problems and solutions to automated spatial predictions. International Journal of Geographical Information Science 13: 143–158.
- Stockwell, D.R.B. & Peterson, A.T. 2002. Effects of sample size on accuracy of species distribution models. Ecological Modelling 148: 1–13.
- Taylor, C.M., Winston, M.R. & Matthews, W.J. 1992. Fish species-environment and abundance relationships in a Great Plains river system. Ecography 16: 16–23.
- Tonn, W. M. 1990. Climate change and fish communities: a conceptual framework. Transactions of the American Fisheries Society 119: 337–351.
- Vander Zanden, M.J., Olden, J.D., Thorne, J.H. & Mandrak, N.E. 2004. Predicting occurrences and impacts of smallmouth bass introductions in north temperate lakes. Ecological Applications 14: 132–148.
- Vieglais, D., Wiley, E.O., Robins, C.R & Peterson, A.T. 2000. Harnessing museum resources for the census of marine life: the FishNet project. Oceanography 13: 10–13.
- Walker, P.A. 1990. Modelling wildlife distributions using a geographic information system: kangaroos in relation to climate. Journal of Biogeography 17: 279–289.
- Wiley, E.O., McNyset, K.M., Peterson, A.T., Robins, C.R. & Stewart, A.M. 2003. Niche modeling and geographic range predictions in the marine environment using a machinelearning algorithm. Oceanography 16: 120–127.
- Zweig, M.H. & Campbell, G. 1993. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clinical Chemistry 39: 561–577.

Appendix A

Rule set:	
IF	$-$ elevation*0.0000 $-$ flow_acc*0.0156 $-$ slope*0.0039 $+$ wetness_ind*0.0313 $-$ precip_july*0.0039 $-$ temp_Jan*0.0000 $-$ temp_July*0.0039 $-$ depth water table*0.0000 $-$ strabler*0.0039 $-$ 0.5000
THEN	
	SP = FILOLINOL
IF	- aspect*0.0039 - flow_acc*0.0039 + wetness_ind*0.0742 - organic_mat*0.0039 - precip_july*0.0039 - potveg*0.0000 -
TUEN	
2 pagatad rango rulo	SP = FILOLINOL
IF NOT	aspect=[-1.00,35679.69] AND wetness_ind=[383.91,1570.60] AND organic_mat=[0.12,3.58]AND precip_july=[50.59,199.41] AND potveg=[1.74,94.00] AND temp_Jan=[0.00,5.00]AND temp_July=[25.00,30.00] AND depth_water_table=[2.35,6.35] AND strahler=[0.98.8.03]
THEN	
4 range rule	
IF	flow_acc=[0.00,0.00] AND slope=[0.00,116.29] AND wetness_ind=[368.59,1333.26]AND precip_july=[99.80,200.00] AND pot- veg=[6.51,6.88] AND annual_temp=[15.00,20.00]AND strahler=[0.98,8.03]
THEN	sp = ABSENCE
5 loait	
IF	- elevation*0.0000 + flow acc*0.0195 - slope*0.1250 + wetness ind*0.0742 - precip july*0.0039 - temp Jan*0.0000 -
THEN	temp_July*0.0039 - depth_water_table*0.0000 - strahler*0.0039 - 0.5000
6 range rule	
IF	aspect=[-1.00,35395.38] AND elevation=[258.11,1215.62] AND flow_acc=[0.00,14678.42]AND slope=[0.00,87.84] AND wet- ness ind=[422 19 1517 01] AND precip july=[99 80 101 58]AND potyeg=[92 90 94 37] AND temp july=[25 00 25 00] AND
