ECOLOGICAL NICHE MODELING AS A PREDICTIVE TOOL: ASIATIC FRESHWATER FISHES IN NORTH AMERICA

By

Pingfu Chen

Submitted to the Department of Ecology and Evolutionary Biology and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy

____

Chairperson, Dr. E. O. Wiley

Committee members

____

Dr. A. Townsend Peterson

____

Dr. John Kelly

____

Dr. Dick Robins

____

Dr. Stephen L. Egbert

Date defended:____________________________

The Dissertation Committee for Pingfu Chen certifies that
this is the approved version of the following dissertation:

ECOLOGICAL NICHE MODELING AS A PREDICTIVE TOOL: ASIATIC FRESHWATER FISHES IN NORTH AMERICA

Committee:

____

Chairperson

____

____

____

____

Date approved:____________________________

ABSTRACT

A major problem with invasive aquatic species is that they are almost impossible to eradicate once established. Hence, the best method to prevent establishment of such species is to assess their invasive potential proactively and respond appropriately. After introduction, the most effective way is to predict their spread, to discover populations early, and to adopt measures to eradicate or at least contain them. This dissertation uses ecological niches modeling to estimate the ecological requirements of 33 Asiatic freshwater fishes from native-range occurrence points, and to use these data to forecast their invasive potential in North America.

The silver carp, bighead carp, grass carp, black carp, northern snakehead, Asian swamp eel and the oriental weather-fish have been introduced into the United States. Native-range niche models for each species predict known occurrences in North American significantly better than null expectations. The silver and bighead carps are predicted to have the potential to spread throughout the eastern U.S. and selected areas of the West Coast. The black carp was predicted suitable throughout the eastern U.S., and the West Coast. The grass carp was to find suitable habitat in a broader area than black carp, and of being able to extend more in the west. The northern snakehead was predicted being able to spread throughout much of the eastern half of the U.S.

The Asian swamp eel was predicted being able to establish populations in the southern U.S., all of the lower Mississippi River drainage, and the West Coast. The oriental weather-fish was predicted suitable in the entire conterminous USA except the Rocky Mountain and desert areas.

Native-range models for the other 25 fishes suggest that *Myxocyprinus asiaticus*, *Channa maculata*, *Sinilabeo decoru*, *Cirrhinus molitorella* are not likely to establish populations and spread broadly in North America. *Siniperca chuatsi*, *Elopichthys bambusa*, *Micropercops swinohonis, Squaliobarbus curriculus*, *Leuciscus waleckii*, and *Rhodeus ocellatus* may be able to become locally established. *Abbottina rivularis*, *Hemiculter leucisculus*, *Hemibarbus labeo*, *H. maculatus*, *Plagiognathops microlepis*, and *Pseudorasbora parva* have the potential to occupy the entire conterminous USA as the common carp has done.

ACKNOWLEDGMENTS

I want to thank many people for their helps and supports during my studies in Lawrence, Kansas, although I will probably never be able to express my gratitude appropriately and to list them all here. First, I would like to thank my dissertation committee members, Edward O. Wiley, A. Townsend Peterson, John Kelly, Dick Robins, and Stephen Egbert. I particularly thank my major advisor, Edward O. Wiley, for accepting me as one of his students, supporting and encouraging me all these years, and for his patience and consideration. His enthusiasm for new technology and his approach and attitude to research always inspire me.

I thank A. Townsend Peterson for his endless advice and patience. I always remember when I first arrived at KU he kindly offered me a copy of his papers and manuscripts which caused my initial interest in ecological niche modeling. I enjoyed his numerous helpful discussions and learned new ideas and modeling skills in his seminars. I feel I can never thank him too much for his helpful suggestions to improve the draft of my dissertation.

John Kelly, as a reader of my dissertation, deserves many thanks for his helpful suggestions. I benefited a lot from his class and his discussion. I also wish to thank Dick Robins for all the patience and kindness he has shown me. He taught me so much about fishes and science, and helped me learn the local community and its culture. I want to thank Stephen Egbert for his supports and patience. I learned a lot in his class and I look forward to including remotely-sensed data for fine-scaled models in my future work.

I wish to thank W. R. Courtenay and J. D. Williams, USGS Florida Integrated Science Centers Center for Aquatic Resources Studies, for their support and updated information on invasive species studies. The dissertation simply would not have been possible without their assistance.

I want to express my heartfelt thanks to C. Zhang, Beijing Institute of Zoology; J. Yang, Kunming Institute of Zoology; S. He and H. Liu, Wuhan Institute of Hydrobiology, Chinese Academy of Sciences; Wenqiao Tan, Shanghai Fisheries University; for allowing access to their museum data and offering support during my visits.

I think I can never thank Gloria Arratia too much. She advised me on my side research on systematics and paleontology. Although my dissertation doesn't include this work, she patiently instructed me how to prepare manuscripts and submit them for publication. She took her time to review and improve my manuscripts. Her advice and helpful suggestions on my work on systematics and paleontology also trained me to do a better job on my dissertation. Hans-Peter Schultze always shows me his smiles and kindness, and I know he is someone I always can count on for any

questions. He has taught me so much about systematics and paleontology, and edited my papers for publication. I here say thank you to him and feel lucky to be friends with him. Collection manager Andy Bentley introduced FishNet to me and answered all kinds of my questions, and it's a great pleasure to work with him on the fish inventory.

There are so many of my fellow graduate students who have provided all kinds of supports during my time here at KU. In the Division of Ichthyology, I am so lucky to work with so many great graduate students: Kevin Tan, Nancy Holcroft Benson, Kris M. McNyset, K. J. Pittman-Irwin, Anna Clarke, Yana Reid, Justin William, Shannon DeVaney, Matt Davis, Francine Abe, and Kathryn Mickle. They have taught me much about myself, and I feel I am so lucky to know them. Kris M. McNyset actually acted as an unofficial adviser at my early school years at KU. She took her precious time to give me advice on everything from my selection of classes to lab log keeping and research on ecological niche modeling. She showed me how to prepare the data and run the GARP program.

There are many great fellow graduate students whom I wish to thank: Hugo Alamillo, Andrew Cambell, Kartrina Gobetz, Jenny Pramuk, Chris Sheil, Davis Steve, Ismael Hinojosa-Diaz, Sean Maher, David McLeod, Jeet Sukumaran, Nick Rasmussen, Cameron Siler, Jamie Oaks, Monica Papes, Richard Williams, Arpad

Nyari, Elisa Bonaccorso, Juan Manuel Guayasamin, Yoshinori Nakazawa, etc. I have learned so many things beyond ichthyology. They made my study and living at KU full of fun, and I so happy to know them and call them friends.

I would like to thank Debby and Mike Andrew for their hosting my family all these years during my study at KU. They helped my family adapt to local community and learn the language and culture. They patiently answered all kinds of questions my family have. I learned a lot of things from them beyond the college campus.

I own my parents, Jinlin Chen and Songmei Wu, my brother, Pinghe Chen, and my sister, Weihong Chen, more than I can ever pay. Without their love and unwavering support I would not be the person I am today. They always encourage me to go ahead to pursue my dream and helped me remember that pursuing happiness itself is the happiest thing.

Finally, I wish to thank my wife, Yan Li, for her standing by me all these years and taking care of the family; my son, Zhen, for being proud of his dad, and my daughter, Ying, for greeting me with a smile when I came home from work. They, to whom this work is dedicated, are my source of strength, and their love and support mean everything for me.

TABLE OF CONTENTS

INTRODUCTION

Introduced species, also called nonindigenous species, are those that humans move from the species' native range to new geographic regions, intentionally or unintentionally. The fates of these organisms vary tremendously. Many, if not most, of these species enter in transport pathway and perish on the way to a new locality (Kruger et al. 1986). Even if they succeed in reaching a new site, they are likely to be eliminated quickly by a multitude of physical or biotic agents and fail to establish. Those that survive to reproduce may last for only a few generations before going extinct. However, a small fraction of these species persists and becomes established. The minimum size, number, and area extent of these populations have no commonly identified thresholds, although a greater number and frequency of new arrivals do raise the probability that a species will become permanently established (Veltman et al. 1996). Among locally established species, a few become more abundant, spread further, and turn out to be invaders.

According to Thienemann (1950) and Balon (1974), transfers of fish species in Europe through human activities may date as far back as Roman times, when the common carp (*Cyprinus carpio*) from the Danube Basin was used for pond culture in Greece and Italy. During the Middle Ages, the common carp was cultured widely in the monastic or village ponds, and inevitably escaped into adjacent rivers and lakes, spreading through much of Europe. From about the middle of the nineteenth century until 1940, fish introduction intensified primarily for sport fishing or for the nostalgia associated with the familiar things as part of the colonial experience. During this period, many introductions were highly inappropriate and did not withstand the rigors of new habitats. Since 1945, however, fish introductions have increased with commercial globalization, easy travel, and development of advanced techniques for artificial breeding. Introduced species have become an international problem: experience has shown that introduced species have caused serious ecological consequences although, undoubtedly, some temporary beneficial effects can be documented. Introduced species have already caused wholesale alternation of the Earth's biota, changing the roles of native species in the communities, disrupting evolutionary processes, and causing radical changes in abundance or even extinctions. Introduced species have been ranked second behind habitat loss as one of four main causes of a global biodiversity crisis, contributing to about 40% of the extinctions of natural faunas recorded since 1750 (Walker and Steffen 1997, Campbell and Reece 2001).

In the United States alone, there are at least 50,000 introduced species, with a cost to the economy of hundreds of billions of dollars in damage and control efforts, and this figure does not include immense costs associated with loss of native species (Pimentel et al. 2000). The Office of Technology and Assessment (1993) reported on 4542 nonindigenous species that have become successfully established in the USA. Included are >2000 plant species, >2000 insect species, 239 plant pathogens, 142 terrestrial vertebrate species, 91 species of freshwater mollusks, and 70 fish species. Since 1980, 205 species have been introduced or detected, 59 of which are causing economic or environmental harm. Potential economic loss from the most harmful 15 species has been estimated at $134 billion (US) (Leach 1995). Fuller et al.(1999) pointed out that all aquatic sub-regions in the United States (USGS 4 digit HUC level) hold nonindigenous fishes. Numbers of fishes introduced beyond their native ranges in U.S. waters have increased steadily in the last century, from <100 species in 1990 to >500 species in 1998 (Nico and Fuller 1999).

Fishes are introduced for various reasons, including sport, food resources, aquaculture, and bio-manipulation. In new surroundings, introduced fishes are freed from predators, parasites, pathogens, and competitors that have kept them in check in the native ecosystem. Once established, these non-native species can create negative impacts. They may upset the delicate balance of the system, reduce native species, and degrade the ecosystem through competition for food, predation, and bringing diseases. In the United States, nonindigenous species have contributed to the listing of at least 160 native species as endangered or threatened (Office of Technology Assessment 1993).

Research history of species' invasion--Extensive research on the ecology of biotic invasions dates back only a few decades (Elton 1958, Salisbury 1961). Although much has been learned, most data is anecdotal, and the field still lacks definitive synthesis, generalization, and prediction. Current hypotheses or generalizations about traits that distinguish both successful invaders and vulnerable communities all concern extraordinary attributes or circumstances of species or communities. Evaluation of these generalizations has been difficult because they are based on post hoc observation, correlation, and classification, rather than experimentation (Mack et al. 2000).

Elton (1958) was one of the first to review and analyze invasions from an ecological viewpoint. He described 241 instances of species introductions worldwide, and reviewed the worst invaders in terms of economic and ecological damage. He referred to "ecological resistance" to invasion, and found that most these invaders occurred in disturbed habitats, hypothesizing that ecological resistance to invasion is lowered in a disturbed system. Sharples (1983) also found that disturbed areas, like islands, which have simplified biotas, make them less resistant to invasion.

Simberloff (1986) disagreed with the “disturbed system” theory of Elton (1958) and others. He put forward a hypothesis that “each potential invader has a probability of successfully colonizing each site, and this probability rests largely on the nature of its habitat requirements and habitat availability at the site, and only secondarily on what other species are present” (Simberloff 1986). Pimm (1991) found, as a generality, that increased species richness and connectance (i.e., the actual, divided by the possible number of interspecific interactions) decreased the chance of invasion; consequently, species-poor communities will have a greater proportion of invaders than species-rich communities. However, nonindigenous species may facilitate each other’s establishment and/or continued existence, instead of interfering with one another. Crosby (1986) depicted the colonization of the Americas, Australia, New Zealand, and the Canary Islands by European plants, pathogens, and animals, including European humans, which constituted a synergistic juggernaut crushing native peoples and their ecosystems. Ricciardi (2001) found that aquatic nonindegenous species in the Great Lakes facilitate, rather than compete with, one another. This argues against the “biotic resistance” theory that states species-rich communities are less vulnerable to invasion because of competition for limiting resources. Moyle and Light (1996) found that if abiotic factors are appropriate for an exotic species in California streams, then that species is likely to invade successfully, regardless of the biota already present. Where exotic species fail to become established in California streams despite repeated invasions, that failure is best attributed to their inability to adapt to abiotic conditions rather than to biotic resistance on the part of their community (Baltze and Moyle 1993). Generally, species invasion appears possible in any kind of ecosystem as long as the species' ecological conditions are met.

Pathways for nonindigenous fish introduction include intentional stocking, ballast water release, illegal biocontrol, contaminated stocking, aquarium releases, and escapes from aquaculture facilities. Along the sequence of invasion transitions, management options become more constrained: once a nonindigenous species is established, eradication is often impossible, and mitigation and control are difficult and expensive, if possible at all. An ounce of prevention is worth a pound of cure. Most invasions begin with the arrival of a small number of individuals (Simberloff 1986, Mack 1995), and the cost of detroying these is usually trivial compared to the cost and effort of later control, after populations have grown and established. For example, the United States and Canada jointly spend about U.S. $15 million annually to control sea lamprey (*Petromyzon marinus*) in the Great Lakes; these costs have been incurred since 1956 and will continue as long as sea lamprey control remains a management goal (Kolar and Lodge 2002). Obviously, the early transition stages of species invasion are the most important for management because they are the stages at which nonindigenous species can be prevented. Efforts to educate consumers and industries and /or the mandatory application of legally binding species-specific risk assessments could greatly reduce the risks from intentional introductions. Researchers should expand their research to address the earlier stages of invasion. After examining some successes and failures in introduced species policy and management, Simberloff (2003) found that the most effective way to deal with invasive introduced species is to discover them early and attempt to eradicate or at least contain them before they spread. Sound prevention policies need to take into account the potential regions or drainage systems in which the invader are most likely to become established.

Ecological niche and geographic distributions of species--The niche concept is important in understanding broad patterns in the diversity, distribution, and abundance of species. Grinnell (1917) presented an early concept of an ecological niche, defined as the ranges of ecological conditions within which a species is able to maintain populations, emphasizing a place or "recess" in the environment that has the potential to support a species. Hutchinson (1957) defined the fundamental niche of a species as an "n-dimensional hypervolume," every point in which corresponds to a state of the environment which would permit a species to persist indefinitely. A species may be excluded from parts of its fundamental niche because of competition and other biotic interactions. The reduced hypervolume is then termed the realized niche (Hutchinson 1957).

The recent concepts of metapopulations, source-sink dynamics and dispersal limitation make the relationship between niche and distribution more complicated. Pulliam (1988) differentiated between source habitats, where local reproduction exceeds local mortality, and sink habitats, where the opposite holds. Sink habitats, by definition, do not have "conditions necessary and sufficient for a species to carry out its life history" (James et al. 1984); however, large numbers of individuals may occur in sinks because of immigration from source areas (Pulliam 1988). Since a species may frequently be found in unsuitable sites where environmental conditions do not permit it to persist indefinitely in the absence of continued immigration, it has been said that the realized niche is often larger than the fundamental niche, or the range of conditions actually experienced by the species is greater than the range of conditions for which birth rates equal or exceed death rate (Pulliam 2000). As some species are "dispersal limited" (Cain et al. 1998, Clark 1998), they often do not reach, and are therefore often absent from, suitable areas. The theory of metapopulations posits that populations frequently go locally extinct and that, even at equilibrium, only a fraction of suitable habitat will be occupied.

The fundamental ecological niche of a species is a critical determinant of its distribution; as such, it is defined in multidimensional ecological space (MacArthur 1972). Hutchinson (1957) focused the niche concept more on the role of a species within a local community. These concepts can be generally refined by distinguishing between the fundamental and realized niches (Hutchinson 1957), the latter taking into consideration the effects of history and interactions among species. Although modeling fundamental niches based on actual distributional data (drawn from the realized niche) may seem counterintuitive, the broad spectrum of community backgrounds present across a species's geographic distribution allows some degree of insight into this more basic concept of a species's ecological requirements. Although only the realized niches are observable in nature, by examining species across their entire geographic distributions, species's distributional possibilities can be observed against varied community backgrounds, and thus a view of the fundamental ecological niches can be assembled (Peterson et al. 1999, Soberón and Peterson 2005, Soberón 2007).

Information management systems such as geographic information systems (GIS) have been used widely for spatial mapping in conservation biology and ecology. In a GIS framework, multivariate modeling of species' occurrences can be used to predict species' distributions (Nix 1986, Austin et al. 1990, Walker and Cocks 1991, Peterson et al. 1999, Stockwell 1999, Peterson et al. 2002a, Peterson et al. 2002b, Peterson 2003, Wiley et al. 2003, Iguchi et al. 2004, Elith et al. 2006, Guisan et al. 2006, Austin 2007, Chen et al. 2007). These methods use species occurrence data and various GIS data sets summarizing ecological dimensions as explanatory variables to define the species niche in ecological space.

Modeling of species' ecological niches--Ecological niche modeling studies have three basic components: a data set describing the occurrence of the species of interest and a data set of putative explanatory variables; a mathematical model that relates the occurrence data to the explanatory variables; and an assessment of the utility of the model developed in terms of a validation exercise or an assessment of model robustness. Numerous approaches have been used to predict potential distributions based on models of a species's ecological niche. For example, BIOCLIM (Nix 1986) utilizes a boxcar environmental envelope algorithm to identify locations presenting environmental conditions that fall within the environmental range recorded for present occurrences. Specifically, the minimum and maximum values for each environmental predictor are identified to define the multidimensional environmental box where the element is known to occur. Study area sites that have environmental conditions within the boundaries of the multidimensional box are predicted as potential sites of occupancy. Since this method is known to be sensitive to outliers, the predicted distribution is often calculated by disregarding 5% of the lower and higher values for each environmental predictor variable and termed the 'core bioclimate,' and represents the 5-95 percentile limits of the multidimensional environmental box. One can be more or less restrictive by selecting broader or narrower percentile limits to define the environmental conditions where the element is predicted to occur. BIOCLIM, however, is sensitive to outliers and sampling bias, and does not address potential correlations and interactions among environmental variables (Farber and Kadmon 2003).

DOMAIN (Carpenter et al. 1993) uses a point-to-point similarity metric (Gower metric) to assign a classification value to a potential site based on its proximity in environmental space to the most similar positive occurrence location. Similarity between the site of interest and each of recorded present occurrence location is calculated by summing the standardized distance between the two points for each predictor variable. The standardization is achieved by dividing the distance by the predictor variable range for the presence sites, equalizing the contribution from each predictor variable. The standardized distance is subtracted from 1 to obtain the complementary similarity. Values are constrained between 0 and 1 for points within the environmental range of the species occurrences and negative values for sites that fall outside the range. DOMAIN prediction values are the maximum similarity that could be obtained between the site of interest and the set of known occurrences. Predictions are not to be interpreted as predictions of probability of occurrence, but as a measure of classification confidence. It does not address potential correlations and interactions among environmental variables, and gives equal weight to all environmental variables. It is difficult to perform with large sample sizes of occurrence data.

Logistic regression models are commonly used statistical methods for predicting probability of presence/absence. Generalized additive models (GAM) and Generalized linear models (GLM) are used extensively in species' distribution modeling because of their strong statistical foundation and ability to realistically model ecological relationships (Austin 2002). GLMs provide the ability to model response error distributions that are not normally distributed or may not have constant variance functions (McCullagh and Nelder 1989). GLMs consist of a response variable, predictor variables and a link function that describes the relationship between the expected response values and the predictors. This relationship is represented by a linear function in which each predictor is weighted by an estimated coefficient. The inverse logistic transformation of this function is the estimated probability of a positive occurrence and can be expressed as a map of the predicted distribution of occurrence. GLMs assume a linear relationship between the response and its predictors, which is often unrealistic. If a more appropriate shape for the relationship between the response and a predictor is known, the predictor can be transformed and added to the model as an additional predictor term. However, in most cases the relationship expressed in the data is not known a priori or the relationship is too complex to be represented as a higher-order polynomial, usually some combination of linear, quadratic and /or cubic terms.

Generalized additive models (GAM) are a non-parametric extension to GLM in which relationships between predictors and response are represented by a series of non-parametric smoothing functions instead of coefficients (Hastie and Tibshirani 1990). Because of their great flexibility, GAMs are more capable of modeling complex ecological response shapes than GLMs (Manel et al. 1999, Elith et al. 2006). Since the shape of the relationship between the response and the predictors is estimated from the occurrence data, GAMs are extremely data hungry, requiring large sample sizes to estimate these relationships in a manner that generates a model that can make accurate predictions outside the model data.

MaxEnt utilizes a statistical mechanics approach called maximum entropy to make predictions from incomplete information (Phillips et al. 2004, Phillips et al. 2006). MaxEnt estimates the most uniform distribution (maximum entropy) of the occurrence points across the study area given the constraint that the expected value of each environmental predictor variable under this estimated distribution matches its empirical average (average values for the set occurrence data). Similar to logistic regression, MaxEnt weights each environmental variable by a constant. The probability distribution is the sum of each weighed variable divided by a scaling constant to ensure that the probability values range from 0-1. The program starts with a uniform probability distribution and iteratively altering one weight at a time to maximize the likelihood to reach the optimum probability distribution. Continuous environmental data can also be entered as quadratic features and product features, thereby adding further constraints to the estimation of the probability distribution by restricting it to be within the variance for each environmental predictor and covariance for each pair of environmental predictors. Since the traditional implementation of maximum entropy is prone to overfitting, a smoothing procedure called regularization in MaxEnt is used to relax constraints on the estimated distribution researching to the exact empirical average but within the empirical error bounds of the average value for a given predictor. The user has the option to alter the parameters of this procedure to potentially compensate for small sample sizes. MaxEnt's predictions for each analysis cell are 'cumulative values' representing as a percentage the average probability value for the current analysis cell and all other cells with equal or lower probability values. The cell with a value of 100 is the most suitable, while cells close to 0 are the least suitable within the study area.

Recently, artificial intelligence has been applied to the generation of species distribution models, like artificial neural networks, fuzzy logic and genetic algorithms (Stockwell 1999). They are becoming popular because of their ability to elucidate complex patterns, particularly when relationships between variables are non-linear, or data sets violate assumptions inherent in statistical approaches, or interactions between relevant variables predominate.

Genetic Algorithm for Rule-set Production (GARP), originally developed by David Stockwell, at the San Diego Supercomputer Center, is a genetic algorithm that creates an ecological niche model for a species that represents the environmental conditions where that species would be able to maintain populations . GARP uses as input a set of point localities where the species is known to occur and a set of geographic layers representing the environmental parameters that might limit the species' ability to survive. GARP tries, interactively, to find non-random correlations between the presence and absence of the species and the values of the environmental parameters, using several types of rules. Each rule type implements a different method for building species prediction models. Currently there are four types of rules implemented: atomic, logistic regression, bioclimatic envelope and negated bioclimatic envelope rules (for detail see http://nhm.ku.edu/desktopgarp). The set of rules is developed through evolutionary refinement (e.g., truncation, point changes, crossing-over among rules) to maximize predictivity, by testing and selecting rules on random subsets of training data sets (Stockwell 1999). GARP model consists of an ordered series of if-then statements that predict either presence or absence.

Potential invasive fishes from Asia--In the United States, most nonindigenous fish species have come from South America, followed by Asia, Africa, and Central America (Fuller et al. 1999). With the constantly increasing trade and travel between the United States and Asia, there is good reason to worry that more nonindigenous fishes from Asia will enter the United States. In this project, I focus on Asian fishes and the target species are selected mainly based on information from the USGS Nonindigenous Fishes Database (http://nas.er.usgs.gov/fishes/index.html), secondarily from literature, reports, and museum collections.

Some Asiatic fishes studied in this analysis, such as the common carp, grass carp and goldfish, are well established already, while some like the silver carp and bighead carp are suspected to be locally established, and some like the northern snakehead were established, thought to be extirpated, and then rediscovered. Most are latent invaders likely to arrive via the aquarium trade, fish farms, or live fish markets. All these fishes must be monitored and analyzed to ensure that proactive, effective management actions are taken before further damage happens to the natural and managed ecosystems.

Consequently, two questions are crucial for effectively managing the increasing Asian fish invasion: (1) Will these Asian fishes continue to spread in North America, or have some of them already occupied most of their potential range? (2) What areas in the conterminous United States do these invasive fishes inhabit, or what areas are most sensitive to these invasive fishes? Practical questions such as these about the control and management of nonindigenous species require reliable, quantitative forecasts of their potential distribution in the United States.

To study these questions, I use the Genetic Algorithm for Rule-set Prediction (GARP) to build ecological niche models for species that are potential nonindigenous species in North America. I use these models to predict distributional potential of each species in the United States. Thus, policymakers could use these critical data to make species-specific introduction policies at different regions or drainages to prevent further damage to natural and managed ecosystems caused by nonindigenous species.

LITERATURE CITED

Austin, M. P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecological Modelling **157**:101-118.

Austin, M. P. 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. Ecological Modelling **200**:1-19.

Austin, M. P., A. O. Nicholls, and C. R. Margules. 1990. Measurement of the realized qualitative niche: environmental niches of five *Eucalyptus* species. Ecological Monographs:161-177.

Balon, E. K. 1974. Domestication of the carp *Cyprinus carpio* L. Royal Ontario Museum, Life Sciences Miscellaneous Publications:1-37.

Cain, M. L., H. Damman, and A. Muir. 1998. Seed dispersal and the Holocene migration of woodland herbs. Ecological Monographs **68**:325-347.

Campbell, N. A., and B. R. Reece. 2001. Biology 6th edition. Benjamin Cummings.

Carpenter, G., A. N. Gillison, and J. Winter. 1993. Domain: a flexible modeling procedure for mapping potential distributions of animals and plants. Biodiversity and Conservation **2**:667-680.

Chen, P., E. O. Wiley, and K. M. McNyset. 2007. Ecological niche modeling as a predictive tool: silver and bighead carps in North America. Biological Invasions **9**:43-51.

Clark, J. S. 1998. Why trees migrate so fast: confronting theory with dispersal biology and the paleorecord. American Naturalist **152**:204-224.

Crosby, A. W. 1986. Ecological Imperialism. The Biological Expansion of Europe, 900-1900. Cambidge University Press, Cambidge.

Elith, J., C. H. Graham, R. P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. M. Overton, A. T. Peterson, S. J. Phillips, K. Richardson, R. Scachetti-Pereira, R. E. Schapire, J. Soberon, S. Williams, M. S. Wisz, and N. E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography **29**:129-151.

Elton, C. S. 1958. The Ecology of Invasions by Animal and Plants. Chapman & Hall, Methuem, London, UK.

Farber, O., and R. Kadmon. 2003. Assessment of alternative approaches for bioclimatic modeling with special emphasis on the Mahalanobis distance. Ecological Modelling **160**:115-130.

Fuller, P. L., L. G. Nico, and J. D. Williams. 1999. Nonindigenous Fishes Introduced into Inland Waters of the United States. American Fisheries Society, Bethesda, Maryland.

Grinnell, J. 1917. Field tests of theories concerning distributional control. American Naturalist **51**:115-128.

Guisan, A., A. Lehmann, S. Ferrier, M. Austin, J. M. C. Overton, R. Aspinall, and T. Hastie. 2006. Making better biogeographical predictions of species' distributions. Journal of Applied Ecology **43**:386-392.

Hastie, T., and R. Tibshirani. 1990. Generalised Additive Models. Chapman and Hall, Lodon.

Hutchinson, G. E. 1957. Concluding remarks. Cold Spring Harbors Symposium on Quantitative Biology **22**:415-427.

Iguchi, K., K. Matsuura, K. McNyset, A. T. Peterson, R. Scachetti-Pereira, K. A. Powers, D. A. Vieglais, E. O. Wiley, and T. Yodo. 2004. Predicting invasions of North American basses in Japan using native range data and a genetic algorithm. Transactions of the American Fisheries Society **133**:845-854.

James, F. C., R. F. Johnson, N. O. Wamer, G. J. Niemi, and W. J. Boecklen. 1984. The Grinnellian niche of the Wood Thrush. American Naturalist **124**:17-47.

Kolar, C. S., and D. M. Lodge. 2002. Ecological predictions and risk assessment for alien fishes in North America. Science **298**:1233-1236.

Kruger, F. J., D. M. Richardson, and B. W. van Wilgen. 1986. Processes of invasion by alien plants. *in* I. A. W. Macdonald, F. J. Kruger, and A. A. Ferrar, editors. The Ecology and Management of Biological Invasions in Southern Africa. Oxford University Press, Cape Town, South Africa.

Leach, J. H. 1995. Non-indigenous species in the Great Lakes: Were colonization and damage to ecosystem health predictable? Journal of Aquatic Ecosystem Health **4**:117-128.

MacArthur, R. H. 1972. Geographical Ecology: Patterns in the Distribution of Species. Harper & Row Publishers, New York.

Mack, R. N., editor. 1995. Understanding the Processes of Weed Invasions: the Influence of Environmental Stochasticity. British Crop Protection Council, Symposium Proceedings No. 64, Brighton, UK.

Mack, R. N., D. Simberloff, W. M. lonsdale, H. Evans, M. Clout, and F. A.Bazzaz. 2000. Biotic invasions: causes, epidemiology, global consequences, and control. Ecological Applications **10**:689-710.

Manel, S., J. M. Dias, S. T. Buckton, and S. J. Ormerod. 1999. Alternative methods for predicting species distribution: an illustration with Himalayan river birds. Journal of Applied Ecology **36**:734-747.

McCullagh, P., and J. A. Nelder. 1989. Generalized Linear Models, Second edition. Chapman and Hall, Lodon, UK.

Moyle, P. B., and T. Light. 1996. Fish invasions in California: do abiotic factors determine success? Ecology **77**:1666-1670.

Nico, L. G., and P. L. Fuller. 1999. Spatial and temporal patterns of nonindigenous fish introductions in the United States. Fisheries **24**:16-27.

Nix, H. A. 1986. A biogeographic analysis of Australian elapid snakes. Pages 4-15 *in* R. Longmore, editor. Snakes: Atlas of elapid snakes of Australian. Australia Government Publishing Service, Canberra, Australia.

Office of Technology Assessment. 1993. Harmful nonindigenous species in the United States. U.S. Government Printing Office, Washington, D.C. .

Peterson, A. T. 2003. Predicting the geography of species's invasions via ecological niche modeling. Quarterly Review of Biology **78**:419-433.

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

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

Peterson, A. T., D. R. B. Stockwell, and D. A. Kluza. 2002b. Distributional prediction based on ecological niche modeling of primary occurrence data. Pages 617-623 *in* J. M. Scott, editor. Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Washington, D.C.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling **190**:231-259.

Phillips, S. J., M. Dudík, and R. E. Schapire. 2004. A maximum entropy approach to species distribution modeling. Pages 655–662 *in* Proceedings of the 21st International Conference on Machine Learning. ACM Press, New York.

Pimentel, D., L. Lach, R. Zuniga, and D. Morrison. 2000. Environmental and economic costs of nonindigenous species in the United States. Bioscience **50**:53-65.

Pimm, S. L. 1991. The Balance of Nature? The University of Chicago Press, Chicago.

Pulliam, H. R. 1988. Sources, sinks and population regulation. American Naturalist **132**:652-661.

Pulliam, H. R. 2000. On the relationship between niche and distribution. Ecology Letters **3**:349-361.

Ricciardi, A. 2001. Facilitative interactions among aquatic invaders: Is an "invasional meltdown" occurring in the Great Lakes? Canadian Journal of Fisheries and Aquatic Sciences **58**:2513-2525.

Salisbury, S. E. 1961. Weeds and Aliens. St. James's Place, Collins, London, UK.

Sharples, F. E. 1983. Spread of organisms with novel genotypes: thoughts from an ecological perspective. Recombinant DNA Technology Bulletin **6**:43-56.

Simberloff, D. 1986. Introduced insects: a biogeographic and systematic perspective. Pages 3-26 *in* H. Mooney and J. A. Drake, editors. Ecology of Biological Invasions of North America and Hawaii. Springer-Verlag, Berlin.

Simberloff, D. 2003. How much information on population biology is needed to manage introduced species? Conservation Biology **17**:83-92.

Soberón, J. 2007. Grinnellian and Eltonian niches and geographic distributions of species. Ecology Letters **10**:1115-1123.

Soberón, J., and A. T. Peterson. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. Biodiversity Informatics **2**:1-10.

Stockwell, D. R. B. 1999. Genetic algorithms II. Pages 123 -144 *in* A. H. Fielding, editor. Machine Learning Methods for Ecological Applications. Kluwer Academic Publishers, Boston.

Thienemann, A. 1950. Die Binnengewasser. Band 18. Vebreitungsgeschichte der Susswassertierwelt Europas. Schweizerbart, Stuttgart.

Veltman, C. J., S. Nee, and M. J. Crawley. 1996. Correlates of introduction sucess in exotic New Zealand birds. American Naturalist **147**:542-557.

Walker, P. A., and K. D. Cocks. 1991. Habitat: a procedure for modelling a disjoint environmental envelope for a plant or animal species. Global Ecology and Biogeography Letter:108-118.

Walker, W. L., and W. Steffen. 1997. An overview of the implication of global change for natural and managed terrestrial ecosystems. Conservation Ecology **1**:available online http://www.ecologyandsociety.org/vol1/iss2/art2/.

Wiley, E. O., K. M. McNyset, A. T. Peterson, C. R. Robins, and A. M. Stewart. 2003. Niche modeling and geographic range predictions in the marine environment using a machine-learning algorithm. Oceanography **16**:120-127.

CHAPTER 1

PREDICTING THE POTENTIAL GEOGRAPHIC DISTRIBUTION OF SILVER AND BIGHEAD CARPS IN NORTH AMERICA

Chapter Abstract

The silver carp and bighead carp (Cyprinidae), native to eastern Asia, have been introduced into the United States in attempts to improve water quality in aquaculture ponds, reservoirs, and sewage pools. Escaped or released specimens from fish farms have been reported in many states, and both species are already locally established and spreading farther. I used the Genetic Algorithm for Rule-set Prediction (GARP) to model the niches of these two carps in their native ranges using general environmental parameters in concert with native distributional data. The results accurately predicted native occurrence data withheld from the modeling process ($p<0.01$). I then projected the niche models onto the North American landscape. Native niche range models predicted known occurrence data from North American introductions significantly better than null expectations ($p<0.001$). Further, the models suggest that both species have the potential of spreading throughout the eastern U.S. and selected areas of the West Coast.

INTRODUCTION

The silver carp (*Hypophthalmichthys molitrix*) and the bighead carp (*H. nobilis*) are large cyprinid fishes native to eastern Asia that were introduced into Arkansas aquaculture in the early 1970s and have since been introduced widely in the U.S. for phytoplankton control (Nico 2005, Nico and Fuller 2005). Both species are thought to deplete plankton stocks for native larval fishes and mussels (Laird and Page 1996), and might be direct competitors for the adults of native species who feed on plankton, such as paddlefish (*Polyodon spatula*), bigmouth buffalo (*Ictiobus bubalus*) and gizzard shad (*Dorosoma cepedianum*). If so, they are not benign and must be considered invasive species and a threat to native species.

