Managing Vertebrate Invasive Species   USDA National Wildlife Research Center Symposia

8-1-2007

INVASIVE SPECIES MANAGEMENT AND RESEARCH USING GIS

Tracy Holcombe
Fort Collins Science Center, U.S. Geological Survey, Fort Collins, Colorado, USA

Thomas J. Stohlgren
Fort Collins Science Center, U.S. Geological Survey, Fort Collins, Colorado, USA

Catherine Jarnevich
Fort Collins Science Center, U.S. Geological Survey, Fort Collins, Colorado, USA

INVASIVE SPECIES MANAGEMENT AND RESEARCH USING GIS

TRACY HOLCOMBE, THOMAS J. STOHLGREN, AND CATHERINE JARNEVICH, Fort Collins Science Center, U.S. Geological Survey, Fort Collins, Colorado, USA

Abstract: Geographical Information Systems (GIS) are powerful tools in the field of invasive species management. GIS can be used to create potential distribution maps for all manner of taxa, including plants, animals, and diseases. GIS also performs well in the early detection and rapid assessment of invasive species. Here, we used GIS applications to investigate species richness and invasion patterns in fish in the United States (US) at the 6-digit Hydrologic Unit Code (HUC) level. We also created maps of potential spread of the cane toad (*Bufo marinus*) in the southeastern US at the 8-digit HUC level using regression and environmental envelope techniques. Equipped with this potential map, resource managers can target their field surveys to areas most vulnerable to invasion. Advances in GIS technology, maps, data, and many of these techniques can be found on websites such as the National Institute of Invasive Species Science (www.NIISS.org). Such websites provide a forum for data sharing and analysis that is an invaluable service to the invasive species community.

Key Words: buffer, early detection, environmental envelope, geographic information systems, GIS, invasive species, regression, thiessen polygons.

INTRODUCTION

Geographic Information Systems (GIS) and Global Positioning Systems (GPS) provide a mechanism to digitally pinpoint a location on earth, view the location on a map, and use the location and ancillary data in spatial analyses. Individuals are able to quickly and easily produce maps and conduct spatial analyses that would otherwise be difficult or possible to produce using a network of satellites, satellite receivers, and mapping software. GIS serves as a data storage and analysis device for spatial data, making data easy to view and manipulate.

Health care, agriculture, and environmental industries are a few of many entities that have been positively influenced with the advent of GIS technologies. Large spatial databases can help companies track their hard goods and allow farmers to determine which areas of their fields need more fertilizer, eliminating the need to add fertilizer to the entire field. Ecological data often contains spatial component. Where an animal spends its time and the patterns of its movements can be important clues to its biology. Biological information of non-native species provides insight to explaining expanding distributions and provides a watch list of spreading invasive species to managers for early detection and rapid response.

GIS can be a useful tool for monitoring invasive vertebrates, especially for early detection and rapid assessment. Species distributions are largely determined by the environment. A growing number of statistical models, called Species Environmental Matching (SEM) models are being used to determine current and potential distributions and abundances of harmful invasive species (Stohlgren and Schnase 2006). SEM models relate observed species distributions to environmental (climatic, topographic, edaphic) envelopes. Then, assuming the same stable relationships, they project species spatial shifts (local, enrichment, or extinction) in response to envelope changes under current conditions. These environmental envelopes, arranged along a gradient from proximal to distal predictors, may have direct or indirect effects on species' establishment and survival (Austin 2002). SEM models are either created in a GIS or can be displayed in GIS to give a visual representation of the environmental envelope and potential habitat or abundance.

An important consideration for invasive species management is that recent invaders may not have filled all suitable habitats, while species naturalized long ago may have filled a larger proportion of suitable habitat. Defining where a species may survive depends heavily on being able to determine

its existing or potential habitat. Technological advances in GIS software make these types of analyses more readily accessible to the general public. Data such as elevation, vegetation type, and climate information are now available for free on the internet, often paired with websites that allow persons to view and utilize the information.