THEN	
7 range rule	
	conset 1004.09 20277 921 AND elevation 1225 74 1194 201 AND flow cos 10.00.0.001AND elevation 120.001 AND
IF	$aspect=[954.06,0277.05] \text{ AND } elevation=[253.74,1104.05] \text{ AND } now_acc=[0.00,0.00] \text{ AND } stope=[0.00,120.05] \text{ AND } wetness_ind=[391.56,1501.70] \text{ AND } organic_mat=[1.67,3.28] \text{ AND } precip_july=[55.34,200.59] \text{ AND } pot-veg=[1.74,94.00] \text{ AND } temp_Jan=[0.00,5.00] \text{ AND } temp_July=[25.00,30.00] \text{ AND } annual_temp=[15.00,20.00] \text{ AND } temp_July=[25.00,30.00] \text{ AND } annual_temp=[15.00,20.00] \text{ AND } temp_July=[25.00,30.00] \text{ AND } annual_temp=[15.00,20.00] \text{ AND } temp_July=[25.00,30.00] \text{ AND } temp_Jan=[0.00,5.00] \text{ AND } temp_July=[25.00,30.00] \text{ AND } temp_Jan=[1.67,3.28] \text{ AND } temp_July=[25.00,30.00] \text{ AND } temp_July=[25.00,30.00] \text{ AND } temp_Jan=[1.67,3.28] \text{ AND } temp_July=[25.00,30.00] \text{ AND } temp_Jan=[1.67,3.28] \text{ AND } t$
тнем	
	sp = Abstrict
IF	elevation=[258.11,1215.62] AND flow_acc=[0.00,14678.42] AND slope=[0.00,87.84]AND wetness_ind=[422.19,1524.67] AND
	$y = [25.00, 29.98]$ AND depth_water_table=[2.93,6.11] AND strahler=[0.98,8.03]
IHEN	sp = PRESENCE
9 range rule	
IF	aspect=[1278.39,30277.83] AND elevation=[235.74,1184.30] AND flow_acc=[0.00,0.00]AND slope=[0.00,120.00] AND wet- ness_ind=[337.97,1524.67] AND organic_mat=[1.67,3.28]AND potveg=[1.74,94.00] AND temp_Jan=[0.00,5.00] AND temp_July=[25.00,30.00] AND annual_temp=[15.00,20.00] AND depth_water_table=[3.04,5.47]
THEN	sp = ABSENCE
10 range rule	
IF	$aspect = [-1.00,35395.38] \text{ AND } elevation = [258.11,1215.62] \text{ AND } flow_acc = [0.00,14678.42] \text{AND } slope = [2.47,91.55] \text{ AND } wetness_ind = [422.19,1517.01] \text{ AND } organic_mat = [0.24,3.56] \text{ AND } precip_july = [99.80,99.80] \text{ AND } temp_Jan = [0.00,5.00] \text{ AND } temp_July = [25.00,25.00] \text{ AND } annual_temp = [15.00,20.00] \text{ AND } depth_water_table = [2.93,6.11]$
THEN	sp = PRESENCE
11 range rule	
IF	elevation=[258.11,1215.62] AND flow_acc=[0.00,14678.42] AND organic_mat=[0.24,3.56] AND precip_july=[99.80,99.80] AND temp_July=[25.00,30.00] AND depth_water_table=[2.93,6.11] AND strahler=[1.01,7.81]
THEN	sp = PRESENCE
12 logit	
IF	- $aspect^{0.0039}$ - $flow_acc^{0.0039}$ - $slope^{0.0039}$ - $potveg^{0.0078}$ - $temp_Jan^{0.0000}$ - $temp_July^{0.0039}$ - $annual_temp^{0.0039}$ - $depth_water_table^{0.0000}$ - $strahler^{0.0039}$ - 0.5000
THEN	sn = PRESENCE
12 rango rulo	
is range ruid IE	apport_[1 00 25205 201 AND algorithm_1250 11 1015 601 AND flow app. 10 00 14670 401 AND wetaport ind 1460 10 1517 011