One of the major problems with invasive aquatic species is that they are almost impossible to eradicate once successfully introduced (Courtenay and Stauffer 1984, Williams and Meffe 2000). One of the best methods to prevent the establishment of such invasive species is to develop methods to assess proactively their threat before introduction. After introduction, the only course open is to predict their eventual range and hopefully adopt measures to stop or slow their dispersal across the landscape.

The strategy is as follows. First, model aspects of the fundamental niche (Grinnell 1917) of a species in its native range using a set of native locality data and a set of globally available environmental data. Because the biotic context of an introduction can only be assessed after a species is introduced (Kolar and Lodge 2002), the aspects of the niche modeled are necessarily abiotic. Second, evaluate the resulting models using additional known native occurrence data withheld from the modeling process. Third, use these models to forecast where it might become established outside its native range by visualizing on the new landscape where the native niche occurs.

Testing such tools is essential if they are to be used for regulatory purposes. One way of testing the predictive ability of models is to seek examples of fishes that have already invaded exotic landscapes and see if native niche models can successfully predict occurrence points where an exotic species has been introduced. If the native model successfully predicts known invasion localities, it is reasonable to think that the prediction might also reveal the potential spread of the invasive across the new landscape. Then conservation biologists can assess the threat of species before they become invasive and take measures to prevent their introduction.

One niche-modeling tool, the Genetic Algorithm of Rule-set Prediction (GARP) (Stockwell and Peters 1999), is commonly applied to such forecasts. For example, Iguchi et al.(2004) used native niche models for largemouth and smallmouth basses, native to eastern North America, to post-predict the successful invasion and spread of both species in Japan and to predict their establishment in the northern Japanese island of Hokkaido (discovered in 2001, now established).

This paper aims to: 1) build niche models for silver and bighead carps in their native ranges in Asia; 2) test the accuracy of each niche model within the native range; 3) project the niche model onto North America to assess the potential range of each species, and 4) test this forecast partly with occurrence data from known introductions. With these goals in mind, I will briefly review the biology and distribution of each species within its native range and comment on introductions.

Silver and bighead carps are similar in many respects. Both are fast growing species that reach upwards of 40 kg. Both are efficient plankton strainers. The silver carp is more specialized. It has unique, sponge-like and porous gill rakers capable of straining phytoplankton down to 4 μm in diameter (Robison and Buchanan 1988). Adults feed primarily on phytoplankton, but larvae feed on zooplankton. The bighead carp is less specialized, having comb-like gill rakers. The bighead carp consumes more zooplankton (Robison and Buchanan 1988), and in its native range is considered a zooplankton feeder (Chen et al. 1998). Although primarily large river species, both easily adapt to lakes and ponds if plankton are available.

The silver carp's native distribution reaches from the Yuanjiang and Pearl rivers in the south to the Heilongjinag River drainage in the north. Through artificial propagation and introduction, it is now ubiquitously found in rivers, streams, lakes and reservoirs in most of China. Spawning usually takes place between April and July in the large rivers such as the Yangtze River. Eggs float at the water surface for about 35 h until they hatch. Silver carps normally take three to four years to reach sexual maturity. Kamilov and Salikhov (1996) found silver carp that had established in the Syr Darya River migrated to the communal spawning grounds during the spring flood in April and May, and spawned in small groups of 15 to 25 fish at dusk and dawn, at water temperatures of 18-20°C.

The silver carp was introduced into the United States through the aquaculture trade in the 1970s. By 1980 it was discovered in natural waters, probably a result of escape from aquaculture facilities (USGS 2004). By1991 this species was established in Sougahatchee Creek, Tallapoosa Drainage, Alabama (Courtenay et al. 1991). Now, it is apparently established in Louisiana and possibly in Illinois; and it has been reported in Alabama, Arizona, Arkansas, Colorado, Florida, Indiana, Kansas, Kentucky, Missouri and Tennessee (USGS 2004).

The bighead carp's native distribution is limited to the middle and lower reaches of the Yangtze, Yellow, and Zhujiang rivers. Through artificial propagation and introduction, it is now distributed widely from Hainan Island in southern China, to the Heilongjiang River drainage in northeastern China. Adults are commonly found in the main rivers, calm river bends, and middle and upper waters of lakes and reservoirs. Young inhabit backwaters adjacent to main river channels. Spawning occurs from April to July when rivers rise abruptly and currents increase after heavy rains when water temperature reaches above 18ºC (Chen et al. 1998). The eggs require approximately 35 hours to hatch. The optimum growing temperature is 25-30ºC. It adapts well to the fertile bodies of water that support large plankton populations. Female bighead carp reach sexual maturity in 4-5 years, with males reaching maturity in as little as three years. Mature weight can reach 35-40 kg in China (Chen et al. 1998).

The bighead was first introduced into the United States in 1972 to improve water quality and increase fish production in culture ponds. In the early 1980s, it was discovered in the Ohio and Mississippi rivers, likely as a result of escape from aquaculture facilities (Jennings 1988). Boschung (1992) thought the bighead carp may have been established in the Sougahatchee Creek and the Yates Reservoir, Tallapoosa drainage, Alabama. Occasionally, escaped or released specimens from fish farms have been reported for Florida, and museum records exist for Louisiana and Mississippi (Courtenay et al. 1991). By 1991 the bighead carp was considered established only in Missouri but has been collected from Alabama, Arkansas, Florida, Illinois, Indiana, Kansas and Kentucky (Courtenay et al. 1991). Current data show that the bighead carp is established in the middle and lower Mississippi and Missouri rivers north to Illinois and Missouri (USGS 2004).

METHODS

Environmental Data Sources--Numerous environmental variable data in the form of digital raster grids are available from the United States Geological Survey (USGS: http://www.usgs.gov), including the hydrological Hydro-1K dataset (http://edcdaac.usgs.gov/gtopo30/hydro/) and the Intergovernmental Panel on Climate Change worldwide Climate Data Distribution Centre (http://ipcc-ddc.cru.uea.ac.uk/index.html). In this analysis, 15 environmental variables common to both Asia and North America were used for analysis (see Table 1), which well summarize aspects of topography (elevation, topographic index, flow accumulation, , slope and aspect from USGS Hydro-1K data set), percent tree cover (Hansen et al. 2003), and climatic conditions (annual means of diurnal temperature range; frost days;

precipitation; maximum, minimum and mean monthly temperatures; solar radiation; wet days; and vapor pressure; for 1960-1990 from the Intergovernmental Panel on Climate Change Worldwide Climate Data Distribution Centre). The analyses were confined to the study region 24.5988 – 53.7988˚ N, 66.1417 – 125.0217˚ W (North America) and the species' native region – East Asia (18.83 – 50.69˚N, 96.1616 – 145.7416˚E). The environmental data sets were converted to a resolution of 0.01˚ for model building.

Occurrence Data Sources--Occurrence data for the silver carp and the bighead carp in East Asia were obtained from the Wuhan Institute of Hydrobiology, Beijing Institute of Zoology, Kunming Institute of Zoology, Chinese Academy of Sciences, and scientific literature such as the provincial fish faunas in China, FishNet (http://speciesanalyst.net/fishnet/), and FishBase (http://www.fishbase.org/search.html). Occurrence data for Asian records were georeferenced using the Geonames Query web tool (http://gnpswww.nima.mil/geonames/GNS/index.jsp). In all cases, points outside the known native range were excluded from the training data pool. Duplicate occurrence points were also removed from the data pool, and only verified, unique occurrence points were used for modeling. In total, I obtained 149 and 108 unique occurrence points, respectively, for the silver carp and the bighead carp, from throughout their native distributions.

Occurrence data for both species in the conterminous United States were obtained from USGS Nonindigenous Aquatic Species database (http://nas.er.usgs.gov). Occurrence localities were georeferenced using USGS Geographic Names Information System (http://geonames.usgs.gov/gnishome.html). Township-section-range data were georeferenced using a conversion engine developed by the Montana State University Environmental Statistics Group (www.esg.montana.edu/gl/trs-data.html). Ambiguous records or unspecific localities were excluded from analysis. I obtained 70 and 156 unique occurrence points, respectively, for the silver carp and the bighead carp in the conterminous United Sates.

Evaluating Environmental Variables-- The environmental variables were subjected to a jackknife procedure, which allows exclusion of environmental variables that can lead to spurious overfitting. Hence, for N environmental coverages, N analyses are run using all combinations of N-1 environmental coverages. Then, coverages are evaluated via correlations between inclusion/exclusion of the environmental variables and the average omission error (i.e., predicting absence at sites of known presence). Environmental variables correlated with increased omission error were excluded from further analysis, following Peterson and Cohoon (1999).

Model Building--The native-occurrence data for each species were randomly divided into two data sets. The training data set was used in the modeling process. It consisted of 122 occurrence points for the silver carp and 88 occurrence points for the bighead carp. The validation data set, consisting of 27 native-range occurrence points for silver carp and 20 occurrence points for bighead carp, was withheld entirely from the modeling process and used to test the model set generated by the algorithm.

The use of GARP in ecological niche modeling has been discussed in detail in e. g., Anderson et al. (2003), Drake and Bossenbroek (2004), Feria and Peterson (2002), Peterson (2001), Levine et al. (2004), Tsoar et al. (2007) and Wiley et al. (2003). In essence, GARP attempts to find nonrandom associations between environmental conditions and the known presence of a species by evolving rules that predict presence or absence of the species. To accomplish this task, GARP uses a subset of the training data to formulate a rule and the rest to test the ability of the rule to predict the "new" data. In this study 80% of the training data were used in each model iteration to evolve rules and 20% were used to test rules. As rules are generated, the expectation is that the differences between one round of prediction and the next will decrease, converging on the same predictive efficiency. The investigator can specify this convergence limit (0.01 in this study). Modeling continues until the convergence limit is reached, or a number of iterations specified by the investigator are run (in this study 1,000 times, which was never reached before convergence).

GARP will produce as many models as the investigator specifies. Because of stochastic elements in the process, some replicate models will be much better than others. Two criteria are used to evaluate model quality, omission error and commission "error." Omission error occurs when a model does not predict one or more of the known points. Commission "error" is complex. It is calculated as a function of the area predicted "present" but in which there are no occurrence points. Commission contains both true error (species should not be present but was predicted present), and apparent commission error (species may be present, but site was not sampled, or inhabited because of species' dispersal ability). Among a set of models, those that have low omission error and cluster around the median commission error relative to all models are preferred as a trade off between over and under predicting unknown occurrences (Anderson et al. 2003). This option was used to select a 10-best model set from all the models generated by the algorithm. The "best subsets" option is an implementation of criteria outlined by Anderson (2003). The investigator specifies the number of models to be run and the thresholds of omission and commission errors before the analysis begins. GARP then selects the set of models specified by the investigator that meet the criteria. In this study, I made GARP generate 200 models for each species and set 20 models under the omission threshold of 0% of omission. Of these 20 low omission models, I selected 10 models with a commission closest to the median commission. A minimum of 10 models is necessary to perform statistical tests of the result.

Model Evaluation--The 10 best model set for each species was evaluated using the validation data with a Receiver Operating Characteristic (ROC) analysis, a method designed to evaluate the specificity (absence of commission error) and sensitivity (absence of omission error) of a diagnostic test (Zweig and Campbell 1993, Fielding and Bell 1997). It has been proven to be an efficient approach to test the statistical accuracy of the 10 best model set by Iguchi et al (2004), Wiley et al (2003), and McNyset (2005), and more generally in niche modeling by Elith et al (2006). The area under the curve (AUC) in a ROC analysis is a measure of predictive accuracy for the model set as a whole. For example, if the AUC is 0.50, then the best-model set is performing no better than random, but if the AUC is significantly better than 0.5 (as judged by a z-test), then the result is significant. The higher the value of AUC, the better the model set; and a perfect prediction would have value of 1.0 (Hanley and McNeil 1982). The maximum AUC is achieved when all of the validation data points fall in pixels where all 10 best models predict presence, but it is influenced by the relative extent of the area predicted "present" compared to the total landscape examined and how the species is distributed over the landscape (Wiley et al. 2003, Peterson et al. 2008). Thus, it is meaningless to use the AUC values to compare the model quality across taxa. The accuracy of the 10 best model set for each species was also calculated as the percentage of validation points within10 best models. When all validation points are successfully predicted by all 10 best models, the accuracy is 100%.

RESULTS

Starting with all 15 environmental variables (Table 1-1) the jackknife process excluded average annual wet day frequency, elevation, and average annual diurnal temperature range for the silver carp, and annual average minimum temperature and average annual temperature range for the bighead carp. The remaining variables for each species were used to build the 10 best-subset models. The 10 best-subset models were projected onto the Asian (Figures 1-1B and 1-2B) and North American landscapes (Figures 1-1A and 1-2A). The niche models for silver carp and bighead carp predict independent occurrence data significantly better than random expectation over both the native landscape and the conterminous USA (Table 1-2).

Silver carp models predict the potential for establishment of populations throughout the eastern U. S. and southeastern Canada, with the potential to spread via the Missouri River to the West Coast (Figure 1-1 A). All 70 known occurrence points in the conterminous United States were correctly predicted by at least 9 of the 10 best models (Table 1-2), yielding an AUC score of 0.8019. The accuracy based on these 70 known occurrence points is 93.70%. Native niche models forecast a smaller potential distribution for the bighead carp in North America, with a predicted range from the lower reaches of the Mississippi River drainages and southeastern U.S. north to New Jersey (Figure 2A). Of 156 occurrence points known for bighead carp in the USA, 112 were predicted by ≥9 of 10 best models, indicating a highly significant prediction (AUC=0.8172), although 23 were not predicted by any of the 10 best models (Table 1-2).

DISCUSSION

The predicted distribution for silver carp in the conterminous U.S. is ecologically consistent with its native distribution. The silver carp is native to lowland rivers and spawning in wild populations is confined to major river channels. Recruitment is low in smaller rivers, such as the Qiantang River, in eastern China, although fish with mature gonads are found there. Recruitment near coastal areas is limited because eggs are washed to sea before hatching (Mao and Xu 1991). The major tributaries of the

Mississippi River drainage provide the required spawning and hatching habitats and all these areas are predicted as suitable by the native-range niche models. Therefore, there is good reason to anticipate that silver carp will become established in these areas and spread further.

The native-range niche models also predicted that areas in Oregon, northern California, northern Idaho, and eastern Montana, as well as areas of southern Canada, are suitable for silver carp, so attention should be paid to these areas. The Columbia and Snake river drainages may provide the required spawning habitats for silver carp, as has been the case in northwestern China, where it has successfully colonized the main course of the Hanshui River since it was introduced into northwestern China (Shaanxi Aquaculture Institute and Biology Department of Shaanxi Normal University 1992).

This model predictions do not include the Great Lakes proper. Kolar and Lodge (2002) predicted that silver carp are not a threat in the Great Lakes, using a generalized risk assessment approach and statistical models of fish introduction. However, we do find most of the Great Lakes drainages to be suitable, and considering that silver carp have quickly spread through the upper Mississippi-Illinois river system, special attention should be paid to the lakes themselves.

The predicted potential range for bighead carp in the conterminous U.S. is much more limited than that for silver carp. This is to be expected, as the native range of this species is more limited than that of the silver carp, and many of the records for this species outside its native range are known introductions. Further, although the bighead carp has been introduced widely in China, but it is rarely able to sustain breeding populations in areas outside its native range. The predicted areas in the conterminous U.S. can be interpreted as areas where this species is potentially capable of establishing breeding populations. In the model evaluation, the 23 of 156 points that were not predicted by native-range niche models can be interpreted as inhabitable areas where bighead carp may not be able to maintain breeding populations. According to Dill and Cordone (1997), there is evidence that the California ponds containing bighead carp have spilled since 1989, opening an outlet for bighead carp to gain access to the Sacramento River. This area is predicted inhabitable by native-range niche models and bighead carp may have been established there. A single fish was discovered in West Virginia in 1977 (USGS 2004), also confirming the prediction of native niche models that this area is suitable for bighead carp.

LITERATURE CITED

Anderson, R. P., D. Lew, and A. T. Peterson. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. Ecological Modelling **162**:211-232.

Boschung, H. T. 1992. Catalogue of freshwater and marine fishes of Alabama. Bulletin of the Alabama Museum of Natural History **14**:1-266.

Chen, Y., X. Chu, Y. Luo, Y. Chen, H. Liu, M. He, P. Yue, S. He, W. Chen, and R. Lin, editors. 1998. Fauna Sinica: Osteichthyes -- Cypriniformes II. Science Press, Beijing, China.

Courtenay, W. R., Jr., D. P. Jennings, and J. D. Williams. 1991. Appendix 2. Exotic Fishes. Pages 97-107 *in* C. R. Robins, R. M. Bailey, C. E. Bond, J. R. Brooker, E. A. Lachner, R. N. Lea, and W. B. Scott, editors. Common and Scientific Names of Fishes from the United States and Canada. American Fisheries Society Special Publication 20, Bethesda, Maryland.

Courtenay, W. R., Jr., and J. R. Stauffer, Jr. 1984. Distribution, Biology, and Management of Exotic Fishes. Johns Hopkins University Press, Baltimore, Maryland.

Dill, W. A., and A. J. Cordone. 1997. History and status of introduced fishes in California. California Department of Fish and Game Fish Bulletin **178**:1871-1996.

Drake, J. M., and J. M. Bossenbroek. 2004. The potential distribution of zebra muzzels in the United States. Bioscience **54**:931-941.

Elith, J., C. H. Graham, R. P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. M. Overton, A. T. Peterson, S. J. Phillips, K. Richardson, R. Scachetti-Pereira, R. E. Schapire, J. Soberon, S. Williams, M. S. Wisz, and N. E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography **29**:129-151.

Feria, T. P., and A. T. Peterson. 2002. Using point occurrence data and inferential algorithms to predict local communities of birds. Diversity and Distributions **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. Environmental Conservation **24**:38-49.

Grinnell, J. 1917. Field tests of theories concerning distributional control. American Naturalist **51**:115-128.

Hanley, J. A., and B. J. McNeil. 1982. The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve. Radiology **143**:29-36.

Hansen, C. M., R. S. DeFries, J. R. G. Townshend, M. Carroll, C. Dimiceli, and R. A. Sohlberg. 2003. Global Percent Tree Cover at a Spatial Resolution of 500 Meters: First Results of the MODIS Vegetation Continuous Fields Algorithm. Available from http://glcf.umiacs.umd.edu/data/treecover/index.shtml.

Iguchi, K., K. Matsuura, K. McNyset, A. T. Peterson, R. Scachetti-Pereira, K. A. Powers, D. A. Vieglais, E. O. Wiley, and T. Yodo. 2004. Predicting invasions of North American basses in Japan using native range data and a genetic algorithm. Transactions of the American Fisheries Society **133**:845-854.

Jennings, D. P. 1988. Bighead carp (*Hypophthalmichthys nobilis*): a biological synopsis. U.S. Fish and Wildlife Service **88**:1-35.

Kamilov, B. G., and T. V. Salikhov. 1996. Spawning and reproductive potential of the silver carp *Hypophthalmichthys molitrix* from the Syr Darya River. Journal of Ichthyology **36**:600-606.

Kolar, C. S., and D. M. Lodge. 2002. Ecological predictions and risk assessment for alien fishes in North America. Science **298**:1233-1236.

Laird, C. A., and L. M. Page. 1996. Non-native fishes inhabiting the streams and lakes of Illinois. Illinois Natural History Survey Bulletin **35**:1-51.

Levine, R. S., A. T. Peterson, and M. Q. Benedict. 2004. Geographic and ecologic distributions of the *Anopheles gambiae* complex predicted using a genetic algorithm. American Journal of Tropical Medicine & Hygiene **70**:105-109.

Mao, J., and S. Xu, editors. 1991. The Fauna of Zhejiang Province: Freshwater Fishes. Zhejiang Publishing House of Science and Technology, Hangzhou, China.

McNyset, K. M. 2005. Use of ecological niche modelling to predict distributions of freshwater fish species in Kansas. Ecology of Freshwater Fish **14**:243-255.

Nico, L. 2005. *Hypophthalmichthys molitrix*. *in* Nonindigenous Aquatic Species Database, Gainesville, FL. Retrived from http://flgvwdmz014.er.usgs.gov/queries/FactSheet.asp?speciesID=549 on 24 February 2005.

Nico, L., and P. Fuller. 2005. *Hypophthalmichthys nobilis* *in* Nonindigenous Aquatic Species Database, Gainesville, FL. Retrived from http://flgvwdmz014.er.usgs.gov/queries/FactSheet.asp?speciesID=551 on 24 February 2005.

Peterson, A. T. 2001. Predicting species' geographic distributions based on ecological niche modeling. Condor:599-605.

Peterson, A. T., and K. C. Cohoon. 1999. Sensitivity of distributional prediction algorithms to geographic data completeness. Ecological Modelling **117**:159-164.

Peterson, A. T., M. Papes, and J. Soberón. 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. Ecological Modelling **213**:63-72.

Robison, H. W., and T. M. Buchanan. 1988. Fishes of Arkansas. University of Arkansas Press, Fayetteville, Arkansas.

Shaanxi Aquaculture Institute, and Biology Department of Shaanxi Normal University. 1992. Fishes of Shaanxi Province. Shaanxi Science and Technology Press, Xi'an, China.

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

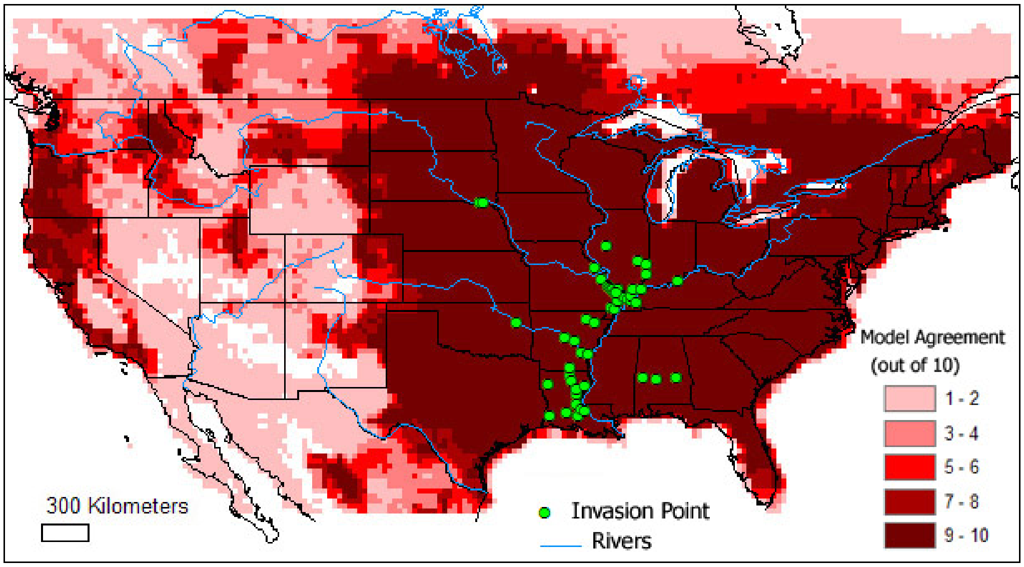
Tsoar, A., O. Allouche, O. Steinitz, D. Rotem, and R. Kadmon. 2007. A comparative evaluation of presenceonly methods for modelling species distribution. Diversity and Distributions **13**:397-405.

USGS. 2004. Nonindigenous Aquatic Species. Biological Resource Division, U.S. Geological Survey (USGS). *in*. Retrieved from http://nas.er.usgs.gov/ on 18 December 2004.

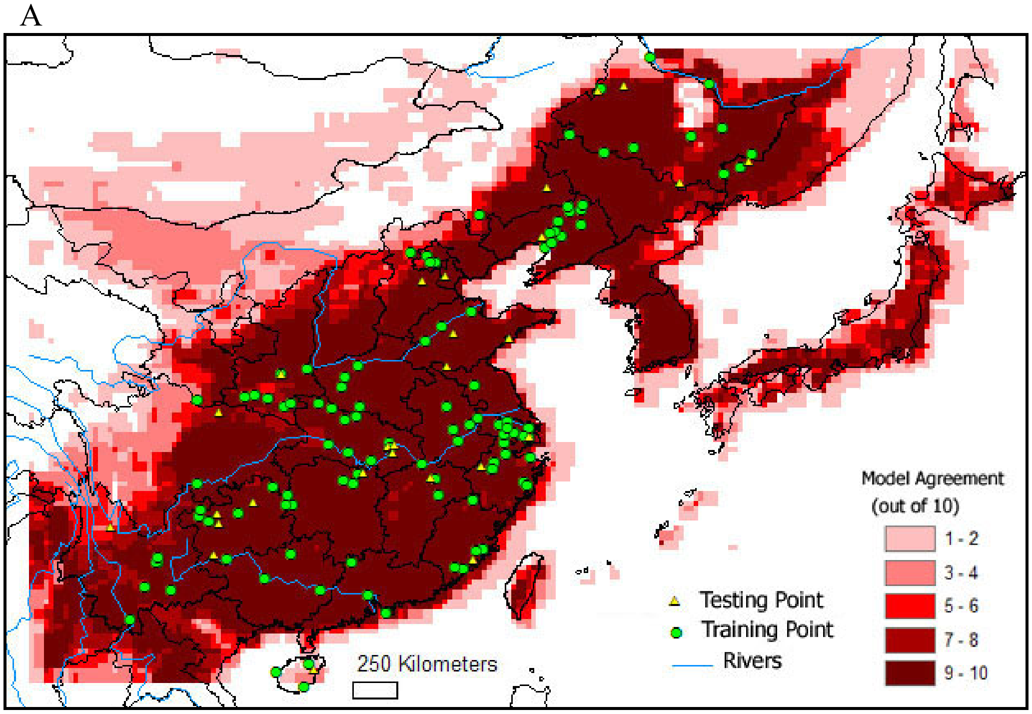
Wiley, E. O., K. M. McNyset, A. T. Peterson, C. R. Robins, and A. M. Stewart. 2003. Niche modeling and geographic range predictions in the marine environment using a machine-learning algorithm. Oceanography **16**:120-127.

Williams, J. D., and G. K. Meffe. 2000. Invasive nonindigenous species: a major threat to freshwater biodiversity. Pages 67-73 *in* R. A. Abell, editor. Freshwater Ecoregions of North America: A Conservation Assessment. Island Press, Washington, D.C.

Zweig, M. H., and G. Campbell. 1993. Reciver-Operating Characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clinical Chemistry **39**:561-577.

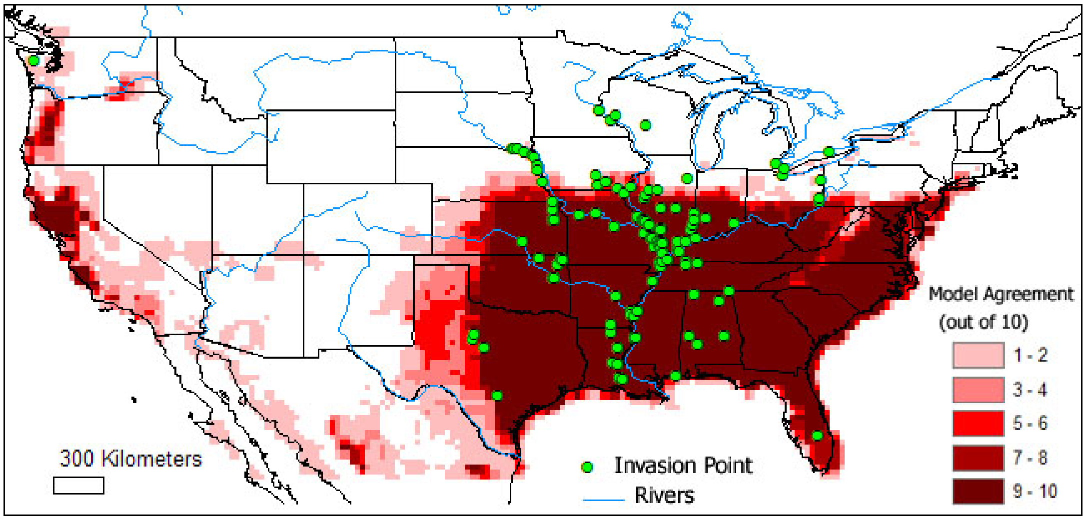


A

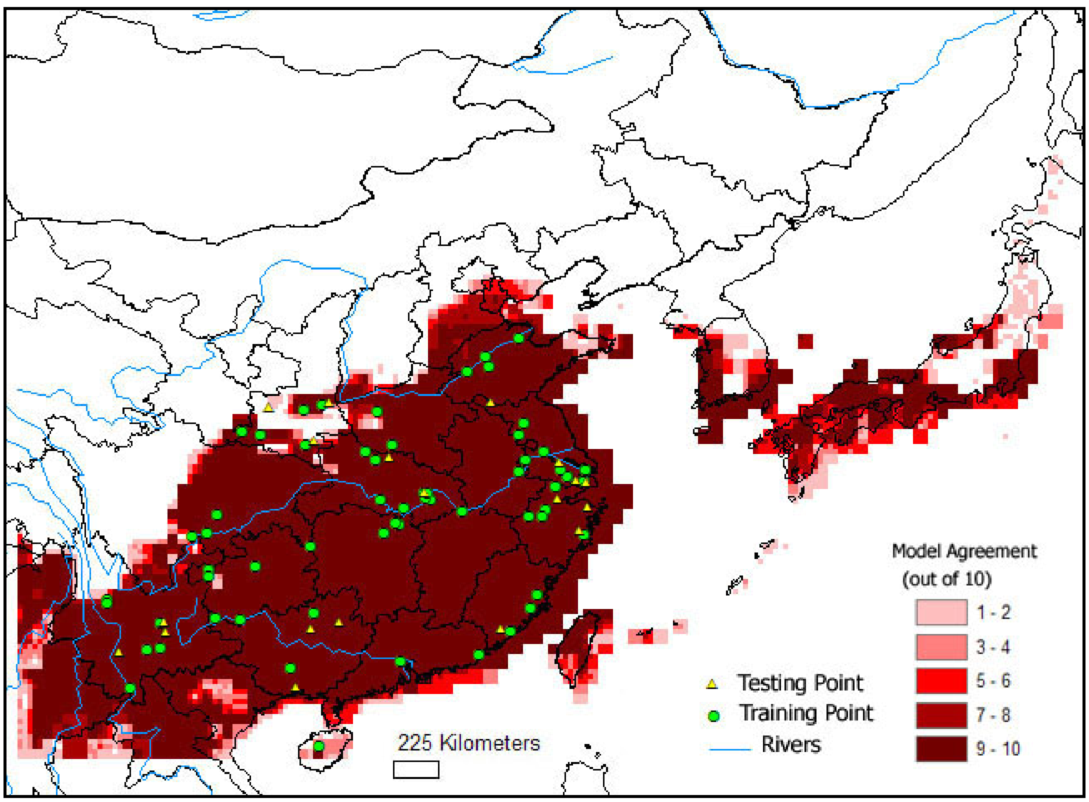


B

Fig. 1-1 A, Native-range model of silver carp projected over the conterminous United States, showing the potential invasive range of northern snakehead; B, Niche model predictions of silver carp over the native landscape, showing probable range. Dark red indicates 9-10 of the 10 best models predicting presence, firebrick 7-8, red 5-6, salmon 3-4, and pink 1-2.



A



B

Fig. 1-2 A, Native-range model of bighead carp projected over the conterminous United States, showing the potential invasive range of northern snakehead; B, Niche model predictions of bighead carp over the native landscape, showing probable range. Dark red indicates 9-10 of the 10 best models predicting presence, firebrick 7-8, red 5-6, salmon 3-4, and pink 1-2.

Table 1-1: Description of environmental layers used in the modeling

Grid name	Description		Source
dtr6190_ann	Diurnal temperature range	1961-1990 annual average	IPCC
frs6190_ann	Ground frost frequency	1961-1990 annual average	IPCC
pre6190_ann	Precipitation	1961-1990 annual average	IPCC
rad6190_ann	Solar radiation	1961-1990 annual average	IPCC
tmn6190_ann	Minimum temperature	1961-1990 annual average	IPCC
tmp6190_ann	Mean temperature	1961-1990 annual average	IPCC
tmx6190_ann	Maximum temperature	1961-1990 annual average	IPCC
vap6190_ann	Vapor pressure	1961-1990 annual average	IPCC
wet6190_ann	Wet day frequency	1961-1990 annual average	IPCC
world_dem	elevation		World Hydro1K dataset
wrld_aspect	aspect		World Hydro1K dataset
wrld_flowa	flow accumulation		World Hydro1K dataset
wrld_slope	slope		World Hydro1K dataset
wrld_topoi	topographic index		World Hydro1K dataset
per_tc	percent tree cover		University of Maryland

Table 1-2: Statistics of model building and evaluation over the native landscape and the conterminous USA. AUC—the area under the curve, SE—standard error, Z—value associated with receiver-operator curve analysis.

Species	Landscape	Training points	Testing points	Model subsets 1-2	3-4	5-6	7-8	9-10	Accuracy	AUC	SE	Z
Bighead carp	Native	88	20	0	0	0	0	19	95%	0.8433	0.0549	8.462**
	USA		156	5	3	6	7	112	76.47%	0.8172	0.0208	24.833**
Silver carp	Native	122	27	0	2	0	1	24	99.71%	0.8019	0.0512	8.3079**
	USA		70	0	0	0	0	70	93.70%	0.8189	0.0309	14.52**

** p<<0.001

CHAPTER 2

PREDICTING THE POTENTIAL GEOGRAPHIC DISTRIBUTION OF BLACK AND GRASS CARPS IN NORTH AMERICA

Chapter Abstract

Black carp and grass carp resemble each other and require similar spawning habitats. Both have been found in the open waters of the USA. In this analysis I use the Genetic Algorithm for Rule-set Prediction (GARP) to find nonrandom associations between environmental variables and the known native presence of the black and grass carps by evolving rules that predict presence or absence of the species to identify suitable areas for these carps on North America. Validated with the independent occurrence points, predictions of the niche models for both black and grass carps are significantly better than random expectations over both the native landscape and the conterminous United States. Black carp was predicted being able to establish populations throughout New England, Mid-Atlantic States, southern states and Midwest states, with the potential to spread via the Missouri River to the West. The native-range niche models forecast a larger potential distribution for the grass carp in North America, with potential range extending farther west.

INTRODUCTION

The black carp, *Mylopharyngodon piceus*, is a large cyprinid, often exceeding 1 m in standard length. The reported maximum individual is about 2 m in total length and weighs over 70 kg. The black carp mainly inhabits the middle and lower layers of waters, and rarely swims to water surface. The optimum water temperature is 22-28˚C. In its growing season it stays in river bends, lakes and ancillary waters, mainly feeding on mollusks and crustaceans, and survives the winter by staying in deep water (Chen et al. 1998). The larvae of black carps feed on zooplankton and fingerlings, and they start to feed on small mollusks and crustaceans when they reach 15 cm in length. The powerful molar-like pharyngeal teeth and the hard callous pad permit the adult black carp to crush the thick shells of large mollusks.