HOW GIS CAN BE USED
View Data
A very basic and effective way to use GIS is to view data. Many datasets are very large and difficult to visualize as a table of numbers. When viewed spatially, these data often make more sense. Stohlgren et al. (2006) combined native and non-native fish datasets from NatureServe and the United States Geological Survey (USGS) Florida Integrated Science Center's Non-Indigenous Aquatic Species database to examine numbers of native and non-native species in each 6-digit HUC area (Figure 1). Without performing any statistical analyses, they found a large majority of native fish in the US are centered throughout the mid-west and south-central US. The non-native fish are found primarily in the western and eastern US. These patterns were ascertained without conducting any statistical analysis, proving that displaying the data in a spatial format can be a useful endeavor even without detailed analyses.

Data Summary
GIS can be helpful for summarizing large datasets for modeling habitat quality and distribution. Data layers, such as Digital Elevation Models (DEM), are often used in modeling because they provide a large amount of environmental information. DEMs are available free via the internet, often at either 10- or 30-meter resolutions.

While these resolutions provide a lot of information at a fine scale, this scale is not always necessary or desired. In certain situations, GIS can be used to summarize data, often simplifying the data into the resolution of interest. The Spatial Analyst module of ArcGIS 9 (ESRI 2004) has a zonal statistics function that calculates raster layer summary statistics for a large polygonal area like a county. This module will extract the average, minimum or maximum, and range for each polygon. Additionally, GIS can be used to extract the value for a specific point in a DEM so that the entire surface does not have to be stored. GIS functions like these make the retrieval of dependant data for models readily accessible.

Field Data - Points, Lines, and Polygons
Spatial field data can be displayed and managed in a GIS. The data are stored in one of three formats: points, lines, or polygons. Locations of individual organisms are examples of points. These are discrete one-dimensional places in space. A linear representation of interest is a line. Lines include rivers, transects, or roads. Polygons represent an area of interest, like a stand of trees or a lake.

Point data types are an excellent medium for recording presence or absence of a species because they are discreet. Lines give similar amounts of information, again lending themselves well to presence, absence, and additional attributes. Polygons are unique because they cover an area which can contain additional information such as abundance or percent cover. All these data types can be collected in the field using either paper maps or GPS devices to collect the data before downloading it to a computer.
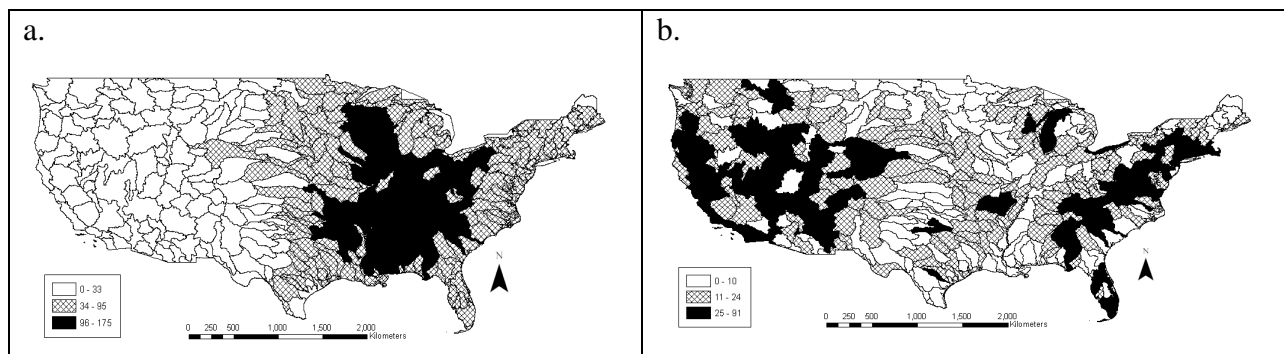
Figure 1. Pattterns of (a.) native and (b.) non-native fish by 6-digit HUC drainage (Stohlgren et al. 2006)

Simple GIS models

GIS can be used to create simple analyses such as buffers and thiessen polygons. A buffer can be created around points, lines, or polygons. It is a new polygon of specified distance from the original feature. Any GIS program can create a buffer around points, lines, or polygons that can be used for various reasons, such as surrogates for habitat for poorly studied species. Buffers can also define potential habitat for species that have a very specific distance they can be from a given feature, such as water. Buffers are a commonly used transformation of spatial data.