IF	aspect= $[-1.00,5359.359]$ AND elevation= $[250.17,1215.05]$ AND how_adc= $[0.00,14076.42]$ AND welless_inte= $[422.19,1517.01]$ AND organic_mat= $[0.24,3.56]$ AND precip_july= $[99.80,99.80]$ AND temp_Jan= $[0.00,5.00]$ AND temp_July= $[24.98,25.02]$ AND annual temp_[15,0.20,001] AND depth water table= $[2,93.6,11]$
THEN	
111⊑N 14 rango rulo	
IH I AIIYE I UIE IE	apport_[1 00 25205 201 AND flow app_[0 00 14679 421 AND close [0 00 97 941 AND support and [420 94 4517 041 AND
IF	$aspect=[-1.00,35395.36]$ AND $now_acc=[0.00,14678.42]$ AND $siope=[0.00,87.84]$ AND wetness_ind=[429.84,1517.01] AND organic_mat=[0.84,3.26] AND potveg=[24.16,94.00] AND temp_Jan=[0.00,5.00] AND temp_July=[25.00,25.00] AND annual_temp=[15.00,20.00] AND depth_water_table=[2.93,6.11] AND strahler=[0.98,8.03]
THEN	sp = PRESENCE

Ecological niche modelling to predict distributions of freshwater fish

Appendix A (Continued)

15 logit	
IF	- elevation*0.0000 - flow_acc*0.0039 - slope*0.1250 + wetness_ind*0.0313 + precip_july*0.0391 - temp_Jan*0.0000 - temp_July*0.0039 - depth_water_table*0.0000 - strahler*0.0039 - 0.5000
THEN 16 range rule	sp = PRESENCE
IF	elevation=[249.17,1220.09] AND flow_acc=[0.00,14678.42] AND slope=[0.00,87.84] AND wetness_ind=[422.19,1517.01] AND potveg=[22.69,94.00] AND temp_July=[24.98,29.94] AND annual_temp=[15.00,20.00] AND depth_water_table=[2.93,6.11] AND strahler=[0.98,8.03]
THEN 17 logit	sp = PRESENCE
IF	- aspect*0.0039 + elevation*0.0078 - flow_acc*0.0039 - slope*0.0039 - wetness_ind*0.0000 - organic_mat*0.0039 - potveg*0.4922 - temp_Jan*0.0000 - temp_July*0.0039 - depth_water_table*0.0000 - 0.5000 sp = PRESENCE
18 range rule	-
	$ \begin{array}{l} \text{AND precip_inly} = [93.6,17,1213.62] \text{ AND now_acc} = [0.00,14073.42] \text{ AND wetress_inl} = [423.64,1332.52] \text{ AND organic_inlat} = [0.24,0.26] \\ \text{AND precip_inly} = [93.80,998.80] \text{ AND potveg} = [27.47,90.32] \text{ AND temp_Jan} = [0.00,5.00] \text{ AND temp_July} = [25.00,30.00] \text{ AND annual_temp} = [15.00,20.00] \text{ AND depth_water_table} = [2.93,6.11] \text{ AND strahler} = [0.98,8.03] \\ \begin{array}{c} \text{AND precip_inly} = [25.00,20.00] \text{ AND depth_water_table} = [2.93,6.11] \text{ AND strahler} = [0.98,8.03] \\ \end{array} $
19 range rule	SP = PRESENCE
IF	elevation=[258.11,1215.62] AND flow_acc=[0.00,14678.42] AND slope=[0.00,87.84] AND wetness_ind=[422.19,1517.01] AND precip_july=[51.19,200.59] AND temp_Jan=[0.00,5.00] AND temp_July=[25.00,25.02] AND annual_temp=[15.00,20.00] AND depth_water_table=[2.93,6.11] AND strahler=[0.98,8.03]
THEN 20 range rule	sp = PRESENCE