The grass carp, *Ctenopharyngodon idella*, is a large species, often reaching over 1 m in total length with the maximum size of 1.5 m in total length and weight 45 kg. It is native to eastern Asia, from the Heilongjian River of far eastern Russian and China, south to Yuanjiang River of Yunnan, in southern China (Chen et al. 1998). This fish typically lives in quiet waters, such as lakes, ponds, pools, and backwaters of large rivers, and individuals don't travel long distances except for the annual spawning migration. Grass carps grow fast, feeding mainly on macrophytes. Sexual maturity is reached at about 4 years, and spawning usually happens between April and

June in large rivers (Chen et al. 1998). Eggs are semibuoyant and are carried by currents until they hatch. At water temperature between 19.4 ˚C–21.2 ˚C, it takes 35-40 hours for eggs to hatch (East China Sea Fisheries Research Institute et al. 1990). It requires long rivers (50 to 180 km) with sufficient discharge (>400 m^3/sec) and velocity (>0.8 m/sec) for successful reproduction (Stanley et al. 1978).

The grass carp have been widely introduced in the United States, Europe, and elsewhere to control unwanted aquatic plants. Grass carps uproot and eat entire plants. They grow to a large size (to 30 kg) and consume 70%-80% of their body weight daily; they can consequently eliminate all macrophytes in a lake, as happened in a Texas lake where 3650 ha of vegetation were eradicated within 2 years (Martyn et al. 1986). Total elimination of macrophytes is undesirable because, among other effects, it results in destruction of critical habitat for invertebrates and juvenile fishes. Feeding preferences by grass carp can also lead to shifts in relative abundances of different species, altering species composition within the plant assemblage. In experimental ponds, grass carp reduced total plant biomass by feeding preferentially on *Chara* sp., *Elodea* sp., and *Potamogeton pectinatus*. Later, total plant biomass increased over original conditions because those plant species avoided by the grass carp (*Myriophyllum and P. natans*) occupied the space vacated by the preferred plants. When grass carp consume submerged vegetation, floating leafed plants can come to dominate (Fowler and Robson 1978, Shireman et al. 1986). Grass carp are known to compete for food with invertebrates such as crayfish, and with other fish species. They cause significant changes in phytoplankton, invertebrate communities, interfere with reproduction of other fish, and modify habitat by destroying vegetation and water quality (Bain 1993, Xie et al. 2001). Once established, grass carp can eliminate vast areas of aquatic plants that are important as fish food and spawning and nursery habitats. Losses of those habitats can potentially reduce recruitment and abundance of native fishes.

This paper aims to: 1) build niche models for black and grass carps in their native ranges in Asia; 2) test the accuracy of each niche model within the native range; 3) project the niche model onto North America to assess the potential range of each species, and 4) partly test this forecast with occurrence data from known introductions.

METHODS

Environmental Data Sources--15 environmental variables common to both Asia and North America were used for this analysis, which summarize aspects of topography (elevation, topographic index, flow accumulation, slope and aspect from USGS Hydro-1K data set; http://edcdaac.usgs.gov/gtopo30/hydro/), percent tree cover

(Hansen et al. 2003), and climatic conditions (annual means of diurnal temperature range; frost days; precipitation; maximum, minimum and mean monthly temperatures; solar radiation; wet days; and vapor pressure; for 1960-1990 from the Intergovernmental Panel on Climate Change Worldwide Climate Data Distribution Centre; http://ipcc-ddc.cru.uea.ac.uk/index.html). The analyses were confined to the region bounded by 24.5988–53.7988˚ N, 66.1417–125.0217˚ W in North America and the native range in East Asia (18.8300–50.6900˚N, 96.1616–145.7416˚E). The environmental data sets were converted to a pixel resolution of 0.01˚ for analysis.

Occurrence Data Sources–Species occurrence data for the black carp and the grass carp in East Asia were obtained from Wuhan Institute of Hydrobiology, Beijing Institute of Zoology, Kunming Institute of Zoology, Chinese Academy of Sciences, and scientific literature such as the provincial fish faunas in China, FishNet (http://speciesanalyst.net/fishnet/), and FishBase (http://www.fishbase.org/search.html). Occurrence data for Asian records were assigned geographic coordinates using the Geonames Query web tool (http://gnpswww.nima.mil/geonames/GNS/index.jsp). Points outside the known native range were excluded from the training data pool, and duplicate occurrence points were removed, keeping only verified, unique occurrence points for modeling. In total, we obtained 73 and 93 unique occurrence points, respectively, for the black carp and the grass carp, from throughout their native distributions.

Occurrence data for both species in the conterminous United States were obtained from the USGS Nonindigenous Aquatic Species database (http://nas.er.usgs.gov). Occurrence localities were georeferenced using USGS Geographic Names Information System (http://geonames.usgs.gov/gnishome.html). Township-section-range data were georeferenced using a conversion engine developed by the Montana State University Environmental Statistics Group (www.esg.montana.edu/gl/trs-data.html). Ambiguous records or unspecific localities were excluded from analysis. In total, 391 occurrence points of grass carp were collected, and 54 of them are thought to represent established populations in the conterminous United Sates and were used to test the models. Five occurrence points were obtained for the black carp, and were used for model validation over the conterminous United Sates.

Evaluating Environmental Variables--The environmental variables were subjected to a jackknife procedure, which allows exclusion of environmental variables that can lead to spurious overfitting. Therefore, for N environmental coverages, N analyses are run using all combinations of N-1 environmental coverages. Then, coverages are evaluated via correlations between inclusion/exclusion of the environmental variables and the average omission error (i.e., predicting absence at sites of known presence).

Environmental variables correlated with increased omission error were excluded from further analysis, following Peterson and Cohoon (1999).

Model Building--The native-occurrence data for each species were randomly divided into two data sets. The training data set was used in the modeling process. It consisted of 53 occurrence points for the black carp and 73 occurrence points for the grass carp. The validation data set, consisting of 20 native-range occurrence points for the black carp and 20 occurrence points for the grass carp, was withheld entirely from the modeling process and used to test the model predictions.

Details on the use of GARP in ecological niche modeling have been presented in numerous publications, such as Anderson et al. (2003), Feria and Peterson (2002), Levine et al. (2004), and Wiley et al. (2003). In essence, GARP attempts to find nonrandom associations between environmental conditions and known occurrences of a species by evolving rules that predict presence or absence of the species. To accomplish this task, GARP uses a subset of the training data to formulate a rule and the rest to internally test the predictive accuracy of the rule. As rules are generated and evolved, the expectation is that the differences between one round of prediction and the next will decrease, converging on a final solution. The investigator can specify this convergence limit (0.01 in this study). Modeling continues until the convergence limit is reached, or a number of iterations specified by the investigator are run (in this study, 1000 times, which was never reached before convergence).

GARP will produce as many models as the investigator specifies. Because of stochastic elements in the process, some of these replicate models will be ‘better’ (i.e., more predictive) than others. Two criteria are used to evaluate model quality, omission error and commission “error.” Omission error occurs when a model fails to predict potential for presence at known occurrence points. Commission “error” is more complex: it is calculated as the proportional area predicted to be suitable but includes both true error (inappropriate conditions predicted as suitable), and apparent commission error (species may be present, but site was not sampled); hence, I refer to it as the “commission index.” Among a set of models, those that have low omission error rates and that are close to the median commission index appear to offer the best predictive ability (Anderson et al. 2003).

As such, I used the best-subsets option of desktop GARP to select the 10 best models from among the models generated by the algorithm. Here, I generated 200 initial models to derive 20 models under an absolute omission error threshold of 0%, from which I selected the 10 with a commission ‘error’ closest to the median. The 10 models were added together, pixel by pixel, to create a final prediction.

Model Evaluation--The models for each species were then evaluated using the validation data via Receiver Operating Characteristic (ROC) analysis, a method designed to evaluate the specificity (absence of commission error) and sensitivity (absence of omission error) of a diagnostic test (Zweig and Campbell 1993, Fielding and Bell 1997). It has been applied to testing the statistical accuracy of GARP results by Iguchi et al (2004), Wiley et al (2003) and Chen et al (2007), and more generally in niche modeling by Elith et al (2006). The area under the curve (AUC) in a ROC analysis is a measure of predictive accuracy for the model set as a whole: if the AUC = 0.50, then the best-model set is performing no better than random, but if the AUC is significantly higher than 0.5 (as judged by a z-test), then the result is significant. The higher the value of AUC, the better the model set, and a perfect prediction would have AUC = 1.0 (Hanley and McNeil 1982). The maximum AUC is achieved when all of the validation data points fall in pixels where all 10 best models predict presence, but it is influenced by the relative extent of the area predicted “present” compared to the total landscape examined and how the species is distributed over the landscape (Wiley et al. 2003). Thus, it is meaningless to use AUC values to compare model quality across taxa. The accuracy of the native-range niche models for each species was also calculated as the percentage of validation points within the10 best models. When all validation points are successfully predicted by all 10 best models, the accuracy is 100%.

RESULTS

Starting with all 15 environmental variables the jackknife procedure excluded flow accumulation for the black carp, and diurnal temperature range for the grass carp. The remaining variables for each species were used to build final models. The final models were projected onto the Asian (Figures 2-1 B and 2-2 B) and North American landscapes (Figures 2-1 A and 2-2 A). Validated with the independent occurrence points, predictions of the niche models for both black carp and grass carp are significantly better than random expectations over both the native landscape and the conterminous United States (see Table 2-1).

Black carp models predict the potential establishment of the species throughout New England, Mid-Atlantic states, southern states and Midwest states, with the potential to spread via the Missouri River to the west (Figure 2-1 A). The niche models in the native landscape are highly significantly predicting the species distribution. The accuracy based on the independent 20 points is 95.0%, and ROC analysis gets an AUC scores 0.8233. Of the five known occurrence points in the conterminous United States four are correctly predicted by all 10 best models, the remaining one by 8 of 10 best models. The predictive accuracy is 96% (Table 2-1).

The native-range niche models forecast a larger potential distribution for the grass carp in North America, with potential range extending farther west (Figure 2-2

A). The model predictions on the native landscape have an accuracy of 92.5%, and an AUC score of 0.7665 based on the 20 independent test points. Forty-nine of the 54 established points over the conterminous United States were predicted by ≥9 of the 10 best-subset models, and the remaining 5 points were predicted by ≥7 of the 10 best-subset models, yielding an AUC score of 0.7643. The predictive accuracy is up to 97.04% (Table 2-1).

DISCUSSION

The predicted distribution for black carp in the conterminous U.S. is ecologically consistent with its native distribution. The native distribution of black carp is generally limited to the lower valleys of large rivers in eastern Asia. The black carp prefers large rivers and lakes with distinctive seasonal conditions, that is, it requires water temperature below 30˚C in summer and slightly higher than 4° in winter. It cannot live in high-gradient rivers in mountainous regions. Black carp needs low water temperature (after summer) to mature, and currents sufficient to stimulate spawning are also necessary. Spawning only occurs in rushing waters with water temperature around 26˚C. The downstream reach also must be long enough for eggs to hatch, and the river also must have bodies of calm waters for fry to feed and grow. The major tributaries of the Mississippi River drainage provide the required spawning and hatching habitats and all these areas are predicted by the 10-best models.

The black carp was brought to the U.S. in the early 1970's from eastern Asia. In the 1980's, the black carp was imported for use as a food fish and to control the spread of trematodes (parasites) in snails at catfish farms. By the 1990s, this species had been used widely in fish farms in several southern states (Nico et al. 2005), all of which are predicted as suitable for black carp by the niche models. Strict regulations on use and transportation of black carps are necessary to prevent them spreading and being established in open waters.

The escape of black carp occurred during a major flood in April 1994 in Missouri when ≥30 black carp escaped with several thousand bighead carp into the Osage River in Missouri. There has been at least one sighting of this species in the wild on Horseshoe Lake, in Alexander County, Illinois. The fish was reported as a sterile triploid and thus linked to aquaculture escape (Chick et al. 2003). However, Nico and others (2005) investigated other reports, and found evidence that wild populations of black carp may have been present in the lower Mississippi River Basin, particularly in and around the Red River of Louisiana, since the early 1990s. Although reproduction in the Mississippi River has not been documented, once established, the black carp may help to destroy already threatened freshwater mussel and snail populations, as well as other invertebrates. Therefore, there is good reason to worry that black carp could become established in all these areas predicted by the niche models and spread further if no appropriate measures are taken.

The predicted potential range for grass carp in the conterminous United States. is much broader than that for black carp. This is to be expected, as the native range of this species is also larger than that of the black carp, and grass carp is able to adapt to and thrive in ponds and rice fields quite readily. The grass carp was imported into Alabama and Arkansas from eastern Asia in 1963 to control aquatic vegetation, and was intentionally released into the wild shortly thereafter (Guillory & Gasaway, 1978). By 1993, grass carp had been established in Arkansas, Kentucky, Illinois, Loisiana, Missouri, Mississippi, Tennessee and Texas (Courtenay, 1993). Evidence of reproduction has been recorded from all of these states except Texas (Fuller et al., 1999). All of these areas are predicted as suitable by the native-range niche models. The niche models also show grass carp can inhabit northeast Texas, and it may have been established there. The spawning habitats of grass carp and black carp are quite similar. The major tributaries of the Mississippi River drainage can offer the required spawning and hatching habitats for grass carp and all these areas are predicted by the predicting models. If no appropriate changes occur in education, policy, and management, there is a high likelihood that grass carp will be established in all these areas.

Most areas of Florida are predicted by less than five of 10 best models, and more southward by fewer models. This is quite similar to the situations in Guangdong Province and Hainan Island in southern China, where distribution of grass carp is limited by its spawning habitats (Pearl River Fisheries Research Institute of Chinese Academy of Fisheries Science et al. 1991).

The niche models also predicted that areas in Washington, Oregon, northern California and northern Idaho are suitable for black and grass carps. Extra attention should be paid to these areas. Although herbivores are rare or lacking in marine and freshwater fish assemblages above 40˚N or below 40˚S (Horn 1989, Wootton and Oemke 1992), and the black carp is also unlikely to survive in the high gradient rivers in mountainous regions as it is in its native landscapes, the Columbia and Snake River drainages may provide the required spawning habitats for black and grass carps.

This modeling results do not include the Great Lakes proper, which is consistent with Kolar and Lodge (2002) who predicted that black carp would not become established in the Great Lakes even if introduced. They used a generalized risk assessment approach and statistical models of fish introduction. However, most of the Great Lakes drainages and southeastern Canada are predicted suitable for black carp and grass carp. This suggests continued attention to the lakes themselves is necessary. Perhaps the characteristics of the great lakes are substantially different from those on which the models are developed, and the niche models may not be robust to such deviations. Or, the Great Lakes cannot maintain populations just like cases in East Asia where grass and black carps in lakes and ponds which are isolated from large rivers are totally dependent on introduction through artificial propagation (Yang 1987).

The robust prediction for black and grass carps in the conterminous United States suggests that ecological niche models can be used as a proactive tool for combating invasive species. This is significant, particularly considering that the environmental data used in this analysis are publicly available, and these data extend worldwide, making it possible to build and project the niche models at any areas of the world. Research on factors limiting fish species' distributions at different spatial scales have found that landscape scale parameters are important in limiting fish distributions (Marsh-Matthews and Matthews 2000). However, the use of these types of data makes it difficult to interpret the results for aquatic species. For instance, water temperature is determined by the interaction of air temperature with several factors such as dams, tree cover, volume and groundwater input (Allan 1995). The factors limiting species distribution are various and that the most important factors for one species may prove of little effect with another species; it is undoubtedly always a combination of factors which accounts for an animal's geographic range in all parts of the periphery of that range (Grinnell 1917).

LITERATURE CITED

Allan, J. D. 1995. Stream Ecology: Structure and Function of Runing Waters. Chapman and Hall, London.

Anderson, R. P., D. Lew, and A. T. Peterson. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. Ecological Modelling **162**:211-232.

Bain, M. B. 1993. Assessing impacts of introduced aquatic species: grass carp in large systems. Environment Management **17**:211-224.

Chen, P., E. O. Wiley, and K. M. McNyset. 2007. Ecological niche modeling as a predictive tool: silver and bighead carps in North America. Biological Invasions **9**:43-51.

Chen, Y., X. Chu, Y. Luo, Y. Chen, H. Liu, M. He, P. Yue, S. He, W. Chen, and R. Lin, editors. 1998. Fauna Sinica: Osteichthyes -- Cypriniformes II. Science Press, Beijing, China.

Chick, J. H., R. J. Maher, B. M. Burr, and M. R. Thomas. 2003. First black carp captured in the U.S. Science **300**:1876-1877.

East China Sea Fisheries Research Institute, Chinese Academy of Fisheries Science, and Shanghai Fisheries Research Institute, editors. 1990. The Fishes of Shanghai Area. Shanghai Scientific and Technical Publishers, Shanghai.

Elith, J., C. H. Graham, R. P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. M. Overton, A. T. Peterson, S. J. Phillips, K. Richardson, R. Scachetti-Pereira, R. E. Schapire, J. Soberon, S. Williams, M. S. Wisz, and N. E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography **29**:129-151.

Feria, T. P., and A. T. Peterson. 2002. Using point occurrence data and inferential algorithms to predict local communities of birds. Diversity and Distributions **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. Environmental Conservation **24**:38-49.

Fowler, M. C., and T. O. Robson. 1978. The effects of food preferences on stocking rates of grass carp (*Ctenopharyngodon idella* Val.) on mixed plant communities. Aquat Botany **5**:261-276.

Grinnell, J. 1917. Field tests of theories concerning distributional control. American

Naturalist **51**:115-128.

Hanley, J. A., and B. J. McNeil. 1982. The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve. Radiology **143**:29-36.

Hansen, C. M., R. S. DeFries, J. R. G. Townshend, M. Carroll, C. Dimiceli, and R. A. Sohlberg. 2003. Global Percent Tree Cover at a Spatial Resolution of 500 Meters: First Results of the MODIS Vegetation Continuous Fields Algorithm. Available from http://glcf.umiacs.umd.edu/data/treecover/index.shtml.

Horn, M. H. 1989. Biology of marine herbivorous fishes. Oceanography Marine Biology: An Annual Review **27**:167-272.

Iguchi, K., K. Matsuura, K. McNyset, A. T. Peterson, R. Scachetti-Pereira, K. A. Powers, D. A. Vieglais, E. O. Wiley, and T. Yodo. 2004. Predicting invasions of North American basses in Japan using native range data and a genetic algorithm. Transactions of the American Fisheries Society **133**:845-854.

Kolar, C. S., and D. M. Lodge. 2002. Ecological predictions and risk assessment for alien fishes in North America. Science **298**:1233-1236.

Levine, R. S., A. T. Peterson, and M. Q. Benedict. 2004. Geographic and ecologic distributions of the *Anopheles gambiae* complex predicted using a genetic algorithm. American Journal of Tropical Medicine & Hygiene **70**:105-109.

Marsh-Matthews, E., and W. J. Matthews. 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.

Martyn, R. D., R. L. Noble, P. W. Bettoli, and R. C. Maggio. 1986. Mapping aquatic weeds with aerial color infrared photography and evaluating their control by grass carp. Journal of Aquatic Plant Management **24**:46-56.

Nico, L. G., J. D. Williams, and H. L. Jelks. 2005. Black carps: Biological Synopsis and Risk Assessment of an Introduced Fish. American Fisheries Society Special Publication 32, Bethesda, Maryland.

Pearl River Fisheries Research Institute of Chinese Academy of Fisheries Science, South China Normal University, Ji-Nan University, Zhanjiang Fisheries College, and Shanghai Fisheries University. 1991. The Freshwater Fishes of Guangdong Province. Guangdong Science and Technology Press, Guangzhou, China.

Peterson, A. T., and K. C. Cohoon. 1999. Sensitivity of distributional prediction algorithms to geographic data completeness. Ecological Modelling **117**:159-164.

Shireman, J. V., D. E. Colle, and D. E. J. Canfield. 1986. Efficacy and cost of aquatic weed control in small ponds. Water Resource Bulletin **22**:43-48.

Stanley, J. G., W. W. Miley, and D. L. Sutton. 1978. Reproductive requirements and likelihood for naturalization of escaped grass carp in the United States. Transactions of the American Fisheries Society **107**:119-128.

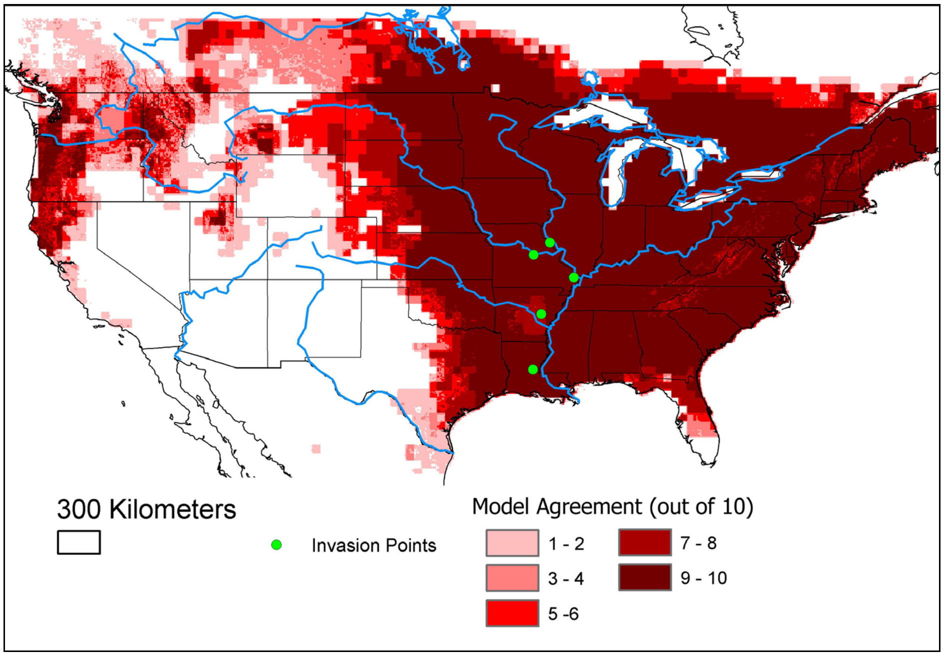
Wiley, E. O., K. M. McNyset, A. T. Peterson, C. R. Robins, and A. M. Stewart. 2003. Niche modeling and geographic range predictions in the marine environment using a machine-learning algorithm. Oceanography **16**:120-127.

Wootton, J. T., and M. P. Oemke. 1992. Latitudinal differences in fish community trophic structure, and the role of fish herbivory in a Costa Rican stream. Environment of Biological Fishes **35**:311-319.

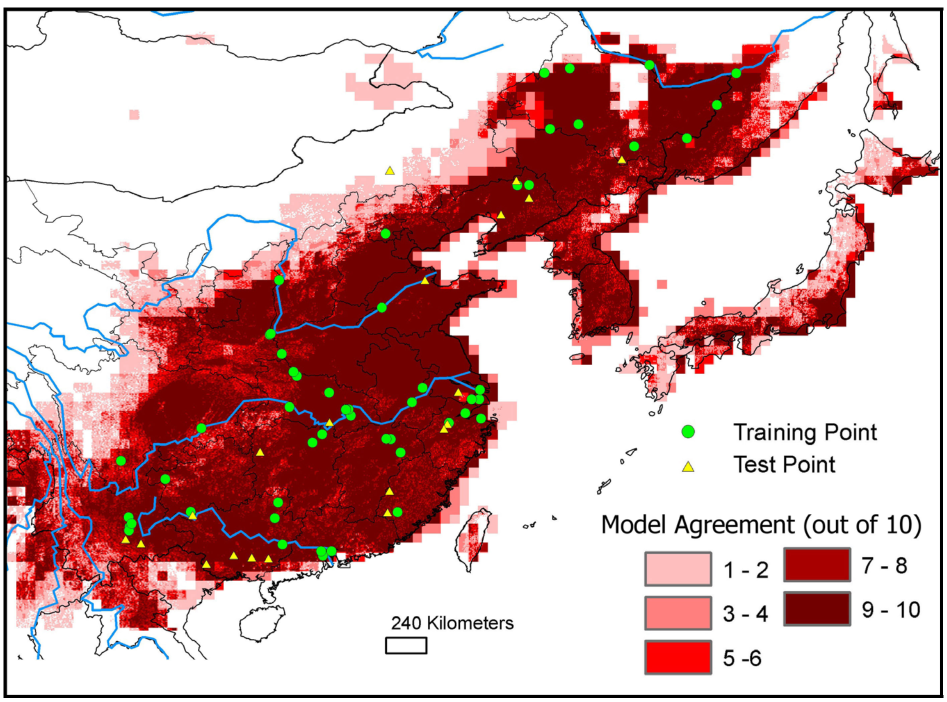
Xie, S., Y. Cui, and Z. Li. 2001. Small fish communities in two regions of Liangzi Lake, China, with or without submersed macrophytes. Journal of Applied Ichthyology **17**:89-92.

Yang, G., editor. 1987. The Fishes of Hubei Province, China. Hubei Science and Technology Press, Wuhan, Hubei.

Zweig, M. H., and G. Campbell. 1993. Reciver-Operating Characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clinical Chemistry **39**:561-577.

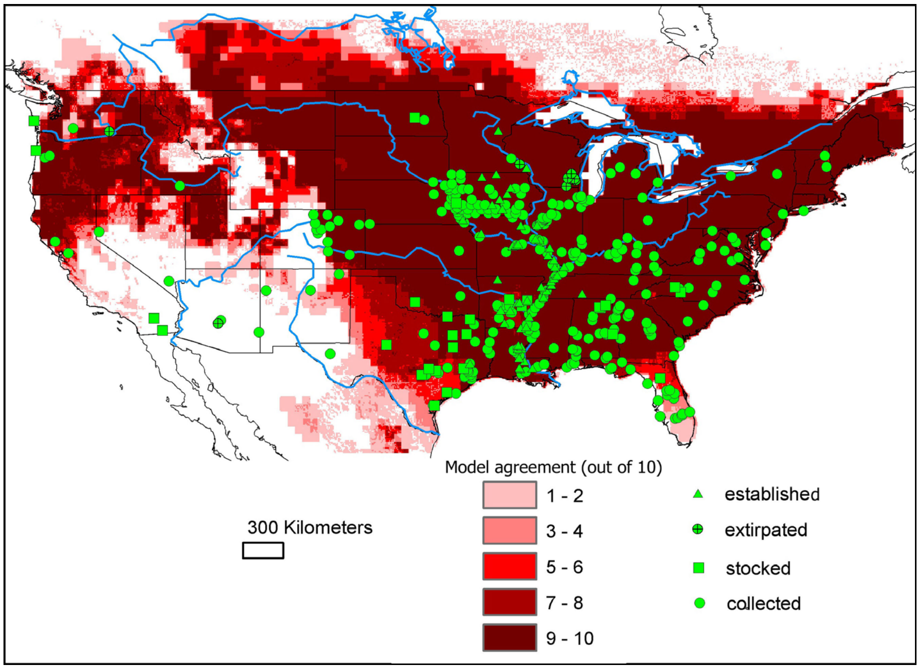


A

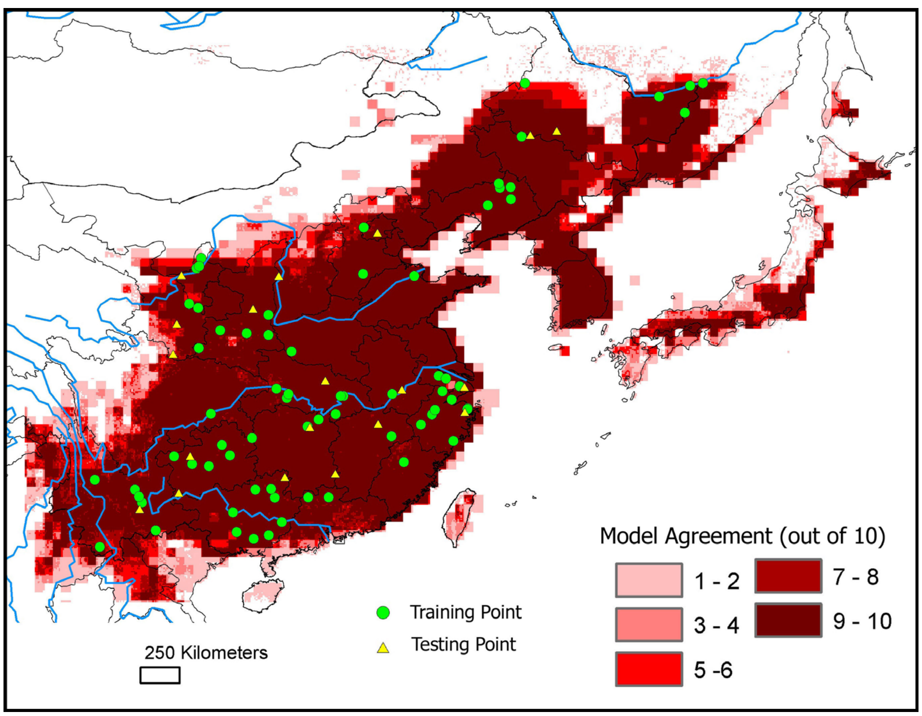


B

Figure 2-1 A, Native-range model of black carp projected over the conterminous United States, showing the potential invasive range of northern snakehead; B, Niche model predictions of black carp over the native landscape, showing probable range. Dark red indicates 9-10 of the 10 best models predicting presence, firebrick 7-8, red 5-6, salmon 3-4, and pink 1-2.



A



B

Figure 2-2 A, Native-range model of grass carp projected over the conterminous United States, showing the potential invasive range of northern snakehead; B, Niche model predictions of grass carp over the native landscape, showing probable range. Dark red indicates 9-10 of the 10 best models predicting presence, firebrick 7-8, red 5-6, salmon 3-4, and pink 1-2.

Table 2-1 Statistics of model building and evaluation over the native landscape and the conterminous USA. AUC—the area under the curve, SE—standard error, Z—value associated with receiver-operator curve analysis.

Species	Landscape	Training points	Testing points	Model subsets 1-2	3-4	5-6	7-8	9-10	Accuracy	AUC	SE	Z
Mylopharyngodon	Native	53	20	0	0	0	0	19	95.00%	0.8233	0.0573	7.4977**
piceus	USA		5		1			4	96%	N/A		
Ctenopharyngodon	Native	73	20	0	0	1	0	18	92.50%	0.7665	0.0616	7.2925**
idellus	USA		54	0	0	0	5	49	97.04%	0.7643	0.0380	12.0620**

** $p<<0.001$

CHAPTER 3

PREDICTING THE POTENTIAL GEOGRAPHIC DISTRIBUTION OF NORTHERN SNAKEHEAD IN NORTH AMERICA

Chapter abstract:

The northern snakehead (*Channa argus*), a voracious predatory fish native to Asia, caused considerable public concern after it was found established in a small pond in Crofton, Maryland, in 2002. Thought having been extirpated from North America, it has now been found several times in the Potomac River and its tributaries. It is critical to know where this fish has the potential to establish populations in order to prepare and plan to prevent its further spread. This paper used the genetic algorithm for rule-set prediction (GARP) to model its ecological niche on its native distributional area using climatic and hydrological information in concert with native range occurrence data. The results predicted native occurrence data withheld from the modeling process accurately (AUC = 0.7992, $P < 0.001$). Upon projecting the native-range niche models onto North American landscapes, Known North America occurrence data were predicted statistically significantly better than random expectations (AUC = 0.7837, $P < 0.001$). Further, the niche models suggest that this species has the potential to spread throughout much of the northeastern, southeastern, and midwestern United States; portions of the Northwest are also predicted as habitat.

INTRODUCTION

With increases in human travel and trade, invasive species have increased tremendously, and they have contributed to declines in native species and to changes in ecosystem function. A major problem with invasive aquatic species is that they are almost impossible to eradicate once successfully introduced (Courtenay and Stauffer 1984, Williams and Meffe 2000). As such, the best method to prevent establishment of such species is to assess their invasive potential proactively before introduction. After introduction, the most effective way is to predict their eventual range, discover them early, and adopt measures to eradicate or at least contain them before they spread across the landscape. This paper uses the genetic algorithm for rule-set production (GARP) to model the ecological niche of northern snakehead from East Asian native distribution points, and uses this ecological niche model to predict the potential invasive range in the conterminous United States.

The northern snakehead (*Channa argus*), a voracious predatory fish of the family Channidae (Perciformes), has been discovered in North American waters (Courtenay and Williams 2004). The northern snakehead is native to China, Russia and Korea, occurring in the middle and lower Heilongjiang (Amur) River basin, Sunghua (Sungari) River basin, Ussuri River basin, Lake Khanka, and across much of the Korean Peninsula (Institute of Hydrobiology Academia Sinica 1976, Ding 1994, Courtenay and Williams 2004). It is widely distributed in the Yangtze River Basin and its tributaries, and has also been reported from the lakes and drainages of Yunnan

Province (Institute of Hydrobiology Academia Sinica 1976). It inhabits shallow ponds or swamps with mud substrate and aquatic vegetation, and is common in canals, reservoirs, lakes, and rivers (Dukravets and Machulin 1978, Dukravets 1992). It can apparently survive out of water for 3-4 days at temperatures ranging 10-15°C.

Juvenile northern snakeheads feed on small crustaceans and fish larvae, while adults feed on fishes, frogs, crustaceans, and aquatic insects. This species has been said to reach sexual maturity in about 3 yr, with a length of 30-35 cm in the Amur and Syr Dar'ya, but some can spawn during the second year (Dukravets and Machulin 1978). This species can spawn one to three, or even up to five, times per year in a circular nest made of aquatic plants in shallow waters, typically in May-June at water temperatures of 18-20˚C. Eggs are pelagic, non-adhesive, spherical, yellow, and about 2 mm in diameter. Numbers of eggs released range 1300 to 15,000. Eggs hatch in 28 hr at 31˚C, 45 hr at 25˚C, and 120 hr at 18˚C. Post-larvae stages feed on plankton, and juveniles on small crustaceans and fish larvae. Once a length of 4 cm has been reached, they begin to feed on fishes and at >13 cm in length fishes dominate (>64%) the diet (Guseva and Zholdasova 1986, Guseva 1990).

Northern snakehead have the traits of most common invaders: high fecundity, short generation time, rapid dispersal, polyphagy, high temperature tolerance, and wide native distribution (Institute of Hydrobiology Academia Sinica 1976, Ding 1994). Thus, this species could turn to be a potentially dangerous invasive species. What is more, Bykhovskaya-Pavlovskaya et al. (1964) cited northern snakehead as hosting 18 parasite species. Should it become established in North American ecosystems, it or its parasite associates could disrupt food webs and ecological conditions, forever changing native aquatic systems. Before being added to the list of injurious wildlife under the Lacey Act in October 2002, which banned import and interstate transport without a U.S. Fish and Wildlife Service permit, northern snakeheads were sold in pet stores and live food fish markets and some restaurants in major U.S. cities, including Boston, New York, and St. Louis. Live specimens have been confiscated by authorities in Alabama, California, Florida, Texas, Virginia, and Washington, where possession of live snakeheads is now illegal (Courtenay and Williams 2004). Although it was captured in waters of the United States as early as 1998, only when it was found established in a 1.8 ha retention pond in Crofton, Maryland (summer of 2002) did it cause considerable public concern. At least 11 more states have specifically prohibited possession of live snakeheads (Courtenay and Williams 2004).

Northern snakehead was thought eradicated from waters of the United States, but it has since appeared in the Potomac River and its tributaries in 2004, and then in Philadelphia and eastern Massachusetts in 2005-2006. Many have been collected (2006 and 2007) in the Potomac Basin, centering around Dogue and Little Hunting Creeks in Virginia and the Anacostia River in Maryland (Fuller and Benson 2007). As such, a thorough understanding of its potential distribution becomes increasingly important.