Another analysis performed by GIS is the creation of thiessen polygons, sometimes known as voronoi polygons. Theissen polygons are created around a group of points, one polygon for each point. The polygons are created around each point in such a way that every location lies within the polygon of the point to which it is nearest. The easiest way to think about thiessen polygons is with fast food delivery areas. A fast food pizza chain would divide a city into thiessen polygons, only delivering to customers that were closer to them than they were to the next restaurant. A wildlife example would be, if there are twelve nests in an area, polygons are formed around those twelve nests so that every place on the landscape falls into the polygon associated with the closest nest. This tool has many applications for studying territorial animals. Nest location data could be used to generate thiessen polygons surrounding each nest. The area surrounding each nest could be an estimate of territory range. Buffers and thiessen polygons are two of the many possible examples of simple operations that can be done using a GIS.

Statistical models

Statistical models use current species distribution data to try and predict potential habitat. Conceptually, the SEM models assume the fitted observational relationships to be an adequate representation of the realized niche of a species under a stable equilibrium or quasi-equilibrium constraint. As such, the SEM model result is only a first approximation of future distributions of individual species (Pearson and Dawson 2003). SEM model results are also determined by other processes such as dispersal, adaptation, competition, succession, fire and grazing pressure (Austin 2002). Still, an integrated model may contribute considerably to a robust early warning system for decision makers to design more effective management and control strategies for harmful invasive species. In short, we will be able to better manage and assess risks associated with harmful invasive species because risk assessments require accurate modeling of current and potential species distributions (Stohlgren and Schnase 2006).

Numerous challenges exist in traditional SEM or niche-based modeling for current and future species distributions (see reviews by Pearson and Dawson 2003, Soberon and Peterson 2005, Elith et al. 2006, Guisan et al. 2006, Heikkinen et al. 2006, Hijmans and Graham 2006, Peterson 2006, Beaumont et al. 2007). These challenges have not prevented scientists and resource managers from refining, testing, and using SEM models in their work. No two SEM models are identical, and each has advantages and disadvantages (Table 1).

Regression Models

Logistic regression is a type of Generalized Linear Model (GLM) appropriate for data with a binary distribution such as species presence or absence (McCullagh and Nelder 1989). The output from logistic regression models can be taken from statistical software and used in GIS to create a visual representation of the model created. We have done this with data obtained from the USGS Florida Integrated Science Center's Non-Indigenous Aquatic Species database on the invasive cane toad (*Bufo marinus*). The cane toad has become established in the US and invaded several watersheds in Florida. We employed logistic regression with Systat 11.0 (SSI 2004) using minimum temperature, minimum radiation, mean temperature, maximum temperature, maximum humidity, and maximum growing degree days as predictor variables to determine how much potential habitat exists for the cane toad in the south-eastern US. We constructed a step-wise GLM, and only minimum temperature was selected as a significant variable. Results of the regression analysis had a high predictive power (McFadden's Rho Squared = 0.92). When the results were imputed in a GIS, the map showed that the cane toad had invaded most of its suitable habitat in the Florida area, with only a few un-invaded areas left in high and medium habitat suitability areas (Figure 2). This result was a first approximation model. More data and ecological information on the cane toad could produce better results in the future.

The Environmental Envelope Model

The Environmental Envelope Model (EEM, Jarnevich et al. 2007) was developed as a rapid assessment technique to estimate the potential