IF	$aspect = [-1.00,35395.38] \text{ AND } elevation = [258.11,1211.14] \text{ AND } flow_acc = [0.00,14678.42] \text{ AND } slope = [0.00,87.84] \text{ AND } wetness_ind = [422.19,1517.01] \text{ AND } organic_mat = [0.24,3.56] \text{ AND } precip_july = [99.80,99.80] \text{ AND } potveg = [22.69,94.00] \text{ AND } temp_Jan = [0.18,5.02] \text{ AND } temp_July = [25.00,30.00] \text{ AND } depth_water_table = [2.93,6.11] \text{ AND } strahler = [0.98,8.03]$
THEN 21 range rule	sp = PRESENCE
IF	$aspect = [-1.00,35395.38] \text{ AND } elevation = [258.11,1215.62] \text{ AND } flow_acc = [0.00,14678.42] \text{ AND } slope = [0.00,87.84] \text{ AND } precip_july = [99.80,99.80] \text{ AND } potveg = [22.69,94.00] \text{ AND } temp_July = [25.00,25.00] \text{ AND } annual_temp = [15.00,20.00] \text{ AND } depth_water_table = [2.93,6.11] } sp = PRESENCE$
22 range rule	appent [142 15 25064 00] AND flow and [0.00 14679 42] AND wetagen ind [414 52 1404 04] AND erganic met [0.24 2 56]
THEN	$ \begin{array}{l} \text{AND precip_july} = [99.80, 99.80] \text{AND potveg} = [22.69, 94.00] \text{AND temp_jan} = [0.00, 5.00] \text{AND temp_july} = [25.00, 25.00] \text{AND annual_temp} = [15.00, 20.00] \text{AND depth_water_table} = [2.93, 6.11] \\ \text{sp} = \text{PRESENCE} \end{array} $
IF THEN	$aspect = [-1.00,35395.38] \text{ AND } elevation = [258.11,1215.62] \text{ AND } wetness_ind = [422.19,1517.01] \text{ AND } organic_mat = [0.24,3.56] \text{ AND } precip_july = [99.80,99.80] \text{ AND } potveg = [22.69,94.00] \text{ AND } temp_Jan = [4.98,5.00] \text{ AND } temp_July = [25.00,30.00] \text{ AND } annual_temp = [15.00,20.00] \text{ AND } depth_water_table = [2.93,6.11] \text{ AND } strahler = [0.98,8.03] $ $sp = PRESENCE$
24 range rule IF	aspect=[-1.00,35395.38] AND precip_july=[99.80,99.80] AND potveg=[22.69,94.00] AND temp_Jan=[0.06,5.00] AND temp_July=[25.00.30.00] AND depth water table=[2.93.6.11] AND strabler=[0.47.7.68]
THEN 25 atomic	sp = PRESENCE
IF	aspect=17483.96 AND elevation=562.37 AND flow_acc=0.00 AND slope=40.83 AND wetness_ind=751.40 AND precip_july=99.80
THEN	sp = PRESENCE
IF NOT	elevation=[226.79,1215.62] AND flow_acc=[0.00,29356.84] AND slope=[0.00,117.53] AND organic_mat=[0.14,3.59] AND potveg=[1.00,94.00] AND temp_Jan=[0.00,5.00] AND strahler=[0.98,8.03]
27 range rule	sp = ABSENCE
IF	aspect=[-1.00,35395.38] AND elevation=[258.11,1215.62] AND flow_acc=[0.00,14678.42] AND slope=[0.00,87.84] AND wetness_ind=[422.19,1517.01] AND organic_mat=[0.24,3.56] AND precip_july=[99.80,99.80] AND potveg=[22.69,94.00] AND temp_Jan=[0.00,5.00] AND temp_July=[25.00,30.00] AND annual_temp=[15.00,20.00] and temp_Second
28 range rule	sp - Theorinol
IF	aspect=[-1.00,35395.38] AND elevation=[258.11,1215.62] AND slope=[0.00,87.84] AND wetness_ind=[422.19,1517.01] AND organic_mat=[0.24,3.56] AND precip_july=[99.80,99.80] AND potveg=[22.69,94.00] AND temp_Jan=[0.00,5.00] AND depth_water_table=[2.93,6.11]
THEN	sp = PRESENCE