METHODS

Environmental Data Sources--Numerous environmental data sets in the form of digital raster grids were used to summarize ecological variation across Asia and North America. In this analysis, 15 environmental variables were used for analysis, which summarize aspects of topography (elevation, topographic index, flow accumulation, slope, and aspect from USGS Hydro-1K data set; http://edcdaac.usgs.gov/gtopo30/hydro/), percent tree cover (Hansen et al. 2003), and climatic conditions (annual means of diurnal temperature range; frost days; precipitation; maximum, minimum and mean monthly temperatures; solar radiation; wet days; and vapor pressure; for 1960-1990 from the Intergovernmental Panel on Climate Change Worldwide Climate Data Distribution Centre; http://ipcc-ddc.cru.uea.ac.uk/index.html). All analyses were confined to the region bounded by 24.5988-53.7988˚ N, 66.1417-125.0217˚ W in North America and the native range in East Asia (18.8300-50.6900˚N, 96.1616-145.7416˚E). All environmental data sets were resampled to a pixel resolution of 0.01˚ for analysis.

Occurrence Data Sources–Native-range occurrence data for northern snakeheads were obtained from the Wuhan Institute of Hydrobiology, Beijing Institute of Zoology, Kunming Institute of Zoology, Chinese Academy of Sciences, and scientific literature such as the provincial fish faunas in China (Institute of Hydrobiology Academia Sinica 1976, Editorial Subcommittee of Fishes of Fujian Province 1984, Yang 1987, Wu 1989, Chen et al. 1990, Colloaborative Surveying Team of Fishery Resources in Yangtze Drainage 1990, East China Sea Fisheries

Research Institute et al. 1990, Ding 1994, Zhang 1995), FishNet (http://speciesanalyst.net/fishnet/), and FishBase (http://www.fishbase.org/search.html). Occurrence data for Asian records were assigned geographic coordinates using the Geonames Query web tool (http://gnpswww.nima.mil/geonames/GNS/index.jsp). Points outside the known native range were excluded from the data pool, and duplicate occurrence points were removed. In total, we obtained 171 occurrence points for northern snakeheads from throughout their native distribution.

Occurrence data for northern snakeheads in the United States were obtained from USGS Nonindigenous Aquatic Species database (http://nas.er.usgs.gov), and were georeferenced using the Geographic Names Information System (http://geonames.usgs.gov/gnishome.html). Township-section-range data were georeferenced using a conversion engine developed by the Montana State University Environmental Statistics Group (http://www.esg.montana.edu/gl/trs-data.html). Ambiguous or unspecific localities were excluded from analysis. We obtained 21 unique occurrence points in the United Sates.

Evaluating Environmental Variables--The environmental variables were subjected to a jackknife procedure, which allows exclusion of environmental variables that can lead to spurious overfitting. Hence, for N environmental coverages, N analyses are run using all combinations of N-1 environmental coverages. Then, coverages are evaluated via correlations between inclusion/exclusion of the environmental variables and the average omission error (i.e., predicting absence at sites of known presence). Environmental variables correlated with increased omission error were excluded from further analysis, following Peterson and Cohoon (1999).

Model Building--The native-occurrence data were randomly divided into two subsets to permit model training (80%) and model validation (20%). Of 171 native-range occurrence points, 137 were used for training, and 34 were withheld from modeling and used to test the model predictions.

The details of use of GARP in ecological niche modeling have been presented in numerous publications, such as Anderson et al. (2003), Feria and Peterson (2002), Levine et al. (2004), and Wiley et al. (2003). In essence, GARP attempts to find nonrandom associations between environmental conditions and known occurrences of a species by evolving rules that predict presence or absence of the species. To accomplish this task, GARP uses a subset of the training data to formulate a rule and the rest to test internally the predictive accuracy of the rule. As rules are generated and evolved, the expectation is that the differences between one round of prediction and the next will decrease, converging on a final solution. The investigator can specify this convergence limit (0.01 in this study). Modeling continues until the convergence limit is reached, or a number of iterations specified by the investigator are run (in this study, 1000 times, which was never reached before convergence).

GARP will produce as many models as the investigator specifies. Because of stochastic elements in the process, some of these replicate models will be 'better' (i.e., more predictive) than others. Two criteria are used to evaluate model quality, omission error and commission "error." Omission error occurs when a model fails to predict potential for presence at known occurrence points. Commission "error" is more complex: it is calculated the proportional area predicted to be suitable but includes both true error (inappropriate conditions predicted as suitable), and apparent commission error (species may be present, but site was not sampled); hence, I refer to it as the "commission index." Among a set of models, those that have low omission error rates and that are close to the median commission index appear to offer best predictive ability (Anderson et al. 2003).

As such, I used the best-subsets option of desktop GARP to select the 10 best models from among the models generated by the algorithm. Here, I generated 200 initial to derive 20 models under an absolute omission error threshold of 0%, from which I selected the 10 with a commission index closest to the median. The 10 models were added together, pixel by pixel, to create a final prediction.

Model Evaluation--Model predictions were then evaluated using the validation data via Receiver Operating Characteristic (ROC) analysis, a method designed to evaluate the specificity (absence of commission error) and sensitivity (absence of omission error) of a diagnostic test (Zweig and Campbell 1993, Fielding and Bell 1997). It has been applied to testing the statistical accuracy of GARP results by Iguchi et al (2004), Wiley et al (2003), and Chen et al (2007), and more generally in niche modeling by Elith et al (2006). The area under the curve (AUC) in a ROC analysis is a measure of predictive accuracy for the model set as a whole: if the AUC = 0.50, then the best-model set is performing no better than random, but if the AUC is significantly higher than 0.5 (as judged by a z-test), then the result is significant. The higher the value of AUC, the better the model set, and a perfect prediction would have AUC = 1.0 (Hanley and McNeil 1982). The maximal value of the AUC score is achieved when all of the validation data points occur in pixels where all 10 of the best models predict presence, but it is influenced by the relative extent of the area predicted "present" compared to the total landscape examined (Wiley et al. 2003).

RESULTS

The jackknife process indicated that all of the environmental variables contributed positively to model quality, so I used all 15 variables to build final models. The sum of the 10 best models could then be visualized across both the native-range and the conterminous United States (Fig.1). Model validations for northern snakehead were highly significant over the native landscape and that of the United States (Table 2). Basically 30 of 34 independent occurrence points in the native landscape were predicted by all 10 best models, yielding an AUC = 0.7992, and 17 of 21 US points were predicted by all 10 best models, yielding an AUC = 0.7837.

The native-range niche models over the conterminous United States shows that the entire Northeast and most of Southeast and Midwest of the United States present suitable conditions for northern snakehead as gauged by the native-range ecological conditions. Oklahoma and northeastern Texas are also habitable for northern snakehead, and more distantly, Washington, northern Idaho, and northeastern California are also predicted to be habitable.

DISCUSSION

The predicted distribution for northern snakehead in the conterminous United States is consistent with its native latitudinal range (24 -53° N) and temperature tolerance (0-30 °C), and indicates broad invasive potential. Northern snakeheads have been imported for sale in live-food fish markets, and have been the most widely available snakehead in the United States until 2002. They have been found alive in fish markets in New York, Houston, St. Louis, Pembroke Pines (Florida), and Orlando before 2002 (Courtenay and Williams 2004). The native-range niche model predicted all of these areas as highly suitable for northern snakeheads, so the likelihood of its becoming established is high if this fish is released in these regions.

Northern California, Oregon, Washington, and Idaho are also predicted to be able to sustain populations of northern snakehead. A shipment of live northern snakeheads bound for a seafood distributor was confiscated in Seattle in 2001 (Courtenay and Williams 2004), so it is quite possible for northern snakehead to establish populations in this area if no further preventive measurements are implemented. A single northern snakehead was captured in 1997 in Spiritwood Lake, in the San Bernardino Mountains, California (Courtenay and Williams 2004), another site predicted as suitable by the independent native-range niche models.

One interesting exception is southern Florida, which is not predicted as habitable for northern snakeheads. This prediction is consistent with observations of the native range, in which the species rarely reaches the upper Beijiang River, in Guangdong Province of China (the southernmost extreme of the distribution of northern snakehead): indeed, the one known specimen from there is likely the result of an introduction (Pearl River Fisheries Research Institute of Chinese Academy of Fisheries Science et al. 1991).

Northern snakehead has been captured in several states in the United States, and modeling shows that it has the potential to spread further. Northern snakeheads were collected from Dogue Creek (in 2004-2006), Massey Creek (in 2004-2005), and Little Hunting Creek (a tributary of the Potomac, in 2004-2006), in Fairfax County, Virginia; and many more have been collected in 2006 and 2007 in the Potomac Basin centering around Dogue and Little Hunting creeks in Virginia and from the Anacostia River in Maryland (Fuller and Benson 2007). There is no doubt that this species is well established in the Potomac River and several of its tributaries in Virginia and Maryland, and according to the predictive models it can spread throughout the Potomac River Basin which includes District of Columbia, Maryland, Pennsylvania, Virginia, and West Virginia. These areas are at the highest risk of invasion. Appropriate actions should be carried out to prevent this species from spreading to adjacent drainages which are predicted suitable for northern snakehead too.

In July 2004, several individuals were captured from a pond in Franklin Delano Roosevelt Park in Philadelphia, Pennsylvania. In the following year, young snakeheads were captured again in the park pond (Fuller and Benson 2007). Although status of established populations is still unknown, all these areas and adjacent are predicted suitable by the native-range models. Therefore, the lower Schuylkill River and Delaware River in Pennsylvania are also at high risk of invasion.

A few individual northern snakeheads have been collected in Massachusetts, New York, Illinois, and North Carolina (Fuller and Benson 2007), which put the Merrimac River basin in Massachusetts and New Hampshire, the coastal drainage in Lower Hudson, the Illinois River basin, and the Santee River basin in North Carolina and South Carolina under direct invasion risk because all these areas are predicted habitable for northern snakehead by the native-range ecological niche models. Specimens have also been captured in California and Florida (Fuller and Benson 2007), areas predicted as relatively unsuitable, indicating it is unlikely to establish populations, and the invasion risk should be low.

The highly significant prediction for the present known distribution of northern snakehead in the U.S. suggested that ecological niche models can be an essential proactive tool for combating species' invasions. First, niche models are based on environmental variables that are available openly, so there is no need to collect the environmental data. Secondly, the occurrence data can also be extracted directly from museum collection and data bases (often online) or from published literature for the native range, so the models do not depend on dense invaded-range occurrence data. Lastly, many of the niche modeling software platforms are available for free (e.g., desktop GARP on http://nhm.ku.edu/desktopgarp/, MARS on http://www.salford-systems.com/mars.php, MaxEnt on http://www.cs.princeton.edu/~schapire/maxent/) and can run on personal desktop computers. All these points ensure that potential distribution analysis and risk assessment can be carried out inexpensively, even before actual introductions happens, or early in the invasion process.

LITERATURE CITED

Anderson, R. P., D. Lew, and A. T. Peterson. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. Ecological Modelling **162**:211-232.

Bykhovskaya-Pavlovskaya, I. E., A. V. Gusev, M. N. Dubinina, N. A. Izyumova, T. S. Smirnova, I. L. Sokolovskaya, G. G. Shtein, S. S. Shil'man, and V. M. Epshtein. 1964. Key to Parasites of Freshwater Fish of the USSR (English translation of Opredelitel' parazitov presnovodnykh ryb SSSR). Israel Program for Scientific Translation, Jerusalem.

Chen, M., H. Tong, T. Yu, and Z. Diao. 1990. The Fish Resources of Qiantang River. Shanghai Scientific and Technological Literature Publishing House, Shanghai, China.

Chen, P., E. O. Wiley, and K. M. McNyset. 2007. Ecological niche modeling as a predictive tool: silver and bighead carps in North America. Biological Invasions **9**:43-51.

Colloaborative Surveying Team of Fishery Resources in Yangtze Drainage. 1990. Fishery Resources of Yangtze Drainage. Ocean Publishing House, Beijing.

Courtenay, W. R., Jr., and J. R. Stauffer, Jr. 1984. Distribution, Biology, and Management of Exotic Fishes. Johns Hopkins University Press, Baltimore, Maryland.

Courtenay, W. R., Jr., and J. D. Williams. 2004. Snakeheads (Pisces, Channidae)--a biological synopsis and risk assessment. U. S. Department of the Interior, USGS.

Ding, R., editor. 1994. The Fishes of Sichuan, China (in Chinese). Sichuan Publishing House of Science and Technology, Chengdu, China.

Dukravets, G. M. 1992. The Amur snakehead, *Channa argus warpachowskii*, in the Talas and Chu river drainages. Journal of Ichthyology **31**:147-151.

Dukravets, G. M., and A. I. Machulin. 1978. The morphology and ecology of the Amur snakehead, *Ophiocephalus argus warpachowskii*, acclimatized in the Syr Dar'ya basin. Journal of Ichthyology **18**:203-208.

East China Sea Fisheries Research Institute, Chinese Academy of Fisheries Science, and Shanghai Fisheries Research Institute, editors. 1990. The Fishes of Shanghai Area. Shanghai Scientific and Technical Publishers, Shanghai.

Editorial Subcommittee of Fishes of Fujian Province, editor. 1984. The Fishes of Fujian Province (Part I). Fujian Science and Technology Press, Fujian China.

Elith, J., C. H. Graham, R. P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. M. Overton, A. T. Peterson, S. J. Phillips, K. Richardson, R. Scachetti-Pereira, R. E. Schapire, J. Soberon, S. Williams, M. S. Wisz, and N. E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography **29**:129-151.

Feria, T. P., and A. T. Peterson. 2002. Using point occurrence data and inferential algorithms to predict local communities of birds. Diversity and Distributions **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. Environmental Conservation **24**:38-49.

Fuller, P. F., and A. J. Benson. 2007. *Channa argus* USGS Nonindigenous Aquatic Species Database. *in* USGS Nonindigenous Aquatic Species Database, Gainesville, FL, retrieved from http://nas.er.usgs.gov/queries/FactSheet.asp?speciesID=2265 on 05 December 2007, Gainesville, FL.

Guseva, L. N. 1990. Food and feeding rations of the Amur snakehead, *Channa argus warpachowskii*, in water bodies in the lower reaches of the Amu Darya. Journal of Ichthyology **30**:11-21.

Guseva, L. N., and I. M. Zholdasova. 1986. Morphoecological characteristics of the snakehead (*Ophiocephalus argus warpachowskii* Berg) as an introduced species in water bodies in the lower delta of the Amy Darya. Pages 98-134 *in* Biological Resources of the Aral Region. FAN, Tashkent.

Hanley, J. A., and B. J. McNeil. 1982. The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve. Radiology **143**:29-36.

Hansen, C. M., R. S. DeFries, J. R. G. Townshend, M. Carroll, C. Dimiceli, and R. A. Sohlberg. 2003. Global Percent Tree Cover at a Spatial Resolution of 500 Meters: First Results of the MODIS Vegetation Continuous Fields Algorithm. Available from http://glcf.umiacs.umd.edu/data/treecover/index.shtml.

Iguchi, K., K. Matsuura, K. McNyset, A. T. Peterson, R. Scachetti-Pereira, K. A. Powers, D. A. Vieglais, E. O. Wiley, and T. Yodo. 2004. Predicting invasions of North American basses in Japan using native range data and a genetic algorithm. Transactions of the American Fisheries Society **133**:845-854.

Institute of Hydrobiology Academia Sinica. 1976. Fishes of the Yangtze River. Science Press, Beijing.

Levine, R. S., A. T. Peterson, and M. Q. Benedict. 2004. Geographic and ecologic distributions of the *Anopheles gambiae* complex predicted using a genetic algorithm. American Journal of Tropical Medicine & Hygiene **70**:105-109.

Pearl River Fisheries Research Institute of Chinese Academy of Fisheries Science, South China Normal University, Ji-Nan University, Zhanjiang Fisheries College, and Shanghai Fisheries University. 1991. The Freshwater Fishes of Guangdong Province. Guangdong Science and Technology Press, Guangzhou, China.

Peterson, A. T., and K. C. Cohoon. 1999. Sensitivity of distributional prediction algorithms to geographic data completeness. Ecological Modelling **117**:159-164.

Wiley, E. O., K. M. McNyset, A. T. Peterson, C. R. Robins, and A. M. Stewart. 2003. Niche modeling and geographic range predictions in the marine environment using a machine-learning algorithm. Oceanography **16**:120-127.

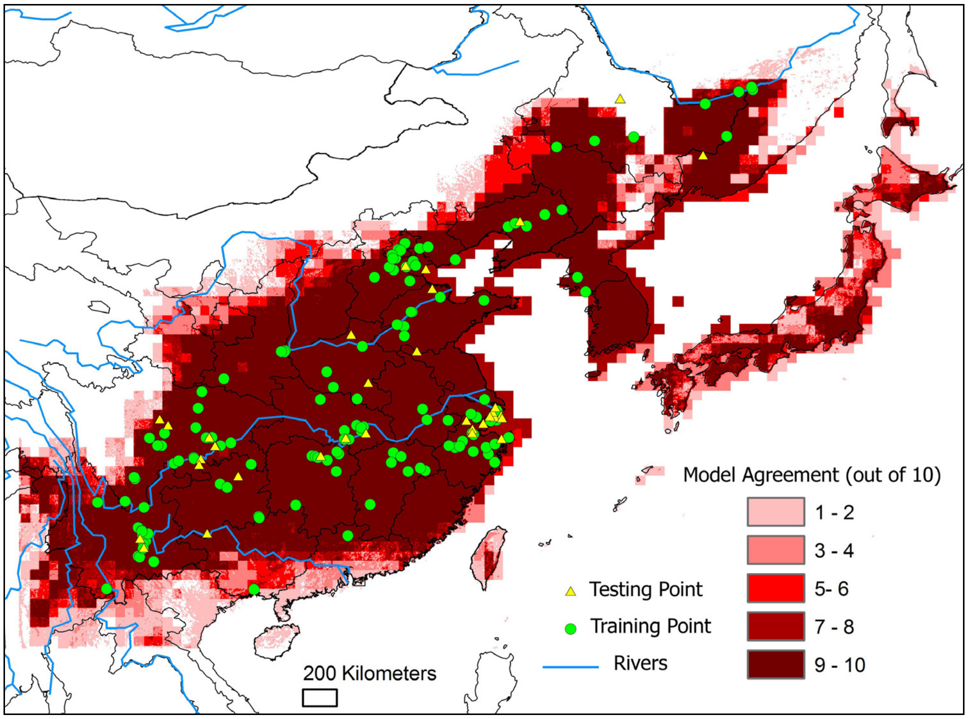
Williams, J. D., and G. K. Meffe. 2000. Invasive nonindigenous species: a major threat to freshwater biodiversity. Pages 67-73 *in* R. A. Abell, editor. Freshwater Ecoregions of North America: A Conservation Assessment. Island Press, Washington, D.C.

Wu, L. 1989. The Fishes of Guizhou Province. The People's Publishing House of Guizhou, Guiyang, China.

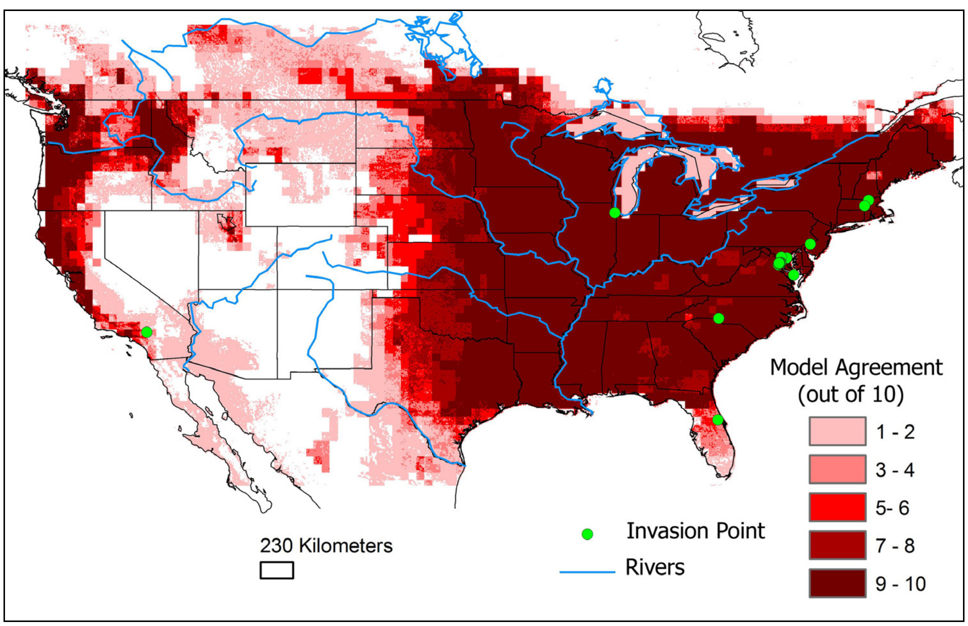
Yang, G., editor. 1987. The Fishes of Hubei Province, China. Hubei Science and Technology Press, Wuhan, Hubei.

Zhang, J., editor. 1995. The Fishes of Heilongjiang Province. Heilongjiang Science and Technology Press, Jilin, China.

Zweig, M. H., and G. Campbell. 1993. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clinical Chemistry **39**:561-577.



A



B

Figure 3-1. A, Niche model predictions over the native landscape, showing probable range; B, Native-range model projected over the conterminous United States, showing the potential invasive range of northern snakehead. Dark red indicates 9-10 of the 10 best models predicting presence, firebrick 7-8, red 5-6, salmon 3-4, and pink 1-2. Green triangles indicate training data used to build models; yellow circles indicate independent validation data.

Table 3-1: Statistics of model building and evaluation over the native landscape and the conterminous United States. AUC—the area under the curve, SE—standard error, *Z*—value associated with receiver-operator curve analysis.

Region	Training points	Testing points	Model subsets 1-2	3-4	5-6	7-8	9-10	AUC	SE	Z
Native	137	34	0	0	0	3	30	0.7992	0.0458	9.8757**
USA		21	0	0	2	0	17	0.7837	0.0596	7.0592**

** p<<0.001

CHAPTER 4

PREDICTING THE POTENTIAL GEOGRAPHIC DISTRIBUTION OF ASIAN SWAMP EEL AND ORIENTAL WEATHER-FISH IN NORTH AMERICA

Chapter Abstract

The Asian swamp eel and the oriental weather-fish are already found in North America. Both are known for their high temperature tolerance, wide native distribution range and their ability to survive for long periods without water. This analysis is to use ecological niche modeling to build the native-range niche models for each species from the environmental variables and species' native occurrence points. The native niche models for both species are highly significant over the native landscape. I then projected the native-range models onto the North American landscape. The niche models predicted the known occurrences of these species significantly better than random expectation ($p<<0.001$), forecasting that Asian swamp eels are able to establish populations in southern United States, Mid Atlantic, all of lower Mississippi River drainage, and the West Coast; and oriental weather-fishes have the potential to cover all of the conterminous United States except the Rocky Mountain and desert areas.

INTRODUCTION

The Asian swamp eel and the oriental weather-fish are already found in North America. Both are known for their high temperature tolerance, wide native distribution range, and their ability to survive for long periods without water. Although the potential risks of these fishes as invaders are largely unknown, they are likely to negatively impact ecosystems as they disperse and become abundant. This analysis is to use ecological niche modeling to (1) model the ecological niches of the Asian swamp eel and the oriental weather-fish from the East Asian native distribution points; (2) test the accuracy of each niche model within the native range; (3) project the niche model onto North America to assess the potential range of each species; and (4) test the forecast with occurrence data from known introductions.

The Asian swamp eel, *Monopterus albus* (Zuiew 1793), is an eel-like fish that lacks scales, and pectoral and pelvic fins. The dorsal, caudal and anal fins are confluent and reduced to a skin fold. The jaws and palate have rows of villiform teeth. The gill openings are merged into single slit underneath the head (Yang 1987). The Asian swamp eel is native to Asia, from India and Burma to China, Japan, Malaysia, and Indonesia, and probably occurs in Bangladesh. It has been widely introduced and established outside its native range. It occurs in streamlets, canals, and estuaries, and is very common in muddy ponds, swamps, and rice fields. It burrows in moist earth in the dry season, and is able to survive for long periods of drought (Mao and Xu 1991). It is capable of living out of water for days, so long as the skin is kept moist. Asian swamp eels are nocturnal predators, devouring fishes, worms, crustaceans, tadpoles and other small aquatic animals (Chen et al. 1990); they also feed on detritus and algae (Yang 1987). It is a protogynous fish. The young are hatched as females, mature as females, and then pass through a non-sexual stage for about a year before transforming into larger males. Spawning occurs from May to August, largely in June and July. Eggs are laid in a bubble nest in shallow water. The nest is typically not attached to vegetation but can float freely at the surface. One or both parents guard the eggs and young (Yang 1987).

The oriental weather-fish, *Misgurnus anguillicaudatus* (Cantor 1842), is an eel-like small cobitid, native to Asia, from Japan, Northern China to Central China and Burma. This species is recognised by its cylindrical body, five pairs of barbels around its mouth and its single short-based dorsal fin. Individuals are found in rivers, lakes, ponds, or silty substrates in low-gradient, shallow water, often in aquatic macrophyte beds such as swamps and rice paddy fields (Mao and Xu 1991, Tabor et al. 2001). The common name refers to this species reportedly becoming restless during changes in barometric pressure and thus having been cited as a harbinger of storms. They feed mostly on small benthic invertebrates, detritus, and algae; Tabor *et* *al*. (2001) found that cladocerans and chironomids were the most frequently ingested prey items collectively. This species does not forage by sight, but rather, requires chemical stimuli to induce feeding behavior (Watanabe and Hidaka 1983). It grows to a maximum length of 25 cm. Up to 2,250 - 3,000 kg of this species can be harvested per hectare in aquaculture farming field (Mao and Xu 1991). It reaches sexual maturity in 2 yr, spawning occurring from April to September (Mao and Xu 1991). This species can tolerate temperature from 2 ˚C to 30 ˚C. Weather-fish are capable of using their intestine as an additional respiratory organ, enabling them to live in oxygen-depleted waters by swallowing air from water surface and to bury themselves in soft substrates in case of long lasting droughts (Mao and Xu 1991).

METHODS

Environmental Data Sources--This analyses used 15 environmental variables common to both Asia and North America, which include aspects of topography (elevation, topographic index, flow accumulation, slope and aspect from USGS Hydro-1K data set; http://edcdaac.usgs.gov/gtopo30/hydro/), percent tree cover (Hansen et al. 2003), and Climatic conditions (annual means of diurnal temperature range; frost days; precipitation; maximum, minimum and mean monthly temperatures; solar radiation; wet days; and vapor pressure; for 1960-1990 from the

Intergovernmental Panel on Climate Change Worldwide Climate Data Distribution Centre; http://ipcc-ddc.cru.uea.ac.uk/index.html). The analyses were confined to the region bounded by 24.5988–53.7988˚ N, 66.1417–125.0217˚ W in North America, and the native range in Asia by 18.8300–50.6900˚N, 96.1616–145.7416˚E for the oriental weather-fish, and by 10.5976˚S-55.1524˚N, 92.8336–145.4336˚E for the Asian swamp eel. All environmental data sets were resampled to a pixel resolution of 0.01˚ for analysis.

Occurrence Data Sources–Native-range occurrence data for the Asian swamp eel and the oriental weather-fish in Asia were obtained from the Wuhan Institute of Hydrobiology, Beijing Institute of Zoology, Kunming Institute of Zoology, Chinese Academy of Sciences, and scientific literature such as the provincial fish faunas in China, FishNet (http://speciesanalyst.net/fishnet/), and FishBase (http://www.fishbase.org/search.html). Occurrence data for Asian records were assigned geographic coordinates using the Geonames Query web tool (http://gnpswww.nima.mil/geonames/GNS/index.jsp). Points outside the known native range were excluded from the data pool, and duplicate occurrence points were removed. In total, we obtained 313 and 400 unique occurrence points, respectively, for the Asian swamp eel and the oriental weather-fish, from throughout their native distributions available.

Occurrence data for both species in the conterminous United States were obtained from USGS Nonindigenous Aquatic Species database (http://nas.er.usgs.gov). Occurrence localities were georeferenced using USGS Geographic Names Information System (http://geonames.usgs.gov/gnishome.html). Township-section-range data were georeferenced using a conversion engine developed by the Montana State University Environmental Statistics Group (www.esg.montana.edu/gl/trs-data.html). Ambiguous records or unspecific localities were excluded from analysis. We obtained 4 and 40 unique occurrence points, respectively, for the Asian swamp eel and the oriental weather-fish in the conterminous United Sates.

Evaluating Environmental Variables--The environmental variable data set for model building was chosen through a jackknife procedure. This procedure was designed to exclude variables which can lead to spurious overfitting by evaluating correlations between inclusion/exclusion of the environmental variables and the average omission error (i.e., predicting absence at sites of known presence) of 20 replicate models. Omission error is high when the model cannot predict known occurrence points. Environmental variables correlated with increased omission error were excluded from further analysis, following Peterson and Cohoon (1999).

Model Building--The native-occurrence data for each species were randomly divided into two subsets to permit model training and model validation. The training data set consisted of 260 occurrence points for the Asian swamp eel and 320 occurrence points for the oriental weather-fish. The validation data set, consisting of 53 native-range occurrence points for the Asian swamp eel and 80 occurrence points for the oriental weather-fish, was withheld entirely from the modeling process and used to test models generated by the algorithm.

The details of use of GARP in ecological niche modeling have been presented in numerous publications, such as Anderson *et al*. (2003), Feria and Peterson (2002), Levine *et al*. (2004), Peterson (2001), and Wiley *et al*. (2003). In essence, GARP attempts to find nonrandom associations between environmental conditions and known occurrences of a species by evolving rules that predict presence or absence of the species. To accomplish this task, GARP uses a subset of the training data to formulate a rule and the rest to internally test the predictive accuracy of the rule. In this study, 80% of the training data were used in each model iteration to evolve rules and 20% were used to test rules. As rules are generated and evolved, the expectation is that the differences between one round of prediction and the next will decrease, converging on a final solution. The investigator can specify this convergence limit (0.01 in this study). Modeling continues until the convergence limit is reached, or a number of iterations specified by the investigator are run (in this study, 1000 times, which was never reached before convergence).

GARP will produce as many models as the investigator specifies. Because of stochastic elements in the process, some of these replicate models will be ‘better’ (i.e., more predictive) than others. Two criteria are used to evaluate model quality, omission error and commission “error.” Omission error occurs when a model fails to predict potential for presence at known occurrence points. Commission “error” is more complex: it is calculated the proportional area predicted to be suitable but includes both true error (inappropriate conditions predicted as suitable), and apparent commission error (species may be present, but site was not sampled); hence, I refer to it as the “commission index.” Among a set of models, those that have low omission error rates and that are close to the median commission index appear to offer best predictive ability (Anderson et al. 2003).

As such, I used the best-subsets option of desktop GARP to select the 10 best models from among the models generated by the algorithm. Here, I generated 200 initial to derive 20 models under an absolute omission error threshold of 0%, from which I selected the 10 with a commission index closest to the median. The 10 models were added together, pixel by pixel, to create a final prediction.

Model Evaluation--The 10 best model set for each species was evaluated using the validation data with a Receiver Operating Characteristic (ROC) analysis, a method designed to evaluate the specificity (absence of commission error) and sensitivity (absence of omission error) of a diagnostic test (Zweig and Campbell 1993, Fielding and Bell 1997). It has been applied to testing the statistical accuracy of GARP results by Iguchi et al (2004), Wiley et al (2003), and Chen et al (2007), and more generally in niche modeling by Elith et al (2006). The area under the curve (AUC) in a ROC analysis is a measure of predictive accuracy for the model set as a whole. The area under the curve (AUC) in a ROC analysis is a measure of predictive accuracy for the model set as a whole: if the AUC = 0.50, then the best-model set is performing no better than random, but if the AUC is significantly higher than 0.5 (as judged by a z-test), then the result is significant. The higher the value of AUC, the better the model set, and a perfect prediction would have AUC = 1.0 (Hanley and McNeil 1982). The maximal value of the AUC score is achieved when all of the validation data points occur in pixels where all 10 of the best models predict presence, but it is influenced by the relative extent of the area predicted “present” compared to the total landscape examined (Wiley et al. 2003). The accuracy of the 10 best model set for each species was also calculated as the percentage of validation points within10 best models. When all validation points are successfully predicted by all 10 best models, the accuracy is 100%.

RESULTS

The jackknife processes were run initially for all 15 environmental variables, and excluded no variables for either the Asian swamp eel or the oriental weather-fish. All 15 variables for each species were thus used to build the 10 best-subset models. The 10 best-subset models were projected onto the Asian (Figures 4-1 B and 4-2 B) and North American landscapes (Figures 4-1 A and 4-2 A). The niche models for both species are highly significant over both the native landscape and the conterminous United States (see Table 4-1).

The Asian swamp eel models predict the potential establishment of populations in southern United States, Mid Atlantic, all of lower Mississippi River drainage, and the West Coast (Figure 4-1 A). The niche models in the native landscape are significantly predicting the species distribution. The accuracy based on the 53 independent points is 96.04%, and ROC analysis gets an AUC scores 0.7682 (Eng 2006). Of the four known occurrence points in the conterminous United States, 3 are correctly predicted by all 10 best models, and one by 3 of 10 best models (Table 4-1).

The native-range models forecast a larger potential distribution for the oriental weather-fish in North America, with a predicted range covering all of the conterminous United States except the Rocky Mountain and desert areas (Figure 4-2 A). The niche models on the native landscape have an accuracy 95.88%, and an AUC score of 0.7459 ($p<<0.001$) based on the 80 independent test points. The oriental weather-fish has become established in southern states, Great Lakes states, and the Pacific states. Thirty-six of the 40 occurrence points on North America were predicted by ≥9 of the 10 best-subset models, yielding an AUC score of 0.7466 ($p<<0.001$). The predictive accuracy was 96.0% (Table 4-1).

DISCUSSION

Generally speaking, species with a successful introduction history are more likely to invade new areas, increasing their range. The Asian swamp eel is a prime example. In the early 1900s, it was introduced to Nara Prefecture in Japan from Korea (Matsumoto et al. 1998). During the late nineteenth century Chin Dynasty, it was introduced to the region near Hami, Xingjiang Autonomous District, northwestern China, by the provincial army (Wang et al. 1994), and to Oahu, Hawaii, by Asian immigrants as a food fish before 1900 (Maciolek 1984, Devick 1991). Brock (1960) stated that it was already established in Hawaii prior to 1900. The small population in the northeastern Australia is believed to be introduced (Merrick and Schmida 1984). The Asian swamp eel is an ecological generalist, widely distributed from China, Japan, to India, to Malaysia and Indonesia, and is often transported as a live market food fish and aquarium fish. The predicted ranges in the conterminous United States are consistent with its wide native distribution.

The Asian swamp eel was probably introduced into North America as an aquarium release. In Georgia, adults were first collected near Atlanta in 1996, although they were likely present since 1990 or before (Starnes et al. 1998). In 1997, two populations were discovered in Florida, one in Manatee County, near Tampa, and the other in Miami-Dade and Broward counties in North Miami. In 1999, a population was found in the Homestead area of Miami-Dade county, near Everglades National Park (Collins et al. 2002, Nico 2006). It also is known from three spring-fed impoundments (Chattahoochee River drainage) at the Chattahoochee Nature Center in Roswell, Fulton County, and the Chattahoochee River National Recreation Area, Gwinnett County, Georgia (Nico 2006). All these localities are predicted as habitable by the native-range models for the Asian swamp eel, and there is a good reason to worry that the Asian swamp eel in both Georgia and Florida will spread to adjacent water bodies and theoretically to all of the potential range unless efficient management practices are implemented.

The potential risk of the Asian swamp eel is largely unknown. However, it is a generalized predator and is likely to impact negatively prey population sizes and the availability of those prey species to larger native fishes, turtles, frogs and wading birds. It may also play a role in altering the habitat beneath ponds and marshy regions where they burrow nests to wait out dry seasons.

Considering its high oxygen and temperature tolerance, wide distribution and polyphagy in native habitats, the larger predicted distribution ranges of the oriental weather-fish in North America is not surprising. This fish has been successfully introduced into several parts of the world for aquaculture purposes, as a bait fish, and as an aquarium fish (Welcomme 1988). It was introduced into Hawaii as food fish before 1900 (Maciolek 1984, Devick 1991), and its use as bait for fish apparently prompted its spread in Hawaii (Brock 1960). This species now occupies streams in Kauai, Maui, and Oahu (Maciolek 1984, Devick 1991).

In the 1930s, this species was introduced into U.S. waters from a local goldfish farm escape in California. In 1968, it was collected from a three-mile reach of the Westminster flood control channel in Orange County, California (St. Amant and Hoover 1969). Additional established populations were discovered upstream from the original collection sites in 1977, and in the adjacent Bolsa Chica Channel in 1979 (Shapovalov et al. 1981). It has also been recorded from the following habitats: Huntington Beach, Orange County, California (Courtenay et al. 1986); the Little Manatee River drainage, Florida (since about 1988, Nico & Fuller, 2006); the Peace drainage, Florida (Nico and Fuller 2006); the Harton Davis Canal, in an irrigation ditch at Eagle State Park, in the Boise River system, Ada County, Idaho (since the middle of 1980s, Idaho Fish and Game, 1990) ; the Clackamas River, Oregon (since the mid 1980s) and the Malheur, Owyhee, and Snake River systems in 1995 (Logan et al. 1996); Burlington Bottoms near the Multnomah Channel of the Columbia River (in 1997), and farther downstream from Scappoose Bay on the Columbia River (in 1994); Washington Lake in Seattle, Washington (since the mid-1990s, Tabor *et al*., 2001); Tulalip Creek, near Marysville, Washington (Nico and Fuller 2006); the North Shore Channel, Cook County, Lake Michigan drainage, Illinois (since 1987), and the Chicago Sanitary and Ship Canal (in 1994; Laird & Page, 1996; Page & Laird, 1993); headwaters of the Shiawassee River, in Oakland and Genesee counties, Michigan (in 1958 and 1959 where it was considered to be established, Schultz, 1960); downstream of a fish farm near Lacombe, Louisiana (Nico and Fuller 2006); and a tributary of Coffee Creek, near the Hiwassee River, Polk County, Tennessee (in April 1995, Nico & Fuller, 2006). All these localities and adjacent areas are predicted by the native-range niche models. If no efficient management practices are implemented, there is a high likelihood that the oriental weather-fish will spread further and become established all over the entire predicted areas of the conterminous United States.

The risks posed by the oriental weather-fish is unknown. Page and Laird (1993) believed that if it becomes more abundant and spreads, it will reduce populations of aquatic insects important as food to native fishes. Maciolek (1984) categorized this species along with several other introduced fishes as species having an intermediate impact on Hawaiian streams after investigating their preferred habitat and diet and their numbers.

LITERATURE CITED

Anderson, R. P., D. Lew, and A. T. Peterson. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. Ecological Modelling **162**:211-232.

Brock, V. E. 1960. The introduction of aquatic animals into Hawaiian waters. Internationale Revue der Gesamten Hydrobiologie **45**:463-480.

Chen, M., H. Tong, T. Yu, and Z. Diao. 1990. The Fish Resources of Qiantang River. Shanghai Scientific and Technological Literature Publishing House, Shanghai, China.

Chen, P., E. O. Wiley, and K. M. McNyset. 2007. Ecological niche modeling as a predictive tool: silver and bighead carps in North America. Biological Invasions **9**:43-51.

Collins, T. M., J. C. Trexler, L. G. Nico, and T. A. Rawlings. 2002. Genetic diversity in a morphologically conservative invasive taxon: multiple introductions of swamp eels to the southeastern United States. Conservation Biology

16:1024-1035.

Courtenay, W. R., Jr., D. A. Hensley, J. N. Taylor, and J. A. McCann. 1986. Distribution of exotic fishes in North America. Pages 675-698 *in* C. H. Hocutt and E. O. Wiley, editors. The Zoogeography of North American Freshwater Fishes. John Wiley and Sons, New York, NY.

Devick, W. S. 1991. Patterns of introductions of aquatic organisms to Hawaiian freshwater habitats. Pages 189-213 *in* New directions in research, management and conservation of Hawaiian freshwater stream ecosystems, Proceedings of the 1990 symposium on freshwater stream biology and fisheries management. Division of Aquatic Resources, Hawaii Department of Land and Natural Resources, Honolulu, Hawaii.

Elith, J., C. H. Graham, R. P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. M. Overton, A. T. Peterson, S. J. Phillips, K. Richardson, R. Scachetti-Pereira, R. E. Schapire, J. Soberon, S. Williams, M. S. Wisz, and N. E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography **29**:129-151.

Eng, J. 2006. ROC analysis: web-based calculator for ROC curves. [updated 2006

May 17]. Available from: http://www.jrocfit.org Johns Hopkins University Baltimore.

Feria, T. P., and A. T. Peterson. 2002. Using point occurrence data and inferential algorithms to predict local communities of birds. Diversity and Distributions **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. Environmental Conservation **24**:38-49.

Hanley, J. A., and B. J. McNeil. 1982. The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve. Radiology **143**:29-36.

Hansen, C. M., R. S. DeFries, J. R. G. Townshend, M. Carroll, C. Dimiceli, and R. A. Sohlberg. 2003. Global Percent Tree Cover at a Spatial Resolution of 500 Meters: First Results of the MODIS Vegetation Continuous Fields Algorithm. Available from http://glcf.umiacs.umd.edu/data/treecover/index.shtml.

Idaho Fish and Game. 1990. Appendix I. A List of Idaho Fishes and Their Distribution by Drainage. Fisheries Management Plan 1991-1995. Idaho Fish and Game.

Iguchi, K., K. Matsuura, K. McNyset, A. T. Peterson, R. Scachetti-Pereira, K. A. Powers, D. A. Vieglais, E. O. Wiley, and T. Yodo. 2004. Predicting invasions of

North American basses in Japan using native range data and a genetic algorithm. Transactions of the American Fisheries Society **133**:845-854.

Levine, R. S., A. T. Peterson, and M. Q. Benedict. 2004. Geographic and ecologic distributions of the *Anopheles gambiae* complex predicted using a genetic algorithm. American Journal of Tropical Medicine & Hygiene **70**:105-109.

Logan, D., E. L. Bibles, and D. F. Markle. 1996. Recent collections of continental exotic aquarium fishes in Oregon and thermal tolerance of Misgurnus anguillicaudatus and Piaractus brachypomus. California Fish and Game **82**:66-80.

Maciolek, J. A. 1984. Exotic fishes in Hawaii and other islands of Oceania. Pages 131-161 *in* W. R. Courtenay, Jr. and J. R. Stauffer, Jr., editors. Distribution, Biology, and Management of Exotic Fishes. The Johns Hopkins University Press, Baltimore, MD.

Mao, J., and S. Xu, editors. 1991. The Fauna of Zhejiang Province: Freshwater Fishes. Zhejiang Publishing House of Science and Technology, Hangzhou, China.

Matsumoto, S., N. Nagai, S. Imanishi, H. Hasuike, and K. Masonori. 1998. An expanding distribution of a swamp eel *Monopterus albus* (Teleostei: Synbranchidae) in Nara Prefecture. Japanese Journal of Ecology **48**:107-116 (in Japanese with English abstract).

Merrick, J. R., and G. E. Schmida. 1984. Australian Freshwater Fishes: Biology and

Management. Griffin Press, Netley, South Australia.

Nico, L. 2006. *Monopterus albus*. *in* USGS Nonindigenous Aquatic Species Database, Gainesville, FL. Retrived from<http://nas.er.usgs.gov/queries/FactSheet.asp?speciesID=974> on 08 September 2006.

Nico, L., and P. Fuller. 2006. *Misgurnus anguillicaudatus*. *in* USGS Nonindigenous Aquatic Species Database, Gainesville, FL. Retrived from <http://nas.er.usgs.gov/queries/FactSheet.asp?speciesID=498> on September 19, 2006

Page, L. M., and C. A. Laird. 1993. The Identification of the Nonnative Fishes Inhabiting Illinois Waters. Center for Biodiversity, Illinois Natural History Survey, Champaign, IL.

Peterson, A. T. 2001. Predicting species' geographic distributions based on ecological niche modeling. Condor:599-605.

Peterson, A. T., and K. C. Cohoon. 1999. Sensitivity of distributional prediction algorithms to geographic data completeness. Ecological Modelling **117**:159-164.

Schultz, E. E. 1960. Establishment and early dispersal of a loach, *Misgurnus anguillicaudatus* (Cantor), in Michigan. Transactions of the American Fisheries Society **89**:376-377.

Shapovalov, L., A. J. Cordone, and W. A. Dill. 1981. A list of freshwater and anadromous fishes of California. California Fish and Game **67**:4-38.

St. Amant, J. A., and F. G. Hoover. 1969. Addition of *Misgurnus anguillicaudatus* (Cantor) to the California fauna. California Fish and Game **57**:330-331.

Starnes, W. C., R. T. Bryant, and G. C. Greer. 1998. Perilous experiment: the Asian rice eel in Georgia. Georgia Wildlife: Urban Wildlife (Natural Georgia Series) **7**:60–70.

Tabor, R. A., E. Warner, and S. Hager. 2001. An oriental weatherfish (*Misgurnus anguillicaudatus*) population established in Washington State. Northwest Science **75**:72-76.

Wang, X., M. Wu, Y. Ma, W. Liu, and W. Shi. 1994. Analysis of karyotype of Monopterus albus in Hami. Journal of Fisheries of China **18**:157-159 (in Chinese).

Watanabe, K., and T. Hidaka. 1983. Feeding behavior of the Japanese loach, *Misgurnus anguillicaudatus* (Cobitididae). Journal of Ethology **1**:86-90.

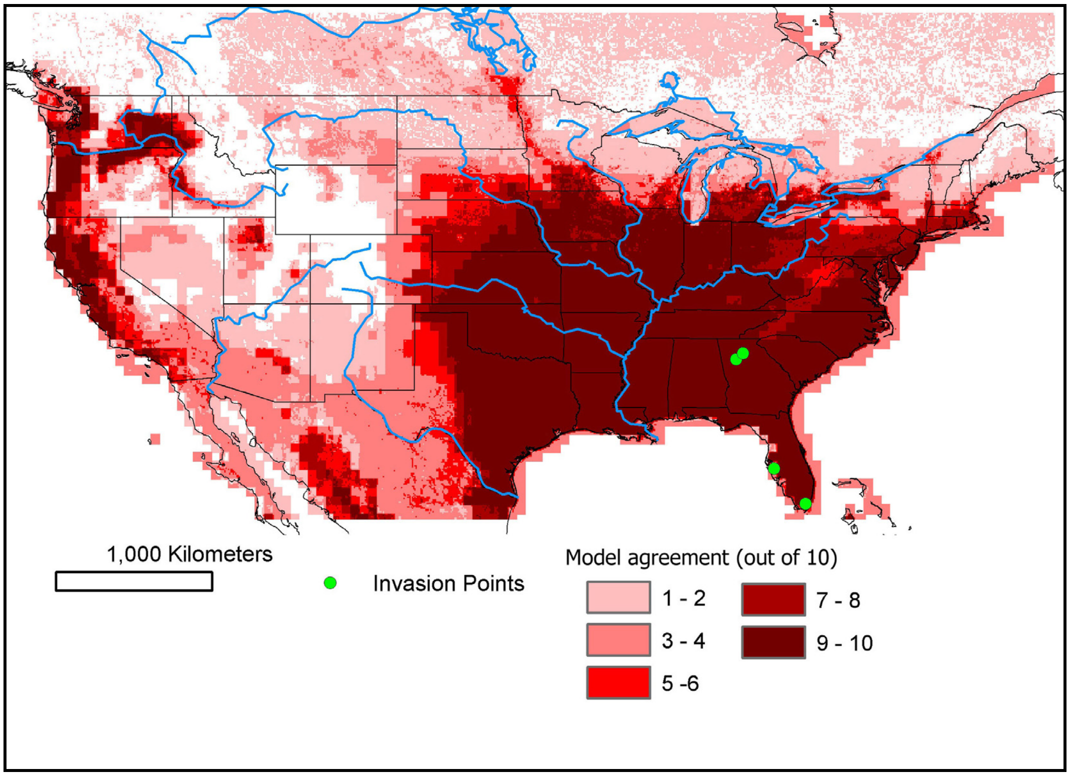
Welcomme, R. L., editor. 1988. International Introductions of Inland Aquatic Species. FAO Fisheries Technical Paper 294. Food and Agriculture Organization of the United Nations (FAO), Rome, Italy.

Wiley, E. O., K. M. McNyset, A. T. Peterson, C. R. Robins, and A. M. Stewart. 2003.

Niche modeling and geographic range predictions in the marine environment using a machine-learning algorithm. Oceanography **16**:120-127.

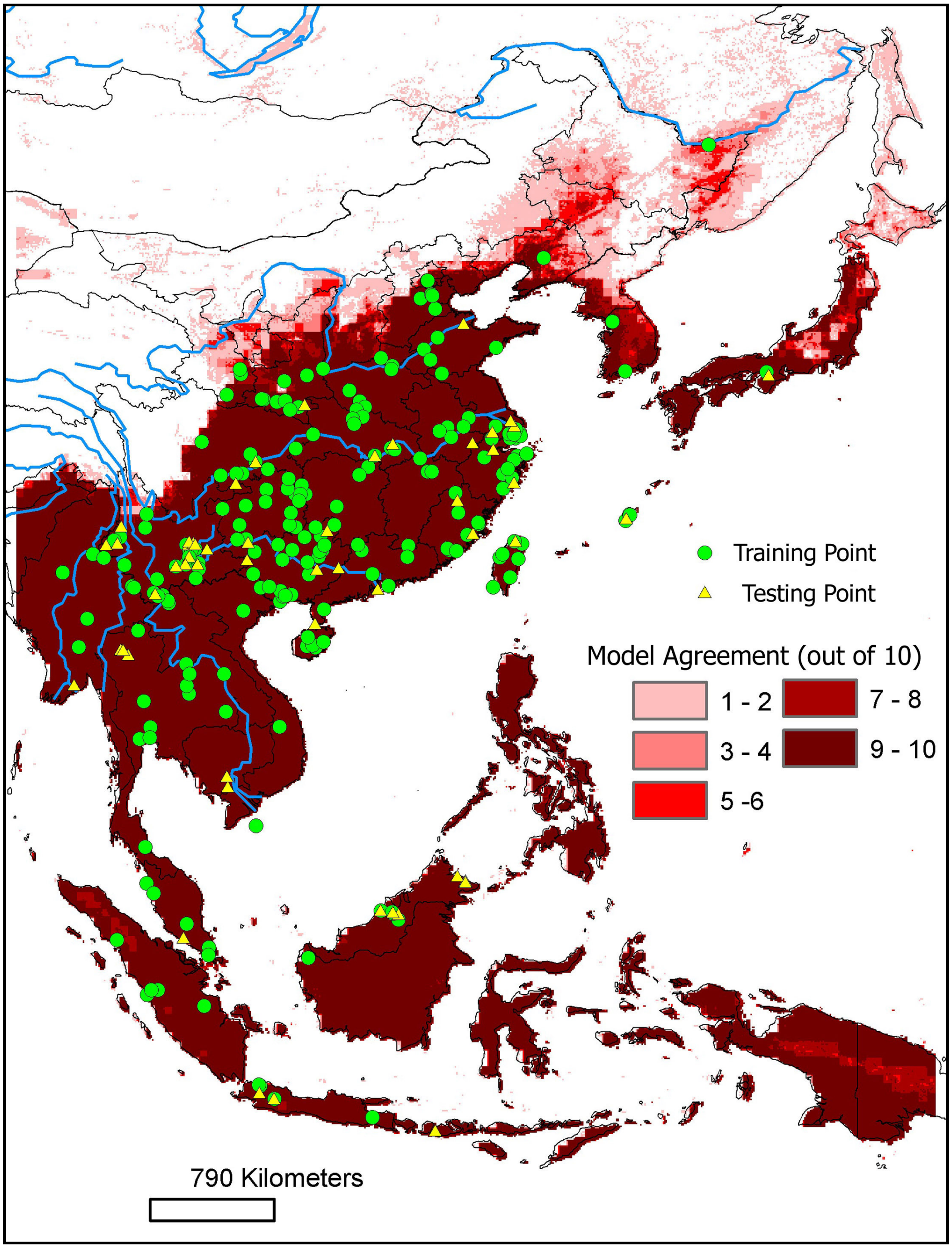
Yang, G., editor. 1987. The Fishes of Hubei Province, China. Hubei Science and Technology Press, Wuhan, Hubei.

Zweig, M. H., and G. Campbell. 1993. Reciver-Operating Characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clinical Chemistry **39**:561-577.



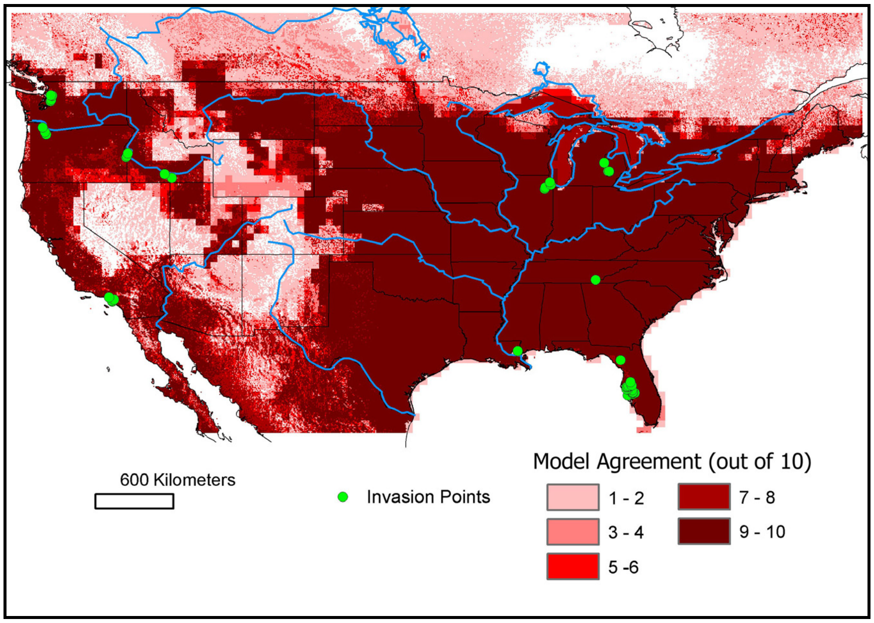
A

Figure 4-1 (continued)

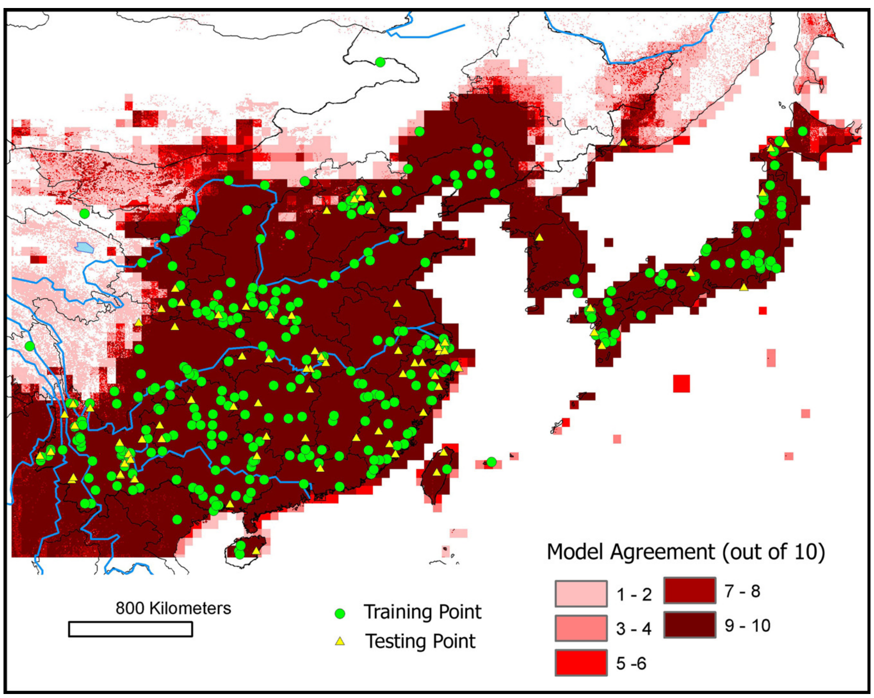


B

Figure 4-1 A, Native-range model of Asian swamp eel projected over the conterminous United States, showing the potential invasive range of northern snakehead; B, Niche model predictions of Asian swamp eel over the native landscape, showing probable range. Dark red indicates 9-10 of the 10 best models predicting presence, firebrick 7-8, red 5-6, salmon 3-4, and pink 1-2.



A



B

Figure 4-2 A, Native-range model of oriental weather-fish projected over the conterminous United States, showing the potential invasive range of northern snakehead; B, Niche model predictions of oriental weather-fish over the native landscape, showing probable range. Dark red indicates 9-10 of the 10 best models predicting presence, firebrick 7-8, red 5-6, salmon 3-4, and pink 1-2.

Table 4-1: Statistics of model building and evaluation over the native landscape and the conterminous USA. AUC—the area under the curve, SE—standard error, Z—value associated with receiver-operator curve analysis.

Species	Landscape	Training points	Testing points	Model subsets					Accuracy	AUC	SE	Z
				1-2	3-4	5-6	7-8	9-10				
Misgurnus	Native	320	80	0	0	1	1	76	95.88%	0.7459	0.0318	15.22**
anguillicaudatus	USA		40	0	1	0	3	36	96.00%	0.7466	0.0450	10.04**
Monopeterus	Native	260	53	0	0	0	0	51	96.04%	0.7682	0.0382	13.154**
albus	USA		4	0	1	0	0	3				

** p<<0.001

CHAPTER 5

PREDICTING THE POTENTIAL GEOGRAPHIC DISTRIBUTION OF 25 ASIATIC INVASIVE FISHES IN NORTH AMERICA

Chapter Abstract

With the constantly increasing trade and travel between United States and Asia, more Asiatic fishes are likely to enter US via the aquarium trade, fish farms, or live fish markets. 25 species selected for this analysis have shown up in US markets or are being considered for importation. I used ecological niche modeling to build niche models for each species and project them to the conterminous United States to forecast the potential distribution ranges and assess their potential invasion risk. Species for which there were independent validation data, model validations were highly significant ($p<<0.001$). The predicting results show that *Myxocyprinus asiaticus*, *Channa maculata*, *Sinilabeo decoru*, and *Cirrhinus molitorella* have a very limited invasive potential in North America, while *Abbottina rivularis*, *Hemiculter leucisculus*, *Hemibarbus labeo*, *Hemibarbus maculatus*, *Plagiognathops microlepis*, and *Pseudorasbora parva* may be able to occupy the entire lower 48 states as the common carp has done, if no other factors could further limit the establishment or spread of these fishes.

INTRODUCTION

Most potential invasive fish species from Asia are not known invaders, but are likely to invade via the aquarium trade, fish farms, or live fish markets. 25 species selected for this analysis are of great concern to management agencies in the U.S. as they have shown up in US markets or are being considered for importation. They are mainly cyprinids, including *Opsariichthys uncirostris, Zacco platypus, Leuciscus waleckii, Squaliobarbus curriculus, Elopichthys bambusa, Megalobrama terminalis, M. amblycephala, Parabramis pekinensis, Hemiculter leucisculus, Distoechodon tumirostris, Plagiognathops microlepis, Abbottina rivularis, Hemibarbus labeo, H. maculatus, Pseudorasbora parva, Rhodeus ocellatus, Cirrhinus molitorella*, *Sinilabeo decorus,* and the Asian catostomid -- *Myxocyprinus asiaticus.* The next group are percifomes, such as *Channa maculata* (Channidae, Perciformes), *Micropercops swinhonis* (Ondontobutidae, Perciformes), *Perccottus glehni* (Odontobutidae, Perciformes), *Siniperca chuatsi* (Sinipercidae, Perciformes). The end group includes *Anguilla japonica* (Anguillidae, Anguilliformes), and *Plecoglossus altivelis* (Osmeridae, Osmeriformes). Little biological data are available for most of these fishes. In this chapter, I first introduce their ecological and biological backgrounds in their native habitats. I then use GARP to build the native-range niche models for each species and project them to the conterminous United States to predict the potential distribution range and assess their potential invasion risk in North American. As there are no occurrence data in North America for these fishes, I include the common carp, *Cyprinus carpio*, which was introduced 100 yr ago and now occurs in all of the lower 48 states, in the modeling analyses, hoping the predicting result and validation for common carp can help test the general modeling method and data applied in this chapter.

Opsariichthys uncirostris **(Cyprinidae, Cypriniformes)**

The three-lips, *Opsariichthys uncirostris*, is a small ferocious fish, native to China, eastern Siberia, northern Korea, and Japan (Froese and Pauly 2006). It inhabits swift mountainous streams or shallow gravel-bottomed waters. It is rarely found in the stagnant lakes or deep pools of rivers. It forms shoals with *Zacco platypus*. The adults feed on small fishes and crustaceans, the juveniles on plankton (Wu et al. 1964). It reaches maturity and starts to spawn in one year (Chen et al. 1998). Spawning usually occurs from March to June (Wu et al. 1964).

Three-lips have been accidentally introduced with Chinese carp fry into the Balykchi fish farm, Tashkent, Uzbekistan in 1960s. Individuals escaped into the Syr Darya River outlet channels, and the species spread into basins of other rivers through numerous transports of fish seed for aquaculture. It is established in Tashkent where it shows higher growth rates and greater fecundity than it does in native habitats.

Adverse ecological impacts have been reported (Froese and Pauly 2006).

Zacco platypus (Cyprinidae, Cypriniformes)

The small freshwater minnow *Zacco platypus* is native to China, Korea and Japan. It is distributed widely in China, from Lancang River, Pearl River in South China, north to the Yangtze River, Yellow River, and Amur River in northeastern China (Chen et al. 1998). It inhabits small gravel-bottomed streams and mainly feeds on crustaceans, small fishes, algae, and organic detritus (Chen et al. 1998). It reaches sexual maturity in one year, and spawning usually occurs between April and June in swift shallow waters. Its invasive potential and ecological impacts on native fauna are unknown.

Leuciscus waleckii (Cyprinidae, Cypriniformes)

The Amur ide, *Leuciscus waleckii*, is small to medium-sized fish native to China, Korea, and Amur River basin. It is common in the Yellow River drainage, and northern rivers and lakes in China (Chen et al. 1998). It prefers rivers with slow current, and feeds on aquatic insect larvae, flying terrestrial insects, algae and organic detritus (Zhang 1995, Froese and Pauly 2006). Individuals spawn at 3 yr, and spawning usually occurs in sand or gravel-bottomed streams during April and May.

Eggs are transparent, yellowish, adhering to sands or gravels (Zhang 1995). Its invasive potential and ecological impacts are unknown.

Squaliobarbus curriculus **(Cyprinidae, Cypriniformes)**

The barbel chub, *Squaliobarbus curriculus*, is native to China, western Korea and Viet Nam. It is found in all drainages except the Tibetan Plateau in China (Chen et al. 1998). It inhabits slow-running rivers and lakes and feeds on algae, aquatic insects and small fishes (East China Sea Fisheries Research Institute et al. 1990). Individuals grow slowly, weighing less than 500 g in three years and reach a maximum weight of this fish could be 2,500 g (Chen et al. 1998). Individuals reach sexual maturity at two years, and spawning usually occurs in June and July (East China Sea Fisheries Research Institute et al. 1990). Its invasive potential and ecological impacts are unknown.

Elopichthys bambusa **(Cyprinidae, Cypriniformes)**

The yellowcheek, *Elopichthys bambusa*, is native to China and Vietnam. It is widely distributed in Pearl River, Yangtze River, Yellow River, Amur River and all eastern drainages in China (Chen et al. 1998). It prefers the middle-upper waters. It is a fast, voracious predator, feeding mainly on other juvenile fishes (Yang 1987). It reaches sexual maturity in 3 yr. with the main growth season between April and June. The maximum weight may reach 50 kg (Yang 1987). It is thought detrimental to other fishes in aquaculture and is expelled from fishing ponds by fish farmers.

The yellowcheek has been incidentally introduced alone or with other fish species to Uzbekistan, and is established there. Ecological impacts are yet unknown (Froese and Pauly 2006).

Megalobrama terminalis (Cyprinidae, Cypriniformes)

This species is native to China, where it is originally found in the Pearl River drainage and drainages of Hanan Island (Chen et al. 1998). It has been successfully introduced to other parts of China (Yang 1987, East China Sea Fisheries Research Institute et al. 1990, Mao and Xu 1991, Zhang 1995). It inhabits middle-lower parts of open quiet waters, feeding on aquatic plants, zooplankton, crustacean and organic detritus (Mao and Xu 1991). Sexual maturity is reached in three years, and spawning occurs in late spring to early summer in the swift currents of upper stream reaches (Mao and Xu 1991). After spawning, individuals return to the middle or lower reaches of rivers, where they overwinter in deep pools or lakes (Mao and Xu 1991). This species has been introduced to Albania, Yugoslavia, Romania, and Hungary, but the status of established populations there and ecological impacts on native species are unknown (Froese and Pauly 2006).

Megalobrama amblycephala **(Cyprinidae, Cypriniformes)**

A native to China, this species mainly inhabits lakes and ponds along the middle-lower Yangtze River valley (Chen et al. 1998). Individuals normally stay in the middle-lower waters, but overwinter in deep waters. This fish is a herbivore, mainly feeding on aquatic plants, such as *Vallisneria asiatic* and *Hydrilla verticillata*, but also consuming vegetable detritus and zooplankton (Yang 1987). It reaches sexual maturity in 2 years, and spawning occurs in the mud-sandy streams or lakes with abundant aquatic vegetables at water temperature of 20-28 ˚C during May and June (East China Sea Fisheries Research Institute et al. 1990). The sticky eggs need to attach to leaves or roots of aquatic plants to hatch (Yang 1987). Development is rapid, with individuals measuring 120-230 mm in one year. The reported maximum weight is about 5 kg (Yang 1987). It has been successfully introduced to other parts of China as an aquacultural fish since the 1960s. It was also introduced to and established in Japan and Taiwan Island in the 1970s (Froese and Pauly 2006).

Parabramis pekinensis **(Cyprinidae, Cypriniformes)**

This species's native range is extensive: from Pearl River, Ming River, Hainan

Island in southern China, to Qiantanjiang River, Yangtze River, Huai River, Yellow River, until up to Liao River, Sungari, Ussuri River, Amur River in northeastern China, and adjacent drainages in Russia and Korea (Chen et al. 1998). It usually dwells in the middle-lower parts of waters. It is an omnivore, feeding on aquatic vascular plants, algae, crustaceans, rotifers, insects and small molluscs (East China Sea Fisheries Research Institute et al. 1990). It reaches sexual maturity in 3 years, and spawning occurs in lakes or rivers with a certain current velocity between May and July (Mao and Xu 1991). Numbers of eggs released ranges between 28,000-90,000 and are 0.9-1.2 mm in diameter. Eggs hatch in 3 days at 21-24 ˚C (East China Sea Fisheries Research Institute et al. 1990). It has been introduced to Albania, Greece, Turkmenistan, Uzbekistan, Romania and Hungary, and it is reported as established in Turkmenistan and Uzbekistan (Froese and Pauly 2006). The ecological impacts are unknown.

***Hemiculter leucisculus* (Cyprinidae, Cypriniformes)**

This is a small fish, native to Vietnam, China, Korea and Russia. It is widely distributed in China, occurring almost in all rivers and lakes (Chen et al. 1998). It is omnivorous, well adapted to both moving and quiet water habitats. Individuals often dwell in the upper waters along shore sides, acting promptly, but overwinter in the deep waters (Mao and Xu 1991). Adults mainly feed on floating algae, detritus of vascular plants, crustaceans, oligochaetes and insects, while the juveniles mainly feed on zooplankton, aquatic insects and mollusks. It reaches sexual maturity in one year, spawning between May and July in slow-flowing lotic or shallow lentic areas. Eggs are released in batches of 8500-12,000 eggs. The viscous eggs usually adhere to weeds or gravels to hatch (Mao and Xu 1991). It reproduces and grows rapidly, providing food not only for humans but for piscivorous fishes (Mao and Xu 1991). The body length of this fish ranges 100-140 mm, but may reach 240 mm (Mao and Xu 1991). It has been introduced to Iran, Afghanistan, and Russia, and is reported being established with adverse ecological impacts (Froese and Pauly 2006).

Distoechodon tumirostris **(Cyprinidae, Cypriniformes)**

This fish is found in Pearl River, Yangtze River, and southeastern seaboard drainages of China (Zhen 1989, Chen et al. 1998). It dwells in the middle and lower part of deep pools along clear swift streams (Mao and Xu 1991). It is often observed in groups browsing on the algae on the rocks or pebbles near the river banks during the hot days of summer or autumn (Mao and Xu 1991). Spawning occurs in the late spring or early summer on the pebbles in the shoals and riffles when rivers rise abruptly after heavy rains. Viscous eggs adhere to pebbles to hatch (Mao and Xu

1991). It spends the winter season by moving down to the deep pools or swimming downstream to the main rivers in groups (Mao and Xu 1991). Its invasive potential and ecological impacts are unknown.

Plagiognathops microlepis (Cyprinidae, Cypriniformes)

A native to China and Russia, this species is found in the Pearl River, Yangtze River, Yellow River, Amur River, and southeastern seaboard drainages of China (Chen et al. 1998). It dwells in the middle and upper waters of the slow-flowing, calm wide river channels in the middle or lower reaches of main rivers or large tributaries (Mao and Xu 1991). It feeds on algae, fragments of plants, organic detritus, also zooplankton and benthonic animals (Mao and Xu 1991). It reaches sexual maturity in 2 y. In the raining season of early summer when rivers rise abruptly, this fish swims in schools upstream to the swift current to spawn on pebbles. Viscous eggs adhere to pebbles to hatch. After spawning, this fish returns to its original habitats. In the late autumn when temperature drops, this fish moves to the deep waters of rivers to spend the winter (Mao and Xu 1991). It can reach 70.0 cm in total length, and weigh 3,000 g (Froese and Pauly 2006). It is a favorite table fish in China, as the flesh is very palatable. Its introduction history and ecological impacts are unknown.

Abbottina rivularis (Cyprinidae, Cypriniformes)

This is a small fish, native to eastern China, Korea and Japan (Froese and Pauly 2006). It is found ubiquitously in all drainages in China except high plateau regions (Chen et al. 1998). It inhabits shallow zones of sluggish rivers, lakes, ponds and ditches with sandy or muddy bottoms (Mao and Xu 1991, Froese and Pauly 2006). It is omnivorous, feeding mainly on cladoceran, copepods and Amphipods, and next on aquatic insects, earth worms and weed shreds (East China Sea Fisheries Research Institute et al. 1990). It reaches sexual maturity in one year (Froese and Pauly 2006). Males dig out nests for spawning in the muddy and sandy bottom, thus this species is nicknamed “sand puffer” in China (Mao and Xu 1991). Berg (1964) reported that the male builds a nest 12-43 cm in diameter on the bottom of the river, at a waterdepth of 8-34 cm, and broods over the spawn. 1,711 eggs were found in a single nest.

This species has been introduced to Kazakhstan, Kyrgyzstan, the Mekong Basin (Kottelat 2001) and the Tedzhen River Basin in Turkmenistan, and is reported as established (Froese and Pauly 2006). Ecological impacts are unknown.

Hemibarbus labeo (Cyprinidae, Cypriniformes)

A large stream fish, native to eastern Asia, this species is found from northern Viet Nam to southeastern Siberia (Froese and Pauly 2006), all drainages of Taiwan

Island, Mingjiang River, Qiantanjiang River, Yangtze River, Yellow River, and Heilongjiang River in China (Chen et al. 1998). It dwells in the sandy bottom of swift streams or rivers (Mao and Xu 1991). The adults feed mainly on molluscs, crustacea and aquatic insects, on various algae, and occasionally on vascular plants, cladocerans, and copepods. The juveniles feed on zooplankton, aquatic insect larvae, and benthonic worms (Chen et al. 1990). It reaches sexual maturity in two years, and spawns on pebbles of upstream channels between April and June (Chen et al. 1990). Most individuals weigh about 200 g, but can reach 3,000 g (Wu 1989). The flesh is very palatable and not bony, and is a famous dish in local wedding parties (Chen et al. 1990). It has been introduced to and established in the Mekong Basin in Laos, and possibly north in the Mekong Basin in China (Froese and Pauly 2006).

***Hemibarbus maculatus* (Cyprinidae, Cypriniformes)**

This species is native to eastern Asia, being found in China, Korea, Japan and Amur River basin. In China, it is widely distributed in all drainages except western plateaus (Mao and Xu 1991). It dwells in the middle and lower depths of rivers, lakes, and reservoirs, feeding on aquatic insects, mollusks, and earthworms (Chen et al. 1990, East China Sea Fisheries Research Institute et al. 1990). Sexual maturity is reached in two years, and spawning occurs from April to early May at water temperature of about 18˚C. The egg load of a normal female is about 30,000-40,000 eggs (Chen et al. 1990). Eggs are laid on soft weeds, and take about 85 hours to hatch at water temperature 17-20˚C (East China Sea Fisheries Research Institute et al. 1990). Average weight is 500 g, but some individuals reach 1000 g. This fish is very palatable, and traditionally served at weeding parties in Tonglu County, Zhejiang Province. Population numbers have declined rapidly in recent years and conservation measures are needed (Chen et al. 1990).

Pseudorasbora parva (Cyprinidae, Cypriniformes)

The topmouth gudgeon, *Pseudorasbora parva* Schlegel, 1842, is a small cyprinid fish, native to Asia, including China, Taiwan Island, southern and central Japan, Korea, and the Amur Basin (Froese and Pauly 2006). It is found ubiquitously in almost all drainages in China, occurring in the middle and upper waters of rivers and lakes, small creeks, ponds, and ditches (Wu 1989, Mao and Xu 1991, Chen et al. 1998). It feeds on small aquatic insects and larvae, copepods, cladoceran, algae, vascular weeds, planktons, and fish eggs (Wu 1989, Chen et al. 1990, Shaanxi Aquaculture Institute and Biology Department of Shaanxi Normal University 1992). Sexual maturity is reached in one year, and spawning occurs between April and June. The eggs are released in batches, ranging 500-1700 (Chen et al. 1990). The viscous ellipsoid eggs are laid neatly in rows on the surface of solid objects, and the male guards the eggs. It takes 44 hours for the eggs to hatch at 22.5-24.5˚C (East China Sea Fisheries Research Institute et al. 1990). This species grows slowly, and individuals are usually lees than 100 mm in standard length. It has little commercial value, but is efficient in controlling flies.

The topmouth gudgeon has been introduced widely (Welcomme 1988), in Iran and Turkmenistan (Froese and Pauly 2006), the Dnieper, Dniester, Danube basins, the Scutari and Prespa lakes, and Aliakmon River in Greece (Reshetnikov et al. 1997). Several countries report adverse ecological impact after introduction, and it is generally regarded as a pest species which competes with the fry of other species due to its high reproductive rate (Froese and Pauly 2006).

***Rhodeus ocellatus* (Cyprinidae, Cypriniformes)**

The rose bitterling, *Rhodeus ocellatus*, is a small cyprinid, native to Asia, in the Mekong, Hangjiang, Pearl, Yangtze and Yellow rivers, as well as Hainan Island (Chen et al. 1998). It is believed originally native only to Taiwan Island, and was introduced into eastern China, the Korean Peninsula, and Japan, areas that were once considered as the native range (Froese and Pauly 2006). This fish uses mussels as spawning sites; its habitats are intimately tied to the distribution of swan mussels (*Anodonta*), freshwater unionid mussels, and other genera of clams including *Pseudanodonta*, *Cristaria*, *Margaritifera* and *Dahurinaia* (Smith et al. 2004). This fish favors heavily vegetated areas of small lakes, ponds, sluggish river backwaters, oxbows, and ditches with fine sandy or thin muddy bottoms (Smith et al. 2004). It is omnivorous, feeding on diatoms, detritus, aquatic insects, crustaceans, fish eggs (Mao and Xu 1991, Holcık 1999). Sexual maturity is usually attained during the second or third year. Spawning mainly occurs in April and May at water temperatures of 12-24°C, although 15-21°C is optimal (Mao and Xu 1991, Holcık 1999). The female develops an ovipositor from the genital opening, up to 6 cm long. The ovipositor lays eggs inside freshwater clams and mussels, using the excurrent siphon as the entry route. Apparently the flow of water out of this siphon encourages egg laying. Before egg laying, the female nudges the clam repeatedly to accustom the mollusk to stimulus so that it does not close up its shell. Males select and defend a particular clam against other males. The female deposits 1-2 eggs at a time and the male sheds sperm which are sucked into the clam on its feeding current. Fecundity is up to 22,136 eggs. Eggs hatch for 2-5 weeks and the young leave the clam after 2 days when the yolk sac has been absorbed (Smith et al. 2004). The eggs and young are protected inside the mussels, while the mussels disperse their own young or glochidia through attachment to the fins of the fish.

This fish is rarely used as food, but an extremely attractive fish in aquarium, graceful in form and movement. It has been introduced and is established in Uzbekistan (Froese and Pauly 2006), reported from the Karakum Canal and Kopetdag Reservoir in Turkmenistan, in the Tedzhen (= Hari) River and Caspian Sea basins of Iran (Coad 2007). Ecological impacts are unknown, but at least one country reports adverse ecological impacts after introduction (Froese and Pauly 2006).

Cirrhinus molitorella (Cyprinidae, Cypriniformes)

This species is native to China and Vietnam, including the Pearl, Yuanjiang, Ming, and Mekong rivers, and Hainan Island (Yue 2000). The distribution in the Ming River is probably as a result of escapes from fish farms (Wu et al. 1964). It dwells in the middle and lower depths of large and medium-sized rivers, feeding on algae, phytoplankton, and detritus. It prefers flowing water and is not known to proliferate in impoundments (Froese and Pauly 2006). This fish cannot tolerate water temperature lower than 7˚C, so it overwinters in deep water, and fish farming is limited to the southern part of Fujiang Province (Wu et al. 1964). When the water temperature is lower than 14˚C, this fish dives to deep water and becomes sluggish; it starts to die when water temperature drops to 7˚C (Institute of Aquaculture of Guangxi Autonomous Region and Institute of Zoology of Academia Sinica 1981).

Sexual maturity is reached in 3 years when the female weighs about 500 g. The spawning period is long, from April and May to August and September (Wu et al. 1964). Individuals congregate on the shoals and ripples in the middle or upper reaches of rivers during the flooding season, chasing each other and making “goo goo” sounds. Spawning occurs when water level abruptly rises and current increases at the optimal water temperature 26-30˚C (Institute of Aquaculture of Guangxi Autonomous Region and Institute of Zoology of Academia Sinica 1981). Eggs are yellow-green, dilating in the water and drifting down along streams; it takes about 24 hours for eggs to hatch (Institute of Aquaculture of Guangxi Autonomous Region and Institute of Zoology of Academia Sinica 1981). The wild fish can grow to 2,000-2,500 g, the maximum up to 8,000 g (Wu et al. 1964). It is an important local aquaculture commercial fish. Its introduction history and ecological impacts are unknown.

***Sinilabeo decorus* (Cyprinidae, Cypriniformes)**

This species is native to southern China, found in Xijiang River and Beijinag River of the Pearl River drainage (Yue 2000); Yuanjiang River, Dongting Lake, Honghu Lake, and their upstream tributaries (Institute of Aquaculture of Guangxi Autonomous Region and Institute of Zoology of Academia Sinica 1981). It inhabits the bottom of swift mountain streams, feeding on algae on rocks and pebbles, or plant detritus (Wu 1989). Sexual maturity is reached in 3 years. Individuals congregate shoals or riffles to spawn between March and April. Eggs are released mostly in night. The eggs are yellowish and immediately sink to the bottom. The fecundity is 17,000-20,000 eggs for the average female. This fish grows slowly, reaching 165-377 mm in standard length. The maximum weight is about 4,000 g (Institute of Aquaculture of Guangxi Autonomous Region and Institute of Zoology of Academia Sinica 1981). The flesh is highly prized, and the species is a locally important commercial fish. Its invasive potentials and ecological impacts are unknown.

Myxocyprinus asiaticus (Catostomidae, Cypriniformes)

The Chinese sucker, *Myxocyprinus asiaticus*, is only found in the Yangtze River and associated lakes. Body color of juvenile fish is brown with three slanted dark bands on side of body; male adults are red, and female adults are dark purple, with a broad rouge-red vertical zone along lateral of body. It prefers flowing water, and is more common in the Yangtze River above Yichang city. The juveniles and adults mostly dwell in middle-lower waters, and the fry prefer living near the surface. Diet consist of benthonic invertebrates, organic detritus in the mud, and aquatic insect larvae, but it also grazes algae on rocks or pebbles (Wu et al. 1964). Sexual maturity is reached in five or six years, and spawning all occurs between March and April in the upstream tributaries of Yangtze River, such as Jinshaojiang River, Mingjiang River and Jialingjiang River (Wu et al. 1964, Yang 1987, Wu 1989). Growth is rapid, reaching 198 mm in the first year, 346 mm in the second year, 496 mm in the third year and 608 mm in the fifth year. Individuals of 2-2.5 kg are common in the market; the maximum reported weight is 50 kg (Yang 1987). Invasive potential and ecological impacts are unknown.

***Channa maculata* (Channidae, Perciformes)**

The blotched snakehead, *Channa maculata*, is native to southern China and northern Vietnam. It is distributed in all drainages south of the Yangtze River Basin, and Hainan Island (Institute of Aquaculture of Guangxi Autonomous Region and Institute of Zoology of Academia Sinica 1981, Courtenay and Williams 2004). It inhabits streams, lakes, ponds, and ditches, and prefers shallow waters with vegetation. It is a fierce predactor, feeding on crustaceans, large insects, frogs, and fishes, and hunting style is described as “hides among rocks or aquatic plants until its prey approaches, then it quickly attacks, kills, and swallows its victim” (Hay and Hodgkiss 1981).

The blotched snakehead builds a circular bubble nest among rooted vegitations. Spawning occurs between April and June in Guangxi, China. The eggs floats in the nest and are guarded by both male and female or only by male fish. The fry are guarded when swimming around for food, but are occasionally eaten by parents when the food is scarce (Institute of Aquaculture of Guangxi Autonomous Region and Institute of Zoology of Academia Sinica 1981). Weight reaches 500-1000 g, with a maximum of 2000-2500g. The flesh is very palatable, and said to be helpful in recovery after a surgical operation (Institute of Aquaculture of Guangxi Autonomous Region and Institute of Zoology of Academia Sinica 1981).

The blotched snakehead has been introduced to Taiwan, Japan, Madagascar, and Hawaii, and is widely established (Courtenay and Williams 2004, Froese and Pauly 2006). It is a valuable food fish in southern China and Taiwan. It has also appeared in a live-food fish market in Boston, Massachusetts (Courtenay and Williams 2004). Ecological impacts are unknown yet.

Micropercops swinhonis **(Ondontobutidae, Perciformes)**

This small benthonic fish is native to China, Korea and Japan (Froese and Pauly 2006). It is widely distributed in China, inhabiting the shallow waters of rivers, lakes, ponds, creeks, ditches, and irrigation channels (Yang 1987). It feeds on crustaceans, small fishes, and zooplankton (Wu 1989, Mao and Xu 1991). In the breeding season, the male makes a nest and territory, and entices the female to the nest for spawning.

Fecundity is about 257 eggs for the average female (Zhang 1995). It has been introduced to Tashkent, Uzbekistan, and adverse ecological impact was reported after introduction (Froese and Pauly 2006).

Perccottus glehni (Odontobutidae, Perciformes)

A small fish, native to northeastern China, northeastern Korea, and the Amur River basin (Froese and Pauly 2006). This species prefers stagnant rivers, lakes and bogs; especially lakes with heavy vegetation. It tolerates low oxygen concentration (Zhang 1995) and feeds on aquatic insects, insect larvae, and small shrimps. Large individuals occasionally feed on small fishes (Zhang 1995). Sexual maturity is reached in two years, with a length of 50-60 mm in standard length. Spawning occurs during May and June at water temperature of 15-20˚C. The fecundity is about 1000 eggs per female; The eggs are elliptic and viscous (Zhang 1995). This is a small fish of little commercial value in its native range. It has been introduced accidentally to eastern Slovakia, Latvia, Poland and Uzbekistan, and is reported as established and causing adverse ecological effects through competition for identical habitats with native fishes (Froese and Pauly 2006).

Siniperca chuatsi (Sinipercidae, Perciformes)

The Chinese perch, *Siniperca chuatsi*, is native to China and the Amur River Basin. It is found in all rivers, streams, lakes and ponds in China (Wu et al. 1964). It prefers stagnant or sluggish waters, especially weedy lakes or ponds (East China Sea Fisheries Research Institute et al. 1990, Mao and Xu 1991). When the temperature drops to 1-5 °C in late autumn this fish moves to deep waters to overwinter. When temperature rises above 15 °C in the spring, the fish moves to shallow waters and starts to feed actively. The Chinese perch has a broad temperature tolerance range, which is why it is found in most parts of China.

The Chinese perch is a demersal stalking piscivore, feeding on live fish, shrimps, and other aquatic invertabrates (Wu et al. 1964). It prefers feeding at night, and usually stays sluggishly in sheltered areas during the daytime. It relies on motion-sensitive vision to follow the prey's movement before leaping forward and snapping at the victim. Therefore, its prey are mainly diurnal fishes whose eyes have color vision and high acuity but cannot function at night. It will not take immobile prey even when they are very close, but as soon as the prey start to move this fish strike. The Chinese perch start to consume fish fry of other species when four or five days old. When the body length reaches 40-50 mm it starts to feed on small shrimps; when reaching 100 mm in total length, it starts to feed on adult fishes (Wu et al. 1964). In the native habitats they prefer fusiform fishes, such as *Megalobrama* *amphylocephalus* and *Carassius auratus*. The Chinese perch grows quickly, reaching 400-600 g in one year and 1000-1500 g in two years.

Sexual maturity is reached in two years for females, with a minimal weight of 160-250 g and body length of 21 cm, while only in one year for males with a minimal weight of 80 g and body length of 16 cm. In southern China, spawning occurs between March and August, while in central China it occurs between May and August. During the breeding season matured fish congregate the river bend or confluence where current is enhanced. Spawning takes place only in running waters at nighttime when temperature is above 21°C. The fecundity is normally 40,000-90,000 eggs (Wu et al. 1964). Eggs are 1.2-1.4 mm in size, semi-pelagic, and hatch in three to four days (Mao and Xu 1991).

The flesh is very palatable, with few inter-muscular bones, and was a well-known food fish even in ancient China where it was extremely popular in the Tang Dynasty (618-907 A.D.) garnishing numerous extols of its color and taste by poets. Its invasive potentials and ecological impacts are unknown.

***Anguilla japonica* (Anguillidae, Anguilliformes)**

The Japanese eel, *Anguilla japonica*, is native to Asia: Japan, Taiwan, Korea, China and northern Philippines (Froese and Pauly 2006). The northern limits are the island of Hokkaido, the coast of the Bohai Sea, and the Liao Ho River. The southern limits are the island of Hainan and the Gulf of Tonkin. It is a catadromous fish, spawning in the sea, but developing and growing in fresh water. Young Japanese eels enter rivers in small schools from February to May, and ascend to the upper reaches of rivers and mountain lakes. Males like to stay in the estuaries and lower reaches of rivers; while females prefer moving upstream to upper reaches of rivers (Mao and Xu 1991). Small eels are able to crawl over land at night or during the rainy season from one place to another. Individuals seek caves or shelters among rocks during daytime to avoid light; at night they come out to feed on benthic crustaceans, small fish, insect larvae, and carcasses of large animals (Mao and Xu 1991). After about 5-6 years in freshwater, when sexual maturity is reached, the eels migrate downstream and enter the sea to spawn from August to October (Mao and Xu 1991). Spawning grounds of Japanese eel are presumed to be in the western Mariana Islands, at a salinity front near 15°N and 140°E (Tsukamoto 1992). The Japanese eel has been introduced to many places, but no reports found it established. It was collected from streams in California, and these specimens are assumed to be the result of aquaculture escape (Welcomme 1988). Ecological impacts are unknown.

***Plecoglossus altivelis* (Osmeridae, Osmeriformes)**

The Ayu, or Ayu sweetfish, *Plecoglossus altivelis*, is native to Japan, Korea and China (Wu et al. 1964, Editorial Subcommittee of Fishes of Fujian Province 1984). It is a typical amphidromous fish, preferring clean river water, and can be found in lakes and rivers, and river reaches from the head to the mouth (Froese and Pauly 2006). The adults usually spawn from August to October, in the lower reaches of rivers. After spawning, most females die while a few survived return to the sea. Larvae enter the sea immediately after hatching and remain there during winter, feeding on plankton (Wu et al. 1964). In springtime, the young (5-7 cm in total length) move upstream, feeding on blue-green algae, diatoms and insects. This fish has small leaf-like teeth loosely attached to the jaw with two ligaments, an adaptation to harvesting algae off pebbles. It is reported it can travel 20 km in one day and is able overcome pretty large obstacles (Wu et al. 1964). Individuals grow fast and typically reache their home range by August. By September and October, they move downstream, congregating lower reaches of rivers to spawn. Spawning occurs in gravel-bottomed, clear, shallow, waters with swift current at dawn or dusk (Wu et al. 1964). Eggs are viscous, forming a dark blue rounded mass, and adhere to gravels to hatch. (Editorial Subcommittee of Fishes of Fujian Province 1984).

This species is small, with natural individuals ranging from 150 to 200 mm in length, rarely reaching 300 mm. However, it is highly esteemed both as commercial food fish and game fish, especially in Japan where it serves as a delicacy. It is cultured commercially everywhere in western Japan. Although this fish is not listed as an endangered fish in China, it has completely disappeared in some streams due to habitat alteration. An introduced population is reported established on Taiwan Island, but introductions have failed in Russia and Hawaii (Froese and Pauly 2006). Ecological impacts are unknown.

METHODS

Environmental Data Sources. 15 environmental variables were used in this analysis, which include aspects of topography (elevation, topographic index, flow accumulation, slope and aspect from USGS Hydro-1K data set; http://edcdaac.usgs.gov/gtopo30/hydro/), percent tree cover (Hansen et al. 2003), and Climatic conditions (annual means of diurnal temperature range; frost days; precipitation; maximum, minimum and mean monthly temperatures; solar radiation; wet days; and vapor pressure; for 1960-1990 from the Intergovernmental Panel on Climate Change Worldwide Climate Data Distribution Centre; http://ipcc-ddc.cru.uea.ac.uk/index.html). All analyses were confined to the region bounded by 24.5988-53.7988˚ N, 66.1417-125.0217˚ W in North America, and the native range in Asia (18.8300-50.6900˚N, 96.1616-145.7416˚E) for all species except

Cirrhinus molitorella (10.5976˚S-55.1524˚N, 92.8336 – 145.4336˚E). The environmental data sets were converted to a resolution of 0.01˚ for the model building.

Occurrence Data Sources. Native occurrence data for each species were obtained from the Wuhan Institute of Hydrobiology, Beijing Institute of Zoology, Kunming Institute of Zoology, Chinese Academy of Sciences, and scientific literature such as the provincial fish faunas in China, FishNet (http://speciesanalyst.net/fishnet/), and FishBase (http://www.fishbase.org/search.html). Occurrence data for Asian records were georeferenced using the Geonames Query web tool (http://gnpswww.nima.mil/geonames/GNS/index.jsp). In all cases, points outside the known native range were excluded from the training data pool, and duplicate occurrence points were removed from the data pool, keeping only verified, unique occurrence points for modeling. The number of native occurrence points per species ranged from 364 for *Pseudorasbora parva* to 36 for *Myxocyprinus asiaticus*.

Evaluating Environmental Variables. The environmental variables were subjected to a jackknife procedure, which allows exclusion of environmental variables that can lead to spurious overfitting. Hence, for N environmental coverages, N analyses are run using all combinations of N-1 environmental coverages. Then, coverages are evaluated via correlations between inclusion/exclusion of the environmental variables and the average omission error (i.e., predicting absence at sites of known presence). Environmental variables correlated with increased omission error were excluded from further analysis, following Peterson and Cohoon (1999).

Model Building. Native occurrence data for each species with over 50 points were randomly divided into two subsets to permit model training (80%) and model validation (20%), with a minimum of at least 20 validation points to insure statistical power. For those species with less than 50 occurrence points, no points set aside for independent validation, all were used for model building.

All native-range niche models were generated using DesktopGARP (http://nhm.ku.edu/desktopgarp/). The details of use of GARP in ecological niche modeling have been presented in numerous publications, such as Anderson et al. (2003), Feria and Peterson (2002), Levine et al. (2004), and Wiley et al. (2003). In essence, GARP attempts to find nonrandom associations between environmental conditions and known occurrences of a species by evolving rules that predict presence or absence of the species. To accomplish this task, GARP uses a subset of the training data to formulate a rule and the rest to internally test the predictive accuracy of the rule. As rules are generated and evolved, the expectation is that the differences between one round of prediction and the next will decrease, converging on a final solution. The investigator can specify this convergence limit (0.01 in this study). Modeling continues until the convergence limit is reached, or a number of iterations specified by the investigator are run (in this study, 1000 times, which was never reached before convergence).

GARP will produce as many models as the investigator specifies. Because of stochastic elements in the process, some of these replicate models will be 'better' (i.e., more predictive) than others. Two criteria are used to evaluate model quality, omission error and commission "error." Omission error occurs when a model fails to predict potential for presence at known occurrence points. Commission "error" is more complex: it is calculated the proportional area predicted to be suitable but includes both true error (inappropriate conditions predicted as suitable), and apparent commission error (species may be present, but site was not sampled); hence, I refer to it as the "commission index." Among a set of models, those that have low omission error rates and that are close to the median commission index appear to offer best predictive ability (Anderson et al. 2003).

As such, I used the best-subsets option of desktop GARP to select the 10 best models from among the models generated by the algorithm. Here, I generated 200 initial to derive 20 models under an absolute omission error threshold of 0%, from which I selected the 10 with a commission index closest to the median. The 10 models were added together, pixel by pixel, to create a final prediction.

Model Evaluation. Model predictions were then evaluated using the validation data via Receiver Operating Characteristic (ROC) analysis, a method designed to evaluate the specificity (absence of commission error) and sensitivity (absence of omission error) of a diagnostic test (Zweig and Campbell 1993, Fielding and Bell 1997). It has been applied to testing the statistical accuracy of GARP results by Iguchi et al (2004), Wiley *et al*. (2003), and Chen *et al.* (2007), and more generally in niche modeling by Elith et al (2006). The area under the curve (AUC) in a ROC analysis is a measure of predictive accuracy for the model set as a whole: if the AUC = 0.50, then the best-model set is performing no better than random, but if the AUC is significantly higher than 0.5 (as judged by a z-test), then the result is significant. The higher the value of AUC, the better the model set, and a perfect prediction would have AUC = 1.0 (Hanley and McNeil 1982). The maximal value of the AUC score is achieved when all of the validation data points occur in pixels where all 10 of the best models predict presence, but it is influenced by the relative extent of the area predicted "present" compared to the total landscape examined (Wiley et al. 2003).

The predictive accuracy of the native-range niche models for each species was also calculated as the percentage of validation points within all 10 best models. When all validation points are successfully predicted by all 10 best models, the accuracy is

100%. Maximum commission index (used as maximum potential distribution index) was calculated as the percentage of the total area predicted present by any of 10 best models, while average commission index (used as average potential distribution index) was calculated as the average percentage of the total area predicted present by all 10 best models. Their difference indicates the amount of overlap inside the 10 best models. Therefore, the native-range models with higher overlap, e.g., little difference between the maximum commission index and average commission index, should be considered more reliable.

RESULTS

Environmental variables excluded from building final niche models for each species as the result of the jackknife procedure are indicated in Table 7-1. The sum of the 10 best models for each species could then be visualized across both the native-range and the conterminous United States (Figure 7-1). For all species for which there were independent validation data, model validations were highly significant over the native landscape ($P<0.01$) (Table 7-2), with AUC values varying from 0.739 for *Abbottina rivularis* to 0.9187 for *Cirrhinus molitorella*. The accuracy ranges from 86.0% for *Elopichthys bambusa* to 99.0% for *Leuciscus waleckii*.

The maximum commission index over the native landscape ranges from 0.1507 for *Perccottus glenii* to 0.7968 for *Pseudorasbora parva*, while the average commission index ranges from 0.1103 for *Perccottus glenii* to 0.6146 for *Pseudorasbora parva*. The maximum commission index over North America ranges from 0.0581 for *Myxocyprinus asiaticus* to 0.9179 for *Pseudorasbora parva*, while the average commission index ranges from 0.0099 for *Myxocyprinus asiaticus* to 0.7318 for *Pseudorasbora parva*. The areas predicted suitable for each species in North America are shown in Figure 7-1.

DISCUSSION

Although there is no invaded-range occurrence data for model validation, the predictive result for common carp is highly significant over the landscape of North America based on the 646 invaded-range occurrence points (AUC=0.7036). The predictive accuracy is 82.04%. This is very encouraging, considering that the majority of the unpredicted points occur in the southwestern U.S. and the Rockey Moutain states where artificial impoundments are common (Figure 5-1 E).

The predicted distribution of each species in North America is consistent with its native-range distribution. For example, *Myxocyprinus asiaticus* has the lowest potential distribution in North America (0.0581, 0.0099), predicted suitable only for the southern United States by ≤3 of 10 native-range models (Figure 5-1 O), and it is natively only found in the Yangtze River and associated lakes. *Pseudorasbora parva* has the highest potential distribution (0.9179, 0.7318), and may dominate all regions in United States except the desert areas if introduced (Figure 5-1 Q). Natively, it is widely distributed in East Asia, found ubiquitously in almost all drainages. *Perccottus glehni*, restricted to northeastern China, was predicted suitable only for the Great Lake states, Minnesota, and from South Dakota north to southern Canada (Figure 5-1 U). *Distoechodon tumirostris*, native to southern China, is predicted suitable mainly for southern United States, and the eastern midwestern states (Figure 5-1 F). *Parabramis pekinensis* is widely distributed in China but restricted to lower plain drainages, and the native-range niche model successfully predicted its potential suitable habitats in North America—in drainages of the lower Mississippi Basin and the Atlantic Coast (Figure 5-1 T).

Abbottina rivularis, *Hemiculter leucisculus*, *Hemibarbus labeo*, *Hemibarbus maculatus*, *Plagiognathops microlepis*, and *Pseudorasbora parva* with a larger commission index than common carp's. The invasion risk of these species may be as high as the common carp in North America. If no other factors could further limit the establishment or spread of these fishes, they all have the potential to occupy the entire lower 48 states. *Pseudorasbora parva* has many characteristics which favor a successful invader, such as resistance to harsh climatic conditions, early sexual maturity, extended breeding season, and broad dietary spectrum. It's quite possible that this fish will spread further as the common carp has. *Abbottina rivularis* and *Hemiculter leucisculus* are ecological generalists, and omnivores, with a short generation time in native landscape, and its invasive potential is unlimited. However, despite of its high fecundity and wide temperature tolerance in native habitats, populations of *Hemibarbus maculatus* has decreased rapidly in China as a result of habitat disturbance and water pollution.

Siniperca chuatsi, *Elopichthys bambusa*, *Micropercops swinohonis, Squaliobarbus curriculus*, *Leuciscus waleckii*, *Zacco platypus*, *Megalobrama amblycephala*, *Megalobrama terminalis, Opsariichthys uncirostris*, and *Rhodeus ocellatus* are fishes with a medium commission index, but they are likely to be locally established in North America. *Siniperca chuatsi*, a stalking piscivore, natively has a broad temperature tolerance range, and was predicted suitable for the entire eastern United States, and the Mississippi River basin (Figure 5-1 X). *Elopichthys bambusa*, a voracious predator, was also predicted suitable for the entire eastern United States, and the Mississippi River Basin (Figure 5-1 G). *Micropercops swinohonis* has been introduced to Tashkent and Uzbekistan, and caused adverse ecological impact (Froese and Pauly 2006). It was predicted suitable for eastern, southern and midwestern states, and as well as on the Pacific coast (Figure 5-1 N). *Squaliobarbus curriculus* has a broad native distribution, and was predicted suitable for eastern, southern and midwestern US, and in the Pacific coast and Columbia-Snake River valley (Figure 5-1 Y). *Leuciscus waleckii*, a cold-water fish in native habitats, was correspondingly predicted suitable for northern US (Figure 5-1 K). *Megalobrama amblycephala* is native only to the middle-lower reaches of Yangtze River (Figure 5-1 V), and *Megalobrama terminalis* is native only to drainages of Pearl River and Hannan Island in China (Chen et al. 1998) (Figure 5-1 L). Both were introduced to other parts of China as aquaculture fishes via artificial breeding (Chen et al. 1998). Their invasive ranges in North America may be overestimated by the native-range models, especially in the northeastern US, and the upper reaches of Mississippi River and its tributaries. *Zacco platypus*, natively having a wide temperature tolerance and short generation time, was predicted suitable mainly for all eastern US drainages east of longitude 100˚ (Figure 5-1 Z). However, considering that this small fish inhabits small gravel-bottomed streams in native habitat (Chen et al. 1998), its potential invasive ranges in North America may be overestimated. *Opsariichthys uncirostris* was predicted suitable for northwestern US and all of the eastern US east of longitude 100˚ (Figure 5-1 P). This predatory fish has a wide temperature tolerance and short generation time in native habitats, and is established in the Syr Darya River, Tashkent, Uzbekistan where it shows a higher growth rates and greater fecundity than it does in native habitats (Froese and Pauly 2006). It deserves special attentions for its likely invasion and possible negative ecological impacts in North America. *Rhodeus ocellatus* is an extremely attractive fish in aquarium, and has been successfully introduced out of its native range. Potential invasion is possible in the southern United States and the lower Mississippi River valley via aquarium trade (Figure 5-1 V).

Myxocyprinus asiaticus, *Channa maculata*, *Sinilabeo decoru*, and *Cirrhinus molitorella*, with a lower commission index, should have a very limited invasive potential in North America. *Channa maculata* and *Sinilabeo decoru* are natively distributed in southern China, and they were predicted habitable only in the Gulf coastal area and Atlantic coast in the southern United States by ≤4 of 10 native-range models (Figure 5-1 C, W). *Cirrhinus molitorella* is natively restricted to southern China and Vietnam, and is very sensitive to water temperature. This species is predicted suitable only for the Gulf coastal region and Florida (Figure 5-1 D). Therefore, there is no need to worry about *Myxocyprinus asiaticus*, *Channa maculata*, *Sinilabeo decoru*, and *Cirrhinus molitorella* establishing populations and spreading in North America.

Plecoglossus altivelis, an amphidromous fish, is predicted suitable for the drainages of southern and eastern US, and the Mississippi River drainages (Figure 5-1 R). The Japanese eel, *Anguilla japonica*, a catadromous fish, is predicted suitable for the drainages of the midwestern, southern, and eastern United States, and portions of the west coast (Figure 5-1 B), but so far no established population has been reported although it was introduced to many places in the world. The risk of invasion for these species should be low in the inland regions.

This study demonstrates the worst invasion scenario for these fishes, as biotic interactions are not available before actual invasion happens, and species' dispersal ability and effects of natural and human disturbance are hard to include in modeling. Generally, the United States is at risk of invasion. The most likely invasive ranges for these Asiatic freshwater fishes include the southern US, and drainages of the Mississippi River and its tributaries.

LITERATURE CITED

Anderson, R. P., D. Lew, and A. T. Peterson. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. Ecological Modelling **162**:211-232.

Berg, L. S. 1964. Freshwater Fishes of the U.S.S.R. and Adjacent Countries. volume 2, 4th edition. Israel Program for Scientific Translations Ltd, Jerusalem.

Chen, M., H. Tong, T. Yu, and Z. Diao. 1990. The Fish Resources of Qiantang River. Shanghai Scientific and Technological Literature Publishing House, Shanghai, China.

Chen, P., E. O. Wiley, and K. M. McNyset. 2007. Ecological niche modeling as a predictive tool: silver and bighead carps in North America. Biological Invasions **9**:43-51.

Chen, Y., X. Chu, Y. Luo, Y. Chen, H. Liu, M. He, P. Yue, S. He, W. Chen, and R. Lin, editors. 1998. Fauna Sinica: Osteichthyes -- Cypriniformes II. Science Press, Beijing, China.

Coad, B. W. 2007. Species Accounts-Cyprinidae-*Rhodeus*. *in* Freshwater Fishes of Iran. available online http://www.briancoad.com/Species%20Accounts/Rhodeus.htm. Access date 22

Feb. 2007.

Courtenay, W. R., Jr., and J. D. Williams. 2004. Snakeheads (Pisces, Channidae)--a biological synopsis and risk assessment. U. S. Department of the Interior, USGS.

East China Sea Fisheries Research Institute, Chinese Academy of Fisheries Science, and Shanghai Fisheries Research Institute, editors. 1990. The Fishes of Shanghai Area. Shanghai Scientific and Technical Publishers, Shanghai.

Editorial Subcommittee of Fishes of Fujian Province, editor. 1984. The Fishes of Fujian Province (Part I). Fujian Science and Technology Press, Fujian China.

Elith, J., C. H. Graham, R. P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. M. Overton, A. T. Peterson, S. J. Phillips, K. Richardson, R. Scachetti-Pereira, R. E. Schapire, J. Soberon, S. Williams, M. S. Wisz, and N. E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography **29**:129-151.

Feria, T. P., and A. T. Peterson. 2002. Using point occurrence data and inferential algorithms to predict local communities of birds. Diversity and Distributions **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. Environmental Conservation **24**:38-49.

Froese, R., and D. Pauly, editors. 2006. FishBase: www.fishbase.org, 10/2006 edition. World Wide Web electronic publication.

Hanley, J. A., and B. J. McNeil. 1982. The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve. Radiology **143**:29-36.

Hansen, C. M., R. S. DeFries, J. R. G. Townshend, M. Carroll, C. Dimiceli, and R. A. Sohlberg. 2003. Global Percent Tree Cover at a Spatial Resolution of 500 Meters: First Results of the MODIS Vegetation Continuous Fields Algorithm. Available from http://glcf.umiacs.umd.edu/data/treecover/index.shtml.

Hay, M. S., and I. J. Hodgkiss. 1981. Hong Kong Freshwater Fishes. The Urban Council, Wishing Printing Company, Hong Kong.

Holcık, J. 1999. *Rhodeus sericeus*. Pages 1-32 *in* P. M. Banarescu, editor. The Freshwater Fishes of Europe 5, I. Cyprinidae. AULA-Verlag, Wiebelsheim.

Iguchi, K., K. Matsuura, K. McNyset, A. T. Peterson, R. Scachetti-Pereira, K. A. Powers, D. A. Vieglais, E. O. Wiley, and T. Yodo. 2004. Predicting invasions of North American basses in Japan using native range data and a genetic algorithm. Transactions of the American Fisheries Society **133**:845-854.

Institute of Aquaculture of Guangxi Autonomous Region, and Institute of Zoology of

Academia Sinica, editors. 1981. The Freshwater Fishes of Guangxi Autonomous Region. The People's Publishing House of Guanxi, Nanning, China.

Kottelat, M. 2001. Fishes of Laos. WHT Publications Ltd., Colombo, Sri Lanka.

Levine, R. S., A. T. Peterson, and M. Q. Benedict. 2004. Geographic and ecologic distributions of the *Anopheles gambiae* complex predicted using a genetic algorithm. American Journal of Tropical Medicine & Hygiene **70**:105-109.

Mao, J., and S. Xu, editors. 1991. The Fauna of Zhejiang Province: Freshwater Fishes. Zhejiang Publishing House of Science and Technology, Hangzhou, China.

Peterson, A. T., and K. C. Cohoon. 1999. Sensitivity of distributional prediction algorithms to geographic data completeness. Ecological Modelling **117**:159-164.

Reshetnikov, Y. S., N. G. Bogutskaya, E. D. Vasil'eva, E. A. Dorofeeva, A. M. Naseka, O. A. Popova, K. A. Savvaitova, V. G. Sideleva, and L. I. Sokolov. 1997. An annotated check-list of the freshwater fishes of Russia. Journal of Ichthyology **37**:687-736.

Shaanxi Aquaculture Institute, and Biology Department of Shaanxi Normal University. 1992. Fishes of Shaanxi Province. Shaanxi Science and Technology Press, Xi'an, China.

Smith, C., M. Reichard, P. Jurajda, and M. Przybylski. 2004. The reproductive ecology of the European bitterling (*Rhodeus sericeus*). Journal of the Zoological

Society of London **262**:107-124.

Tsukamoto, K. 1992. Discovery of the spawning area for Japanese eel. Nature **356**:789-791.

Welcomme, R. L., editor. 1988. International Introductions of Inland Aquatic Species. FAO Fisheries Technical Paper 294. Food and Agriculture Organization of the United Nations (FAO), Rome, Italy.

Wiley, E. O., K. M. McNyset, A. T. Peterson, C. R. Robins, and A. M. Stewart. 2003. Niche modeling and geographic range predictions in the marine environment using a machine-learning algorithm. Oceanography **16**:120-127.

Wu, L. 1989. The Fishes of Guizhou Province. The People's Publishing House of Guizhou, Guiyang, China.

Wu, X., G. Yang, P. Yue, and H. Huang, editors. 1964. The Economic Animals of China -- Freshwater Fishes. Science Press, Beijing.

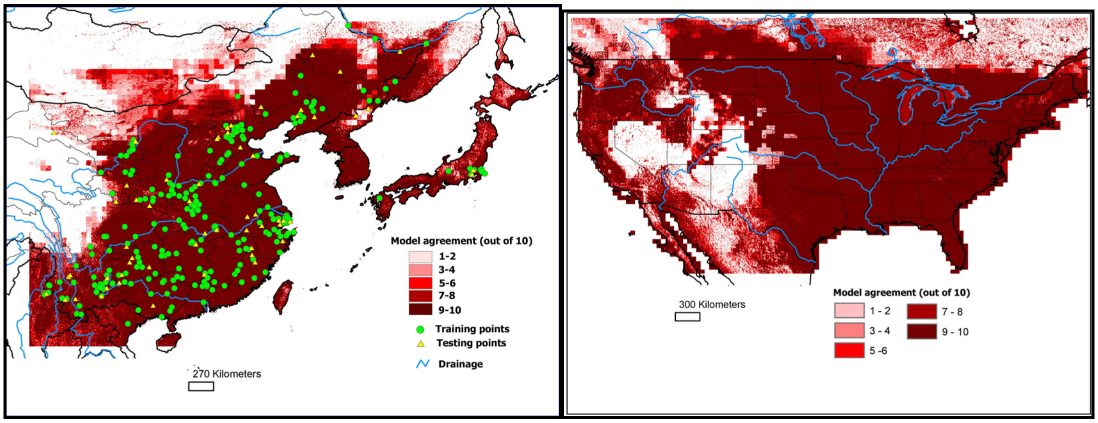
Yang, G., editor. 1987. The Fishes of Hubei Province, China. Hubei Science and Technology Press, Wuhan, Hubei.

Yue, P., editor. 2000. Fauna Sinica, Osteichthyes-Cypriniformes III. Science Press, Beijing, China.

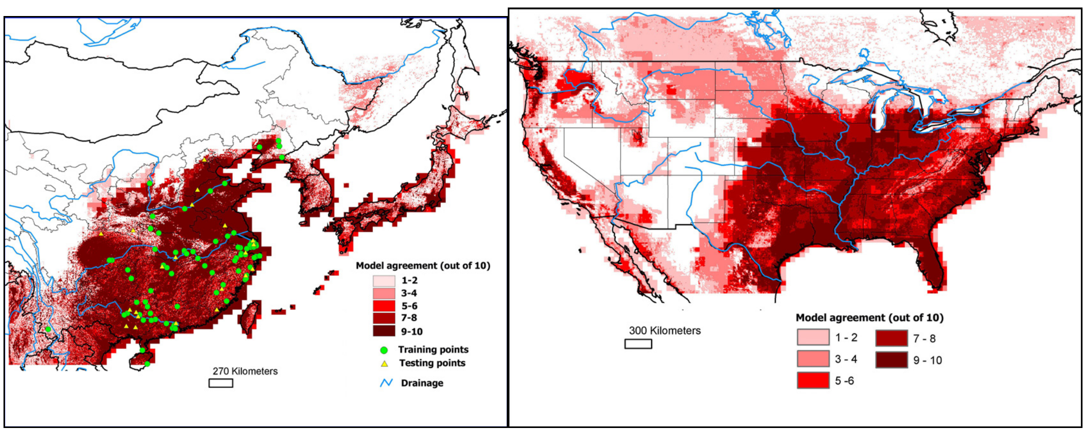
Zhang, J., editor. 1995. The Fishes of Heilongjiang Province. Heilongjiang Science and Technology Press, Jilin, China.

Zhen, Z. 1989. The Fishes of Pearl River. Science Press, Beijing.

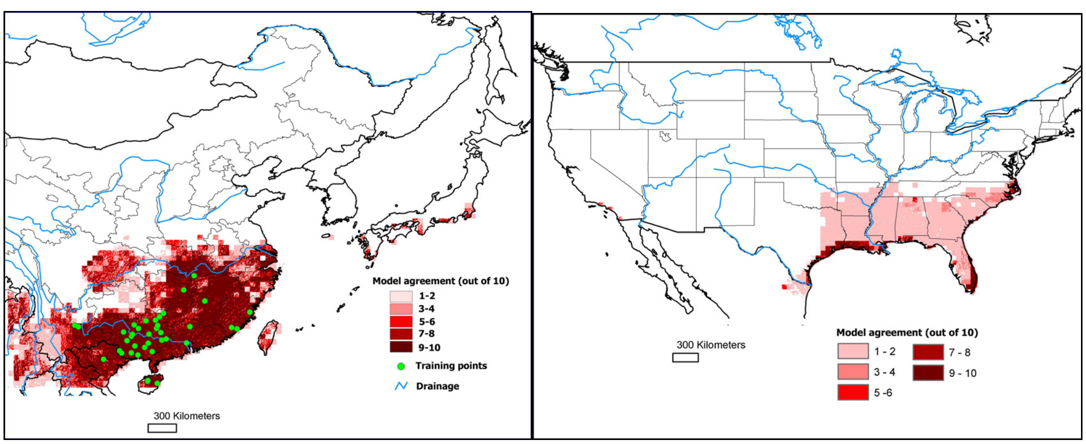
Zweig, M. H., and G. Campbell. 1993. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clinical Chemistry **39**:561-577.



A. ***Abbottina rivularis***

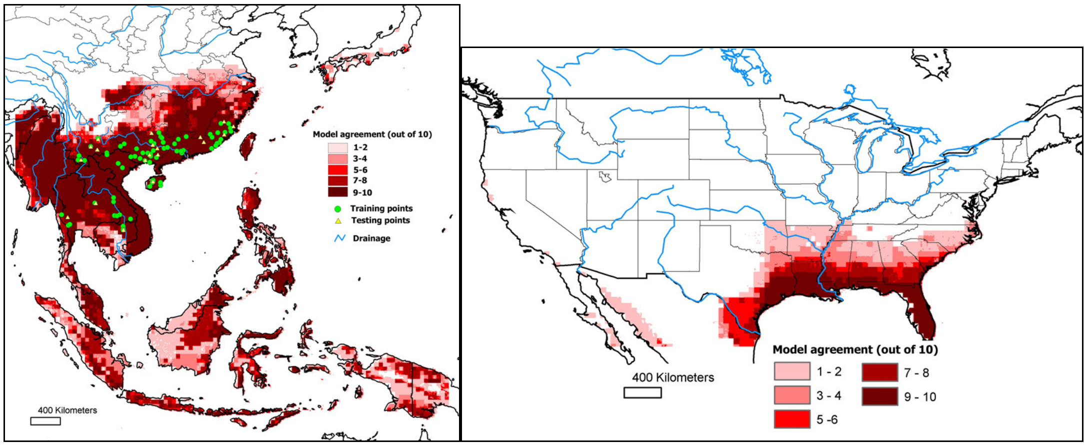


B. ***Anguilla japonica***

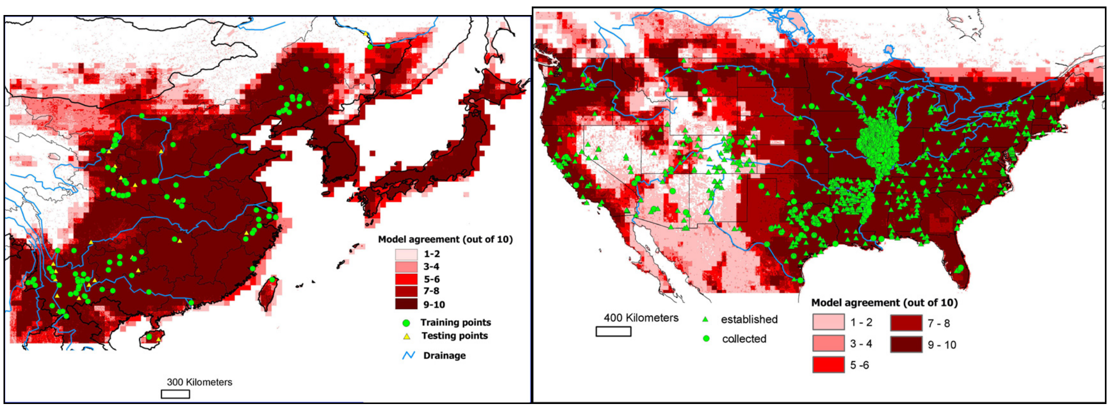


C. ***Channa maculata***

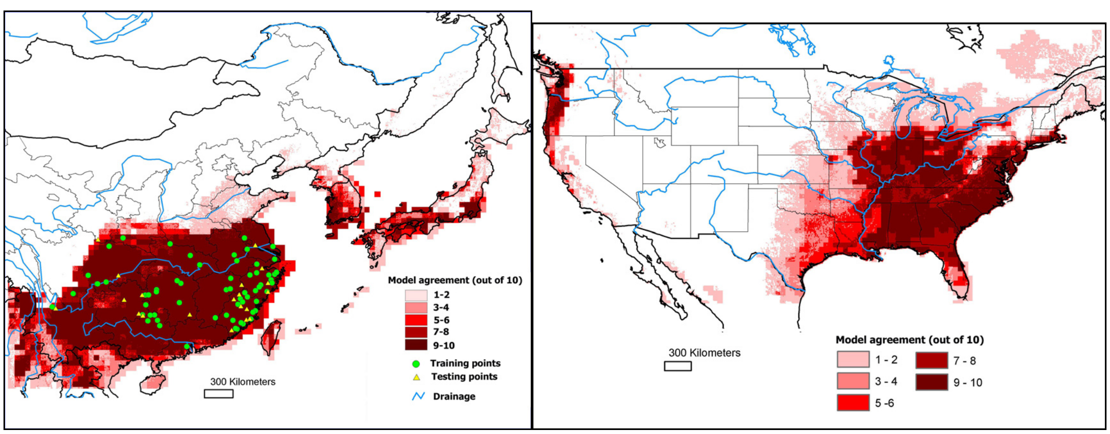
Figure 5-1 (continued)



D. ***Cirrhinus molitorella***

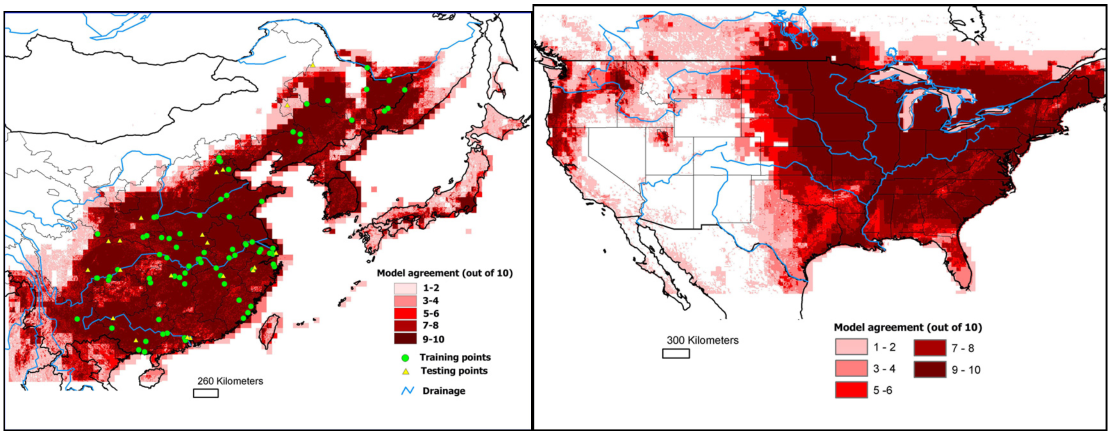


E. ***Cyprinus carpio***

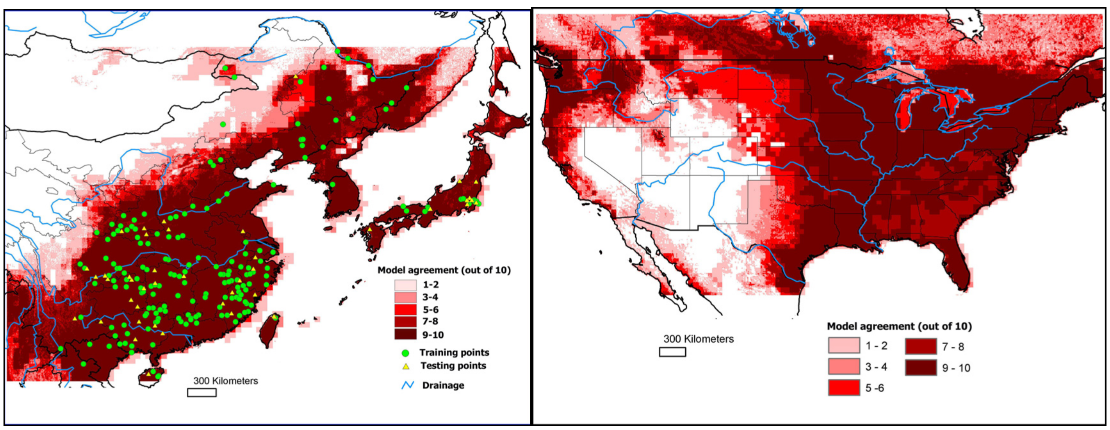


F. ***Distoechodon tumirostris***

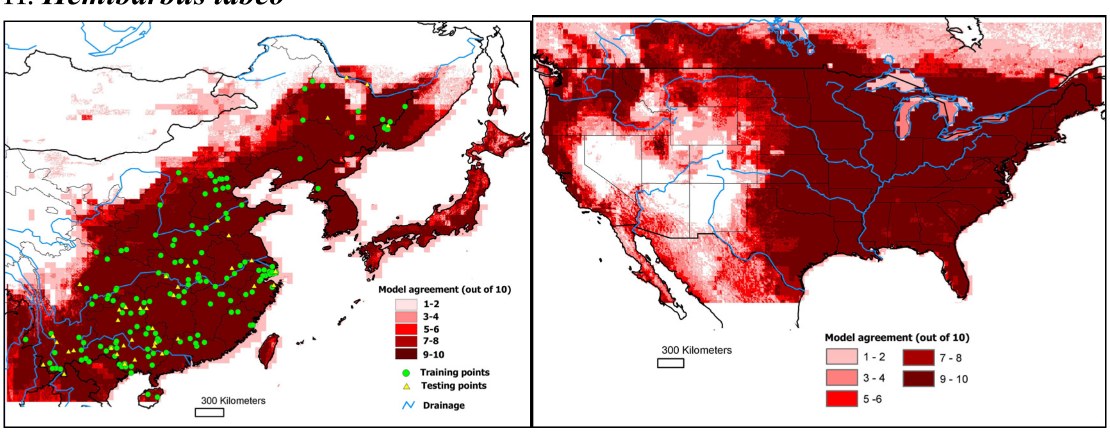
Figure 5-1 (continued)



G. ***Elopichthys bambusa***

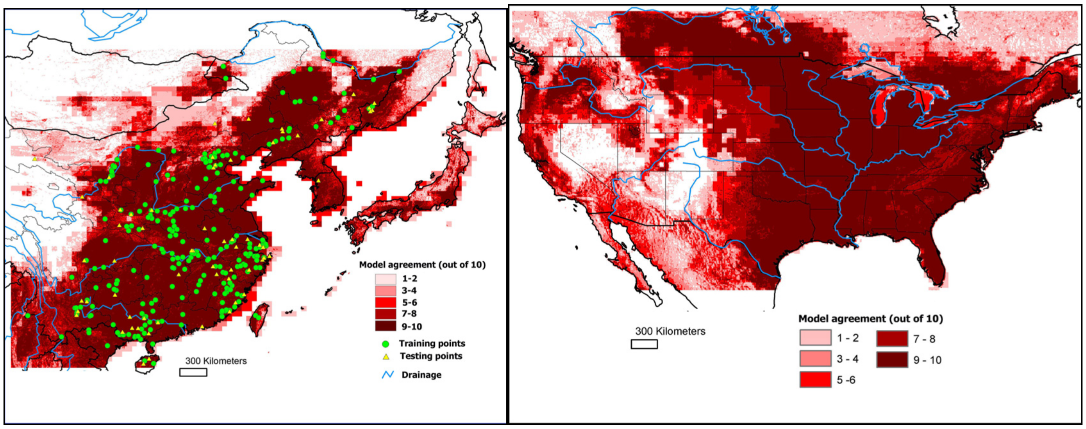


H. ***Hemibarbus labeo***

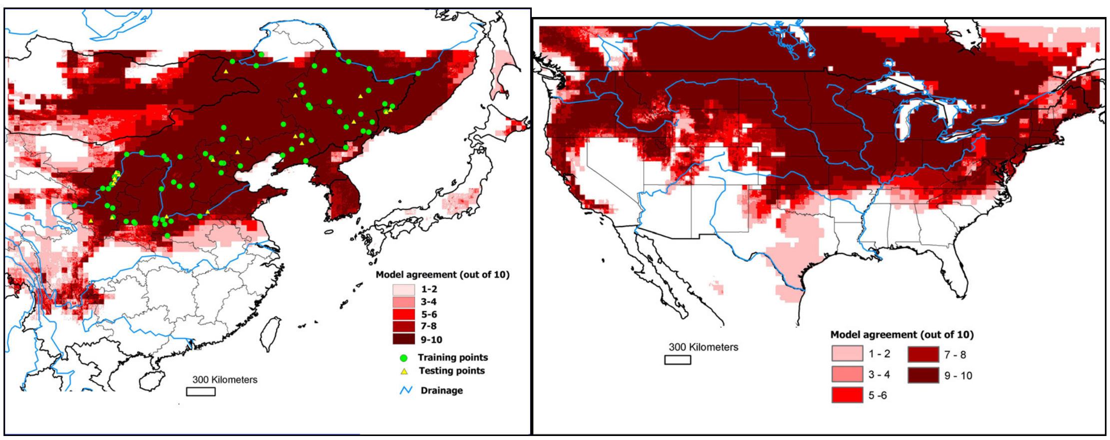


I. ***Hemibarbus maculatus***

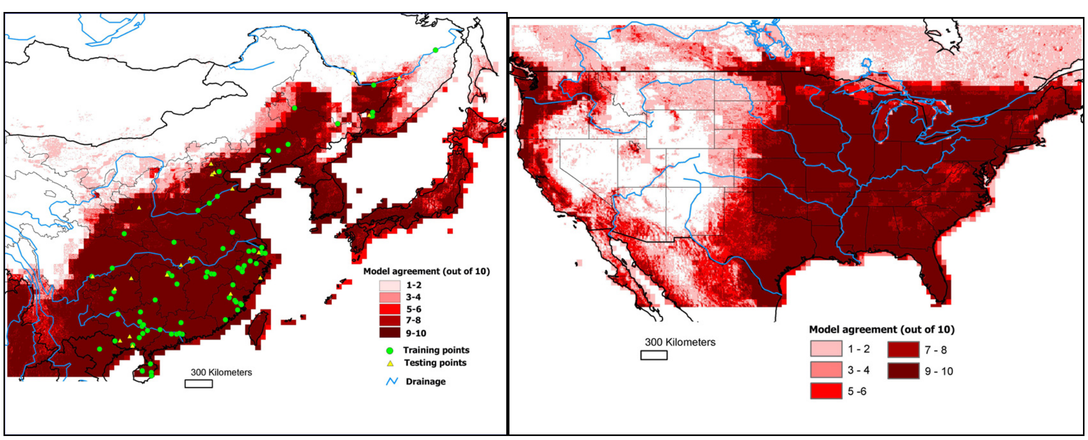
Figure 5-1 (continued)



J. ***Hemiculter leucisculus***

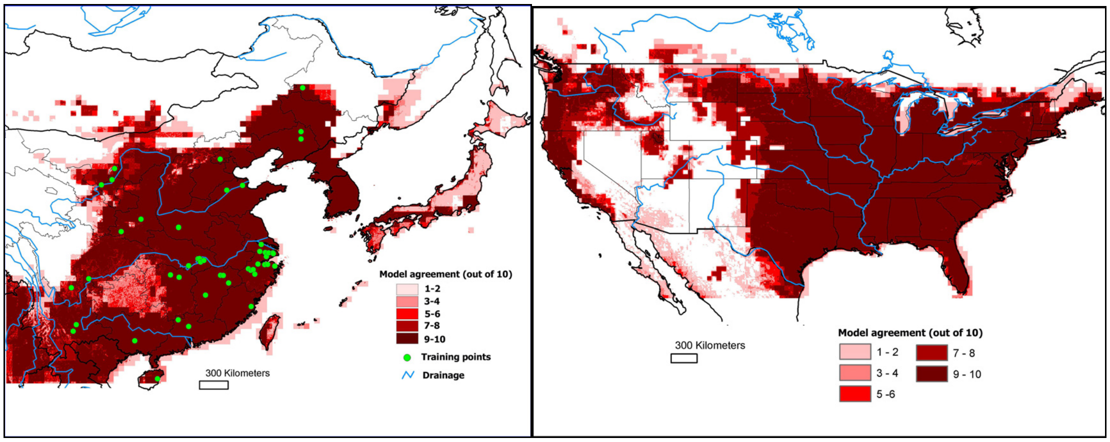


K. ***Leuciscus waleckii***

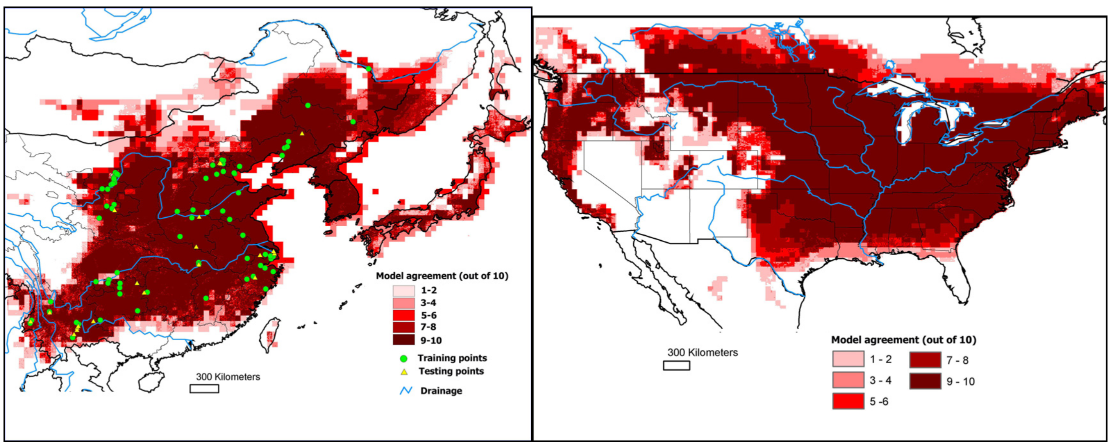


L. ***Megalobrama terminalis***

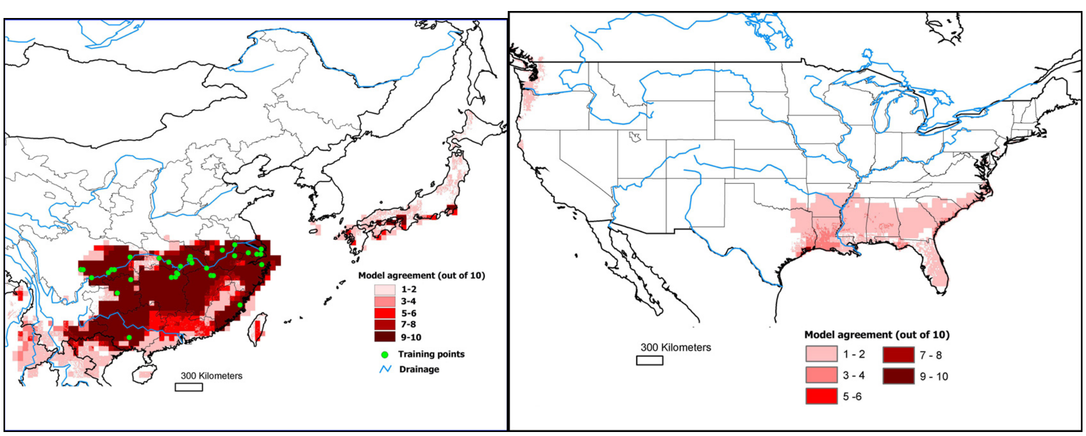
Figure 5-1 (continued)



M. ***Megalobrama amblycephala***

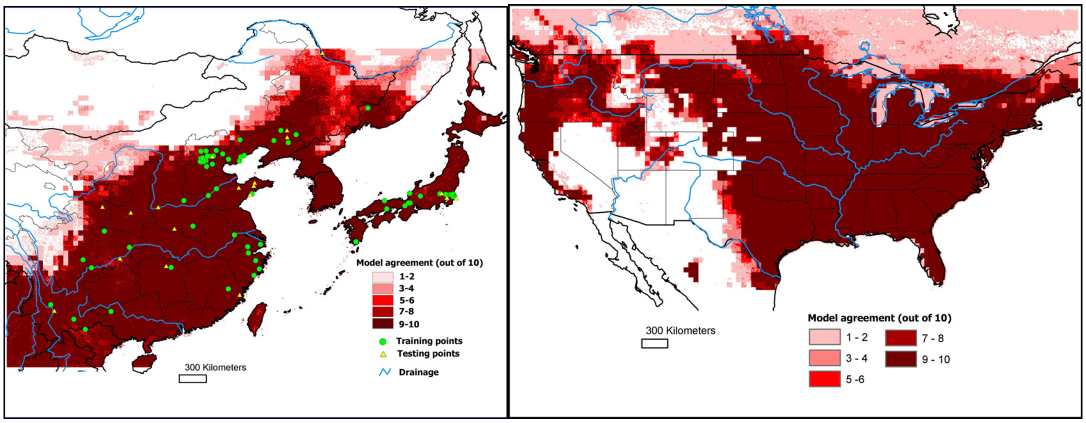


N. ***Micropercops swinohonis***

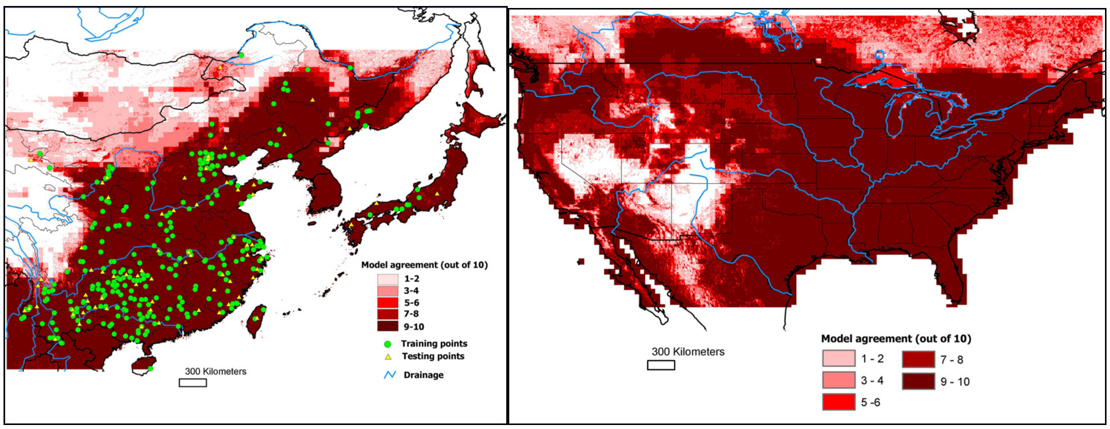


O. ***Myxocyprinus asiaticus***

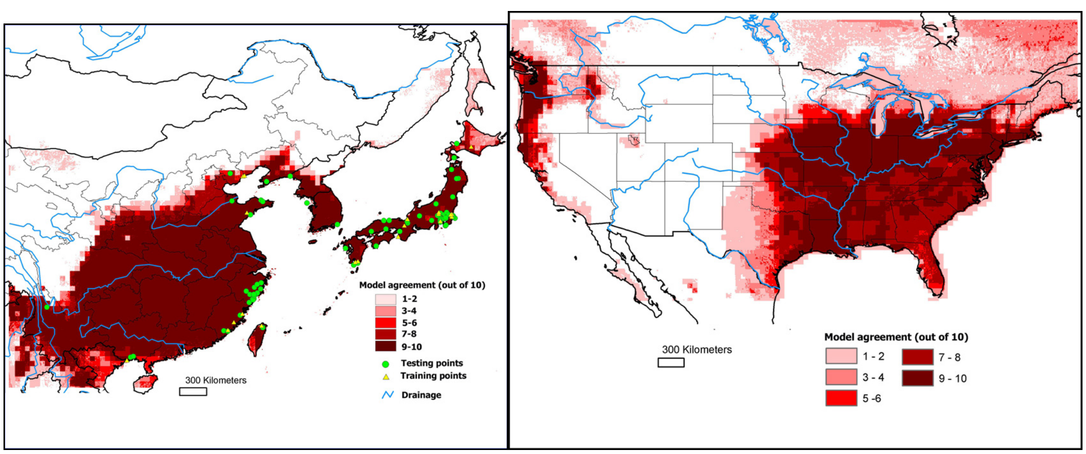
Figure 5-1 (continued)



P. ***Opsariichthys uncirostris***

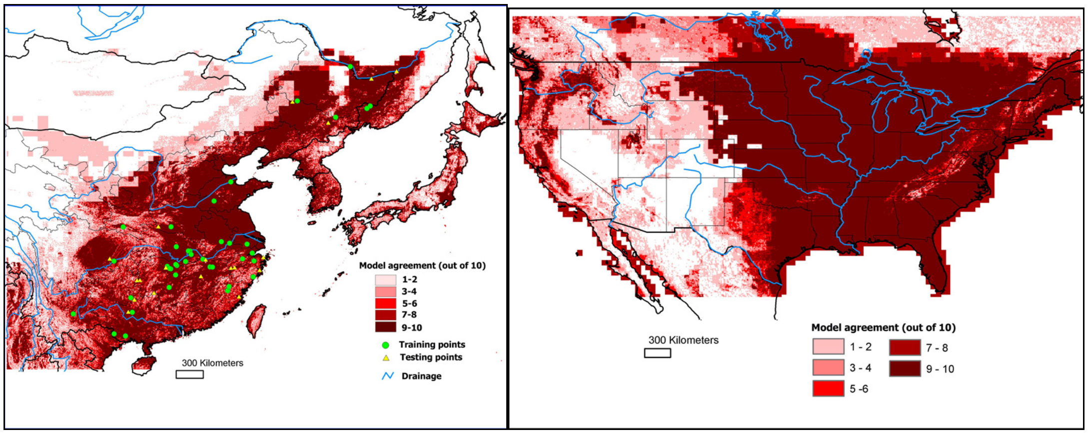


Q. ***Pseudorasbora parva***

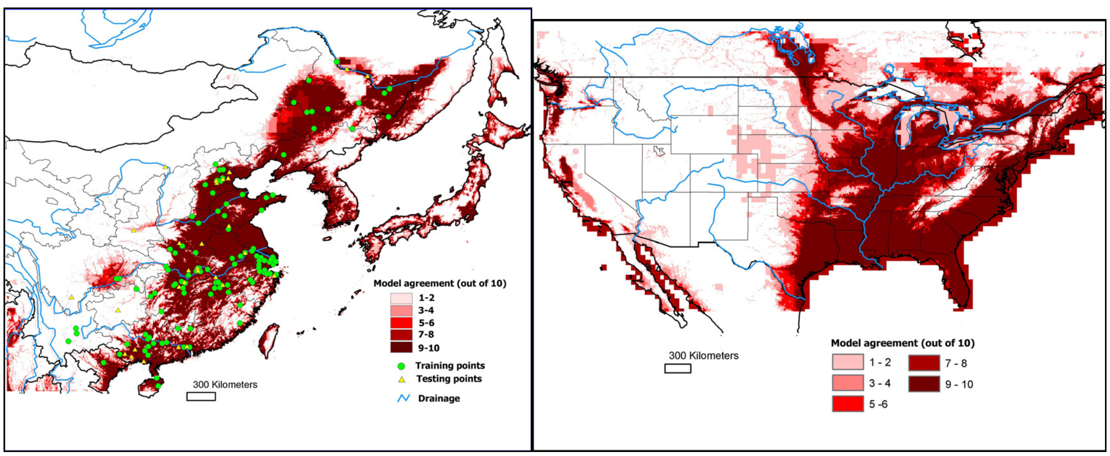


R. ***Plecoglossus altivelis***

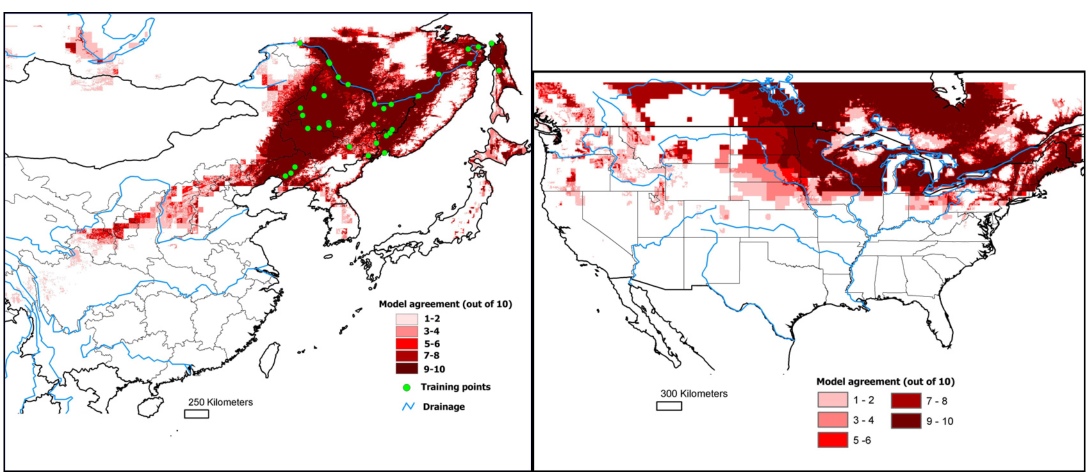
Figure 5-1 (continued)



S. ***Plagiognathops microlepis***

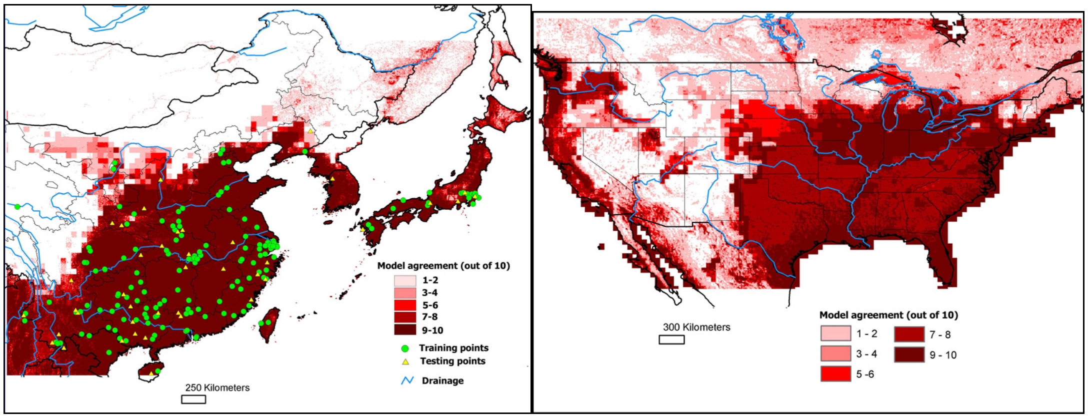


T. ***Parabramis pekinensis***

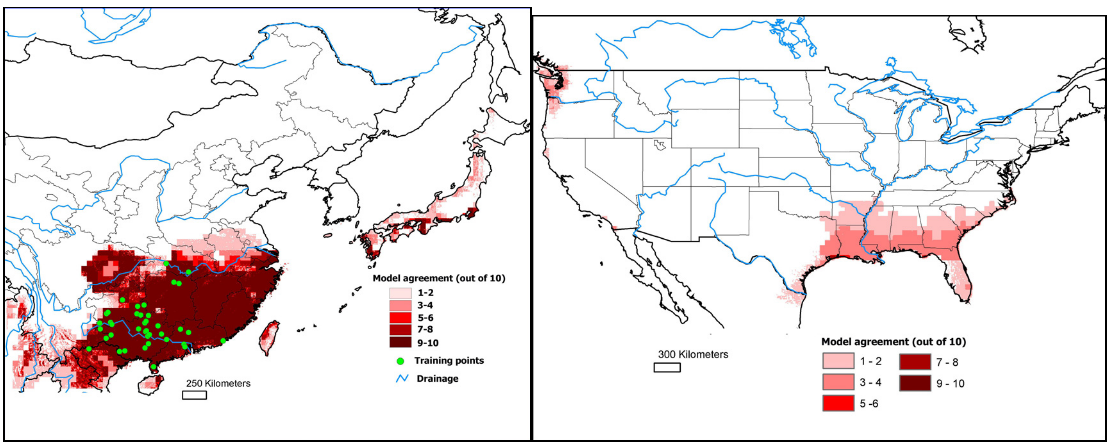


U. ***Perccottus glenii***

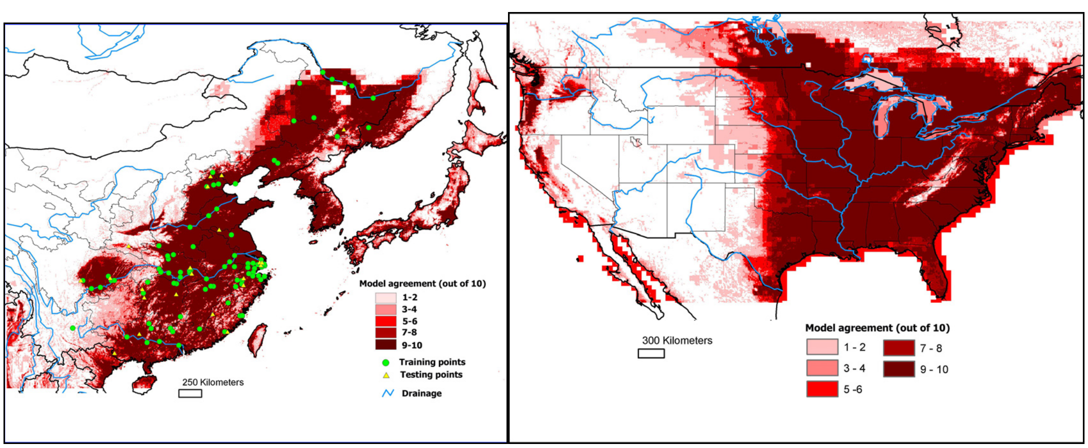
Figure 5-1 (continued)



V. ***Rhodeus ocellatus***

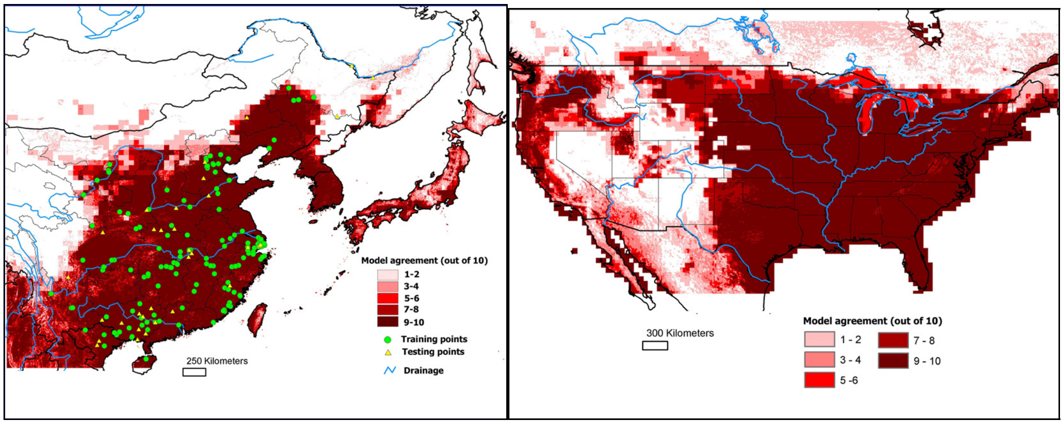


W. ***Sinilabeo decoru***

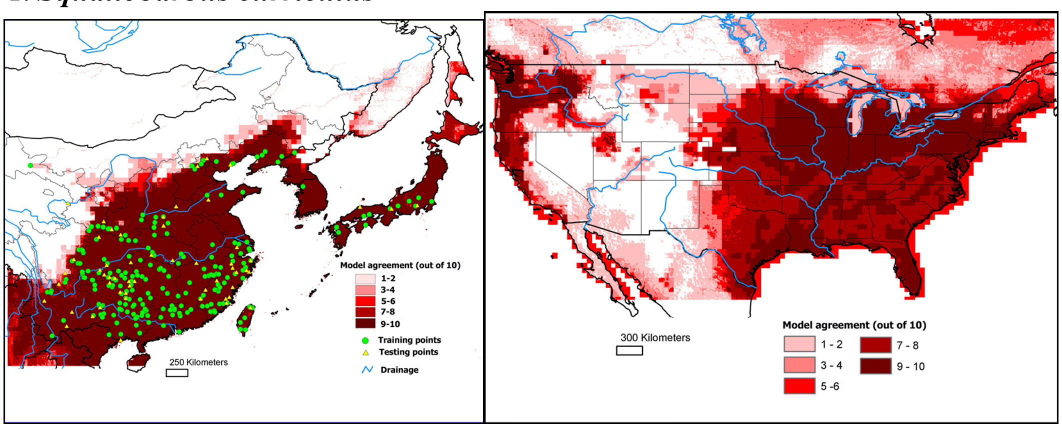


X. ***Siniperca chuatsi***

Figure 5-1 (continued)



Y. ***Squaliobarbus curriculus***



Z. ***Zacco platypus***

Figure 5-1 Niche model predictions over the native landscape (left), showing probable range, and native-range model projected over the conterminous United States (right), showing the potential invasive range of Asiatic invasive fishes. Dark red indicates 9-10 of the 10 best models predicting presence, firebrick 7-8, red 5-6, salmon 3-4, and pink 1-2. Green circles indicate training data used to build models; yellow triangles indicate independent validation data.

Table 5-1: Description of environmental variables used in the modeling. Environmental variables correlated with increased omission error were excluded from building the final niche models, indicated by an "X" for each species.

Variable \ Species	*Abbottina rivularis*	*Anguilla japonica*	*Channa maculata*	*Cirrhinus molitorella*	Cyprinus carpio	*Distoechodon tumirostris*	*Elopichthys bambusa*	*Hemibarbus labeo*	*Hemibarbus maculatus*	*Hemiculter leucisculus*	*Leuciscus waleckii*	*Megalobrama terminalis*	*Megalobrama amblycephala*	*Micropercops swinohonis*	*Myxocyprinus asiaticus*	*Opsariichthys uncirostri*	*Pseudorasbora parva*	*Plecoglossus altivelis*	*Plagiognathops microlepis*	*Parabramis pekinensis*	*Perccottus glenii*	*Rhodeus ocellatus*	*Sinilabeo decorus*	*Siniperca chuatsi*	*Squaliobarbus curriculus*	*Zacco platypus*
Diurnal temperature range																										
Ground frost frequency											X															
Precipitation																										
Solar radiation					X	X													*X*							
Minimum temperature								X																		
Mean temperature																*X*										
Maximum temperature											X					*X*										
Vapor pressure						X																				
Wet day frequency					X																					
elevation		X					X					X				*X*		*X*	*X*							
aspect		X																								
flow accumulation			X								X															
slope																										
topographic index			X			X												*X*								
percent tree cover														X												

Table 5-2. Statistics of model building and evaluation over the native landscape and the conterminous USA. Number of species occurrence points used for building ("training") and validating ("Test") models for each species indicated. There is no independent Test points ("N/A") for those species with less than 50 occurrence points in native landscape. "Maximum Potential distribution" is the percentage of the total area predicted as present by any of the best models. "Average potential distribution" is the average area predicted present by all 10 best models. "Accuracy" is the Percentage of testing data predicted by all 10-best models. AUC -- the area under the curve, which are all significant ($P < 0.01$). SE -- standard error.

Species	Landscape	Training points	Testing points	Maximum commission index	Average commission index	Accuracy	AUC	SE
Abbottina rivularis	Native	243	61	0.7459	0.6120	95.25%	0.7390	0.0367
	USA		N/A	0.8699	0.7146	N/A	N/A	N/A
Anguilla japonica	Native	83	21	0.4514	0.3232	98.10%	0.9047	0.0442
	USA		N/A	0.5878	0.3032	N/A	N/A	N/A
Channa maculata	Native	38	N/A	0.1952	0.1398	N/A	N/A	N/A
	USA		N/A	0.0625	0.0138	N/A	N/A	N/A
Cirrhinus molitorella	Native	74	20	0.3503	0.2410	97.50%	0.9187	0.0423
	USA		N/A	0.1120	0.0557	N/A	N/A	N/A
Cyprinus carpio	Native	107	27	0.7088	0.5523	91.85%	0.7446	0.0548
	USA		646	0.7331	0.5391	82.04%	0.7036	0.0116
Distoechodon tumirostris	Native	66	20	0.3787	0.2685	95.00%	0.8863	0.0487
	USA		N/A	0.3060	0.1515	N/A	N/A	N/A
Elopichthys bambusa	Native	83	20	0.5781	0.4078	86.00%	0.7885	0.0606
	USA		N/A	0.6376	0.4068	N/A	N/A	N/A
Hemibarbus labeo	Native	186	46	0.6857	0.5047	98.91%	0.8182	0.0381
	USA		N/A	0.8053	0.5503	N/A	N/A	N/A
Hemibarbus maculatus	Native	158	39	0.6839	0.5254	97.44%	0.7839	0.0437
	USA		N/A	0.8318	0.5932	N/A	N/A	N/A
Hemiculter leucisculus	Native	253	64	0.7579	0.5435	95.47%	0.7996	0.0334
	USA		N/A	0.8295	0.5826	N/A	N/A	N/A
Leuciscus waleckii	Native	73	20	0.5836	0.4516	99.00%	0.8288	0.0567
	USA		N/A	0.6682	0.5308	N/A	N/A	N/A
Megalobrama terminalis	Native	75	20	0.5948	0.4887	88.50%	0.7528	0.0632
	USA		N/A	0.7004	0.4714	N/A	N/A	N/A
Megalobrama amblycephala	Native	47	N/A	0.5554	0.4363	N/A	N/A	N/A
	USA		N/A	0.5518	0.4335	N/A	N/A	N/A

Table 7-2 (continued)

Species	Landscape	Training points	Testing points	Maximum commission index	Average commission index	Accuracy	AUC	SE
Micropercops swinohonis	Native	77	20	0.6169	0.4543	95.00%	0.8175	0.0579
	USA		N/A	0.6403	0.5120	N/A	N/A	N/A
Myxocyprinus asiaticus	Native	36	N/A	0.2063	0.1394	N/A	N/A	N/A
	USA		N/A	0.0581	0.0099	N/A	N/A	N/A
Opsariichthys uncirostris	Native	68	20	0.6789	0.5417	91.00%	0.7405	0.0656
	USA		N/A	0.7136	0.5174	N/A	N/A	N/A
Pseudorasbora parva	Native	291	73	0.7968	0.6146	94.38%	0.7489	0.0337
	USA		N/A	0.9179	0.7318	N/A	N/A	N/A
Plecoglossus altivelis	Native	68	20	0.4435	0.3710	89.50%	0.7934	0.0602
	USA		N/A	0.5105	0.2790	N/A	N/A	N/A
Plagiognathops microlepis	Native	40	20	0.5561	0.3956	90.50%	0.8144	0.0582
	USA		N/A	0.7540	0.5765	N/A	N/A	N/A
Parabramis pekinensis	Native	122	30	0.3270	0.2545	86.33%	0.8523	0.0439
	USA		N/A	0.4654	0.3036	N/A	N/A	N/A
Perccottus glenii	Native	38	N/A	0.1507	0.1103	N/A	N/A	N/A
	USA		N/A	0.3165	0.2309	N/A	N/A	N/A
Rhodeus ocellatus	Native	159	40	0.5061	0.4215	95.25%	0.8096	0.0415
	USA		N/A	0.6847	0.4329	N/A	N/A	N/A
Sinilabeo decorus	Native	40	N/A	0.2251	0.1641	N/A	N/A	N/A
	USA		N/A	0.0650	0.0145	N/A	N/A	N/A
Siniperca chuatsi	Native	88	22	0.4607	0.3365	95.00%	0.8540	0.0511
	USA		N/A	0.6154	0.4289	N/A	N/A	N/A
Squaliobarbus curriculus	*Native*	141	34	*0.5605*	*0.4473*	92.06%	0.7934	0.0462
	USA		N/A	0.6765	0.5016	N/A	N/A	N/A
Zacco platypus	Native	217	56	0.5117	0.4424	95.89%	0.7940	0.0360
	USA		N/A	0.7359	0.4301	N/A	N/A	N/A

Chapter 6

CONCLUSION AND PROSPECT

Understanding where species occur is fundamental to understanding their biology, and prediction of occurrence is essential for conservation and population management (Rushton et al. 2004). This is particularly the case for invasive species where ecological niche modeling has been applied to provide managers with potential geographic distributions that help make measurements to eradicate them at early stage if possible, or prevent them from spreading further, or to predict sites sensitive to some potential invasions, or to guide site management by manipulating features known to constrain / reduce invasive species; and for endangered species, where knowing what determines distribution is a necessary precursor for actions to mitigate decline or to create new populations through reintroduction (Rushton et al. 2004). Increasingly, ecological niche modeling has been also used in studies investigating the potential impacts of climate change on biodiversity (Peterson et al. 2002, Thomas et al. 2004, Hannah et al. 2005, Thuiller et al. 2005, Araújo et al. 2006).

With the increased availability of remote-sensed data, GIS and statistical packages, it is possible for applied ecologists to use ecological niche modeling to make management actions for conservation in a way that was unprecedented decades of years ago. It is also possible to build and find a (or possibly a small set of) model(s) that is nearer to reality of species distribution.

Accordingly, I have presented a practical application of ecological niche modeling approach, evaluating the invasive potential of Asiatic freshwater fishes in United States. Generally, the United States is at invasion risk from Asiatic fishes. The most sensitive sites to all 33 Asiatic fishes analyzed above in the United States include basins of the Mississippi River and its tributaries, and the eastern United States (Figure 6-1).

Although the potential applications of ecological niche modeling are numerous, there are inherent limitations to our ability to model species' geographic distributions because they are so complex. All model approaches have their limits. These limits also point directions for future development to improve the qualities of ecological niche models. Modeling can never provide a complete substitute for data on species' distribution, demography, abundance, and interaction (Guisan and Thuiller 2005). Efforts should be made to improve species occurrence and environmental data sets within an appropriate scale, and future development should concentrate on developing approaches to overcome the inherent problems in model developing and validation.

Species occurrence data set--One of the major issues in species distribution modeling is collecting data that are of the correct 'range' in both time and space (Vaughan & Ormerod 2003). In an ideal world, the target species would occur at a fixed point in space, and its ecological requirements would be well known and measurable at the same spatio-temporal scales. In reality, these data sets usually fall into two basic types: those that have been collected as part of a survey designed to provide information on target species with the main aim of modeling species–habitat relationship; and those that have been generated as a result of other exercises not specifically associated with distribution modeling. These data are usually collected over long time periods, often using a variety of methods, often in surveys that are not systematic. Historical records, such as herbarium or museum collections, often lack precise details of location: usually they show only proximity to a common site, a valley or village at a scale of a kilometer or more, even a county or a river at a span of ten kilometers or more. These data have been recorded by volunteers, usually without recourse to any predefined sampling strategy. Thus, it is difficulty to apply the usual statistical approaches to analysis their association with environmental variables. This is the case for the occurrence data applied in this study. Absence records, where investigators search but find nothing, are particularly poorly recorded. These issues have severe implications for the success of the modeling effort, irrespective of the approaches used. In many cases, collecting the species occurrence data is more difficult than collecting the associated habitat variables, simply because the target species may move around the landscape. Identifying where the individuals are and what they are using as resources in the landscape still remains a great challenge.

Further, species responses to environment depend on the biotic context. Given the dynamic nature of species' distributions, the effects of natural and human disturbance, and the complicating effects of variation in the speed with which different species re-occupy sites from which they have been displaced, the biotic context is itself dynamic (Barry and Elith 2006). This study didn't consider these factors.

Environmental data set--Measurement of the known potential predictor variables may be difficult and species may have ecological requirements that are unknown or are immeasurable. Modeling data often have frequent limitations: comprehensive, purposeful sampling is rare; conservationists often have to deal with data sets compiled for other purposes; it is difficult to distinguish source or sink habitats; key environmental variables may be undescribed or even unknown.

The collection of habitat data for ecological niche modeling underwent dramatic change in the 1990s when remote-sensed data from satellites became widely available. This, coupled with the increased use of geographical information systems (GIS) to store and manipulate spatial data, led to an expansion in ecological niche modeling. However, as satellite-derived data have a fixed resolution (this may depend on the wavelength of radiation sampled) and may be temporally discordant with much biological recording data. They are not collected specifically for modeling species distributions and can only be used as surrogates for habitat predictors.

Furthermore, at the spatial resolution where biotic processes (for example competition and predation) become important in modeling the species environmental niche, then biotic predictors will be needed. However, such biotic predictors are frequently not available as GIS layers and so cannot be used for spatial prediction. The development of suitable spatial surrogates for such variables is an area that needs more investigation.

Species responses depend on the nature of the environmental predictors and the associated ecological processes. The ideal approach of choosing the appropriate environmental predictors is via a conceptual framework based on known biophysical processes or ecophysical knowledge which allows consideration and selection of appropriate environmental predictors recognizing three types of environmental variables indirect, direct and resource variables. The alternative selection of predictors is based on the availability and experience that the variables show correlations with species distributions and may act as surrogates for more proximal variables (Austin 2007). This study, like most other studies, adopts the latter approach.

Appropriate scale in modeling--Another challenge in ecological niche modeling is in identifying the appropriate scale at which to sample. In the case of non-systematically colleted survey data, the sampling unit is usually some form of grid cell, the size of which is not normally related to any ecological feature of significance to the species concerned (Rushton et al. 2004).

The scale at which data are available can severely restrict the purposes for which the data can be used or place caveats on the usefulness of the results for the intended purpose. Two important aspects of scale are extent and resolution. Extent refers to the area over which a study is carried out, while resolution is the size of the sampling unit at which the data are recorded. For example, if the purpose is to investigate the environmental realized niche of a species then the extent of the study should range beyond the observed environmental limits of the species. If this is not the case, then the species responses are truncated and the actual shape cannot be determined (Austin 2007).

Resolution governs what variables can be measured and what processes can be hypothesized to operate in determining species distribution and abundance. Low resolution (a larger cell size) might result in a more manageable data set if spatial autocorrelation is measured within the species data. Thus, species observations cannot be considered independent. High resolution (a small cell size) might better represent the ecological processes, but the number of species occurrence data could decrease tremendously as fewer occurrences have a high location accuracy associated with grid cells at high resolution. In order to avoid the measurement errors in the model, data of varying spatial accuracy can be manipulated by either (1) aggregating all data in regular grid cells whose size still matches the poorest location accuracy of observed occurrences, or (2) dropping the most inaccurate data (Elith et al. 2002). Many statistical models are constrained by the environmental data in GIS and target species' occurrence data available. Huntley et al. (2004), using the same climatic data and modeling approach for 306 European species representing three major taxa (higher plants, insects and birds), and including species of different life form and from four trophic levels, found model performance was related neither to major taxonomic group nor to trophic level. Their conclusion is only applicable to the data model used. Using only climate predictors at the level of resolution of 50 km x 50 km, only climate effects will be detected. In a study cited by Austin (2002) with resolution 0.1 ha for plants and ca. 10 ha for fauna, plant competition and animal mobility and territories will impact on distribution and interact with climate variables.

Development of ecological niche models--Two key assumptions in statistical modeling are that the data used as predictors are adequate (in the sense that they are true variables determining the species distribution pattern) and that the error structure is appropriate for the data. The first of these assumptions becomes very important if the predictor variables used in modeling are only surrogates for true predictors, as is the case with data derived from remote-sensed imagery. In a logistic regression model the error model can be accepted as appropriate if the residual deviance (unexplained variation) after model fitting is equivalent to the number of degrees of freedom. If the residual deviance is much greater than the degrees of freedom, the data are 'overdispersed'. Overdispersion can arise because there is a structural failure in the model, such as failing to include key predictor variables that are actually driving the response variable, or because the error model is inappropriate for the data (Rushton et al. 2004).

Given the expense of undertaking data collection, many data sets are collected over small areas. In these cases the data sets often show spatial autocorrelation or some other form of nonindependence. Spatial autocorrelation, where the abundance or occurrence of species is correlated with presence and abundance of the species nearby, can affect statistical modeling (Cressie 1993).

With a large suite of predictor variables, it is also possible to 'overfit' to the extent that models often perform very well in the context of the data set used to create them but fail to be robust when used elsewhere. Overfitting obviously has major implications for the applied value of the work, as models only have real utility if they have a general application in areas other than those from which they were created

(Rushton et al. 2004, Peterson et al. 2007).

The probability of a species becoming an invader is very small. Williamson and Fitter (1996) estimated that around 10% (between 5% and 20%) of organisms introduced to a new environment become casual, 10% of these become naturalized and 10% of these naturalized species go on to become pest species. Thus, only 0.1% of the species originally introduced are expected to become invaders. When an event is so rare, it is much harder to forecast which species will become an invader, because the probability of correctly predicting any event is a function, not only of the accuracy of the prediction system, but also of the frequency with which that event occurs at all. This phenomenon is referred to as the "base-rate effect" and has wide applicability for understanding rare events. For example, the ability of a weather forecaster to predict rain with 90% accuracy would sound superficially very impressive. However, if the base-rate probability of rainfall – the average probability of getting rain is as low as 1% of days, then the times that the forecaster make a mistake by identifying dry days as rainy (10% of 99% of days) will overwhelm the very few days (90% of 1%) when they correctly predict a rainy day as rainy. In other words, at such a low base rate for rain, even if forecasters predict rain, we would be far better off ignoring the forecast, unless we had a mortal terror of rain (Matthews 1996). Similarly, because of the rarity of successful invasions compared with the number of imported species, there is a base-rate effect involved in calculating the probability of correctly predicting invasive success (Smith et al. 1999). Smith et al (1999) used the introduction of plant species to Australia as specific examples and a prediction theory analysis of earthquake prediction to explore when people be best advised to ignore the recommendations of a screening system for exotic introductions; they concluded that a pest risk assessment system with an accuracy of 85% would be better ignored, unless that damage caused by introducing a pest is eight times or more of that caused by not introducing a harmless organism that is potentially useful.

Model assessment--The major difficulty with evaluating statistical methods and their compatibility with ecological theory is that the true model is unknown. Comparative evaluations on real data are unsatisfactory because two statistical methods may give different models but both may be half-right. Nevertheless, models have their greatest utility when they can be used to predict and not simply as a means of exploring putative relationships in a data set. It is possible to over-fit models to the extent that they appear to explain variation in the observed data set, but perform poorly when used in other circumstances. One obvious approach is to use thresholds in the predictions, above and below which presence and absence are defined. There are a number of metrics that can be used to compare model predictions and observed data using thresholds. The kappa statistic has been increasingly used in model testing; however, it is sensitive to sample size and fails if one class (the presences or absences) exceeds the other (Fielding & Bell 1997). The other is the use of threshold-independent approaches, such as receiver operator characteristic statistics (ROC plots). These are based on plotting the true positives against the false positive fractions for a range of thresholds in prediction probability. The area under the curve for a ROC plot is taken as a measure of the accuracy of the model that is not dependent on a single threshold.

As the true potential range may differ from the realized range because of dispersal limitation, competition exclusion or other factors (Anderson et al. 2002), evaluating model performance is a complex task and use of observed absences may be misleading (Elith et al. 2006, Peterson et al. 2008).

An ounce of prevention is worth a pound of cure--Most invasions begin with the arrival of a small number of individuals, and the costs of excluding these is usually trivial compared to the cost and effort of later control after populations have grown and have been established.

There are international treaties, such as "the agreement on the Application of Sanitary and Phytosanitary Measures" (SPS), to restrict the movement of biotic invaders among the members of the World Trade Organization (WTO). However, nations always give variances or exceptions based on politico-economic considerations that outweigh biological concerns. Even if a nation attempts to ban importation of a species, its efforts may have to go to the international judgment if the WTO, in its regulatory capacity, rules that the ban is an unlawful or protectionist trade barrier rather than a legitimate attempt to exclude pest (Jenkins 1996). Risk assessments that would estimate the invasive potential of a species proposed for import is suggested (Ruesink et al. 1995). The low base rate at which species become naturalized as well as the low base rate of becoming invaders means that the predictive power of any risk assessment must be very high to identify invaders reliably (Smith et al. 1999). Thus, risk assessment tools are likely to produce high rates of false positives that would not have become invasive. It will remain a great challenge for scientists to identify the few potentially harmful invaders among the potential nonindigenous species.

LITERATURE CITED

Anderson, R. P., M. Laverde, and A. T. Peterson. 2002. Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. Oikos **93**:3-6.

Araújo, M. B., W. Thuiller, and R. G. Pearson. 2006. Climate warming and the decline of amphibian and reptiles in Europe. Journal of Biogeography **33**:1712–1728.

Austin, M. P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecological Modelling **157**:101-118.

Austin, M. P. 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. Ecological Modelling **200**:1-19.

Barry, S., and J. Elith. 2006. Error and uncertainty in habitat models. Journal of Applied Ecology **43**:413-423.

Cressie, N. A. C. 1993. Statistics for Spatial Data. Wiley, New York.

Elith, J., M. A. Burgman, and H. M. Regan. 2002. Mapping epistemic uncertainties and vague concepts in predictions of species distribution. Ecological Modelling **157**:313-330.

Elith, J., C. H. Graham, R. P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A.

Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. M. Overton, A. T. Peterson, S. J. Phillips, K. Richardson, R. Scachetti-Pereira, R. E. Schapire, J. Soberon, S. Williams, M. S. Wisz, and N. E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography **29**:129-151.

Guisan, A., and W. Thuiller. 2005. Predicting species distributions: offering more than simple habitat models. Ecology Letters **8**:993-1009.

Hannah, L., G. F. Midgley, G. Hughes, and B. Bomhard. 2005. The view from the cape: extinction risk, protected areas, and climate change. Bioscience **55**:231-242.

Huntley, B., R. E. Green, Y. C. Collingham, J. K. Hill, S. G. Willis, P. J. Bartlein, W. Cramer, W. J. M. Hagemeijer, and C. J. Thomas. 2004. The performance of models relating species geographical distributions to climate is independent of trophic level. Ecology Letters **7**:417-426.

Jenkins, P. 1996. Free trade and exotic species introductions. Pages 145-147 *in* O. T. Sandlund, P. J. Schei, and A. Viken, editors. Proceedings of Norway / UN Conference on Alien Species. Directorate for Nature Management and Norwegian Institute for Nature Research, Trondheim, Norway.

Matthews, R. A. 1996. Base-rate errors and rain forecasts. Nature **382**:766.

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

Peterson, A. T., M. Papes, and M. Eaton. 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. Ecography **30**:550-560.

Peterson, A. T., M. Papes, and J. Soberón. 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. Ecological Modelling **213**:63-72.

Ruesink, J. L., I. M. Parker, M. J. Groom, and P. M. Kareiva. 1995. Reducing the risk of nonindigenous species introductions--guilty until proven innocent. Bioscience **45**:465-477.

Rushton, S. P., S. J. Ormerod, and G. Kerby. 2004. New paradigms for modelling species distributions? Journal of Applied Ecology **41**:193-200.

Smith, C. S., W. M. Lonsdale, and J. Fortune. 1999. When to ignore advice: invasion predictions and decision theory. Biological Invasions **1**:89-96.

Thomas, C. D., A. Cameron, R. E. Green, M. Bakkenes, L. J. Beaumont, Y. C. Collingham, B. F. N. Erasmus, M. F. d. Siqueira, A. Grainger, L. Hannah, L. Hughes, B. Huntley, A. S. v. Jaarsveld, G. F. Midgley, L. Miles, M. A. Ortega-Huerta, A. T. Peterson, O. L. Phillips, and S. E. Williams. 2004. Extinction risk from climate change. Nature **427**:145-148.

Thuiller, W., S. Lavorel, M. B. Araújo, M. T. Sykes, and I. C. Prentice. 2005. Climate change threats to plant diversity in Europe. Proceedings of the National Academy of Sciences USA **102**:84245-88250.

Williamson, M., and A. Fitter. 1996. The varying success of invaders. Ecology **77**:1661-1665.

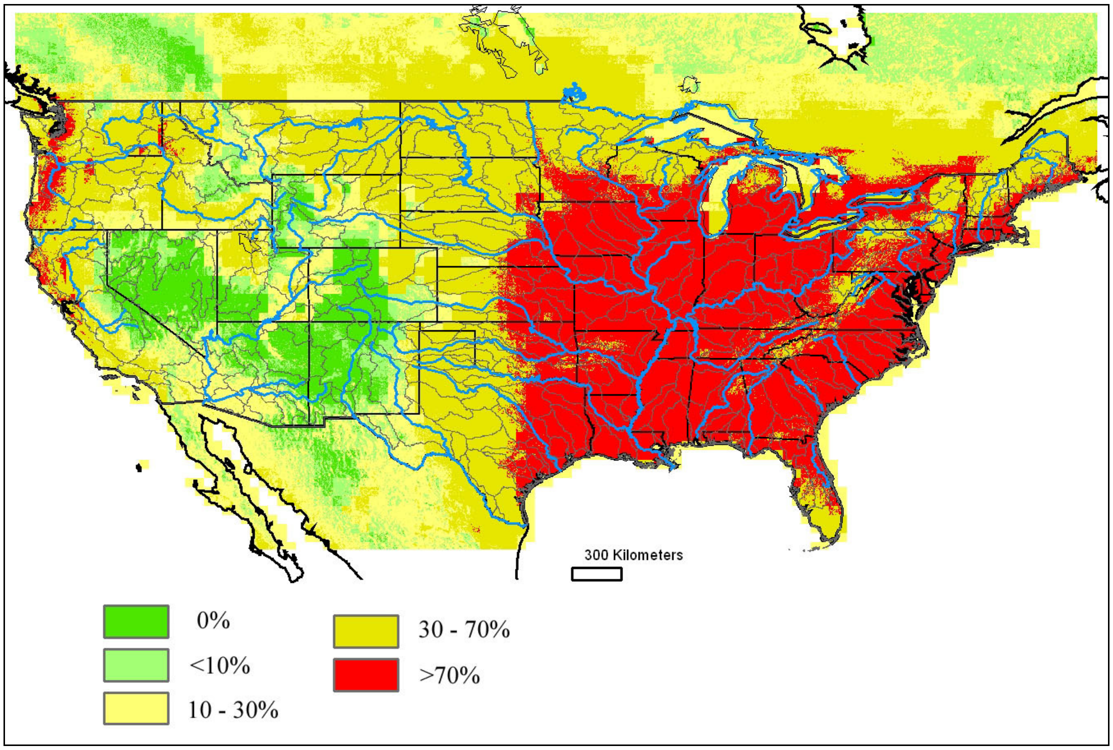


Figure 6-1 Map of sensitive sites to 33 Asiatic potential invasive freshwater fishes in North America. Red indicates ≥70% of 33 species predicted suitable by all 10 best models, yellow 30%-70%, light yellow 10-30%, and green <10%.