**Table 1.** Commonly used species environmental matching models for predicting species distributions.

Model	Citation	Advantages	Disadvantages
Maxent	(Phillips et al. 2006)	Presence only, nonlinear, nonparametric, not sensitive to multicollinearity, provides variables' relative importance (jackknifing), easy to run and takes less time, becoming popular	Presence only (no consideration of absence data)
Classification and Regression Tree (CART)	(Breiman et al. 1984)	Non-parametric, Presence/absence, easy to run and interpret	Absence data needed
Boosted Regression Tree	(Friedman 2001, De'ath 2007)	Non-parametric, Presence/absence, limitations with spatial data	Absence data needed, more statistical details
Logistic Regression	(McCullagh and Nelder 1989)	Widely used, presence/absence	Absence data needed, sensitive to multicollinearity
Least square regression	Most statistics software	Widely used, continuous response variable (e.g., species richness)	Needs continuous response variable, sensitive to multicollinearity, decision about significance level (P value?)
BIOCLIM	(Busby 1991)	Presence only, simple	Presence only, does not use absences, less accurate than other niche models
DOMAIN	(Carpenter et al. 1993)	Presence only, simple	Presence only, does not use absences, less accurate than other niche models
ENFA (Env. Niche Factor Analysis)	(Hirzel et al. 2002)	Presence only	Presence only, does not use absences
Envelope	(Jarnevich et al. 2007)	Presence only or absence only models can be run.	All environmental factors are given equal weighting.

distribution of a species given its present location and associated environmental attributes. It is supported by ArcGIS 9x (ESRI 2004) and will be available on the National Institute of Invasive Species Science (NIISS) website (www.NIISS.org). Envelope models use environmental variables, chosen by the modeler that are relevant to the species of interest or species growth in general, to determine locations within the environmental envelope where the species of interest may be able to become established. The minimum and maximum of each independent variable are noted
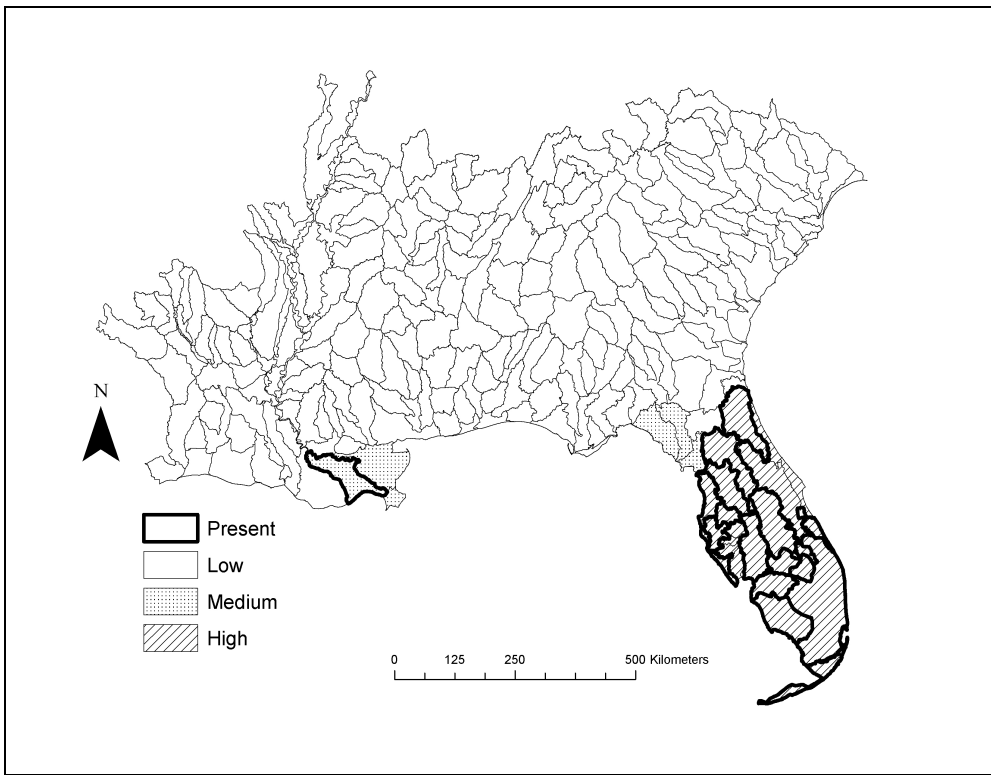
Figure 2. Regression model of the cane toad showing low, medium, and high likelihood of suitable habitat in each 6-digit HUC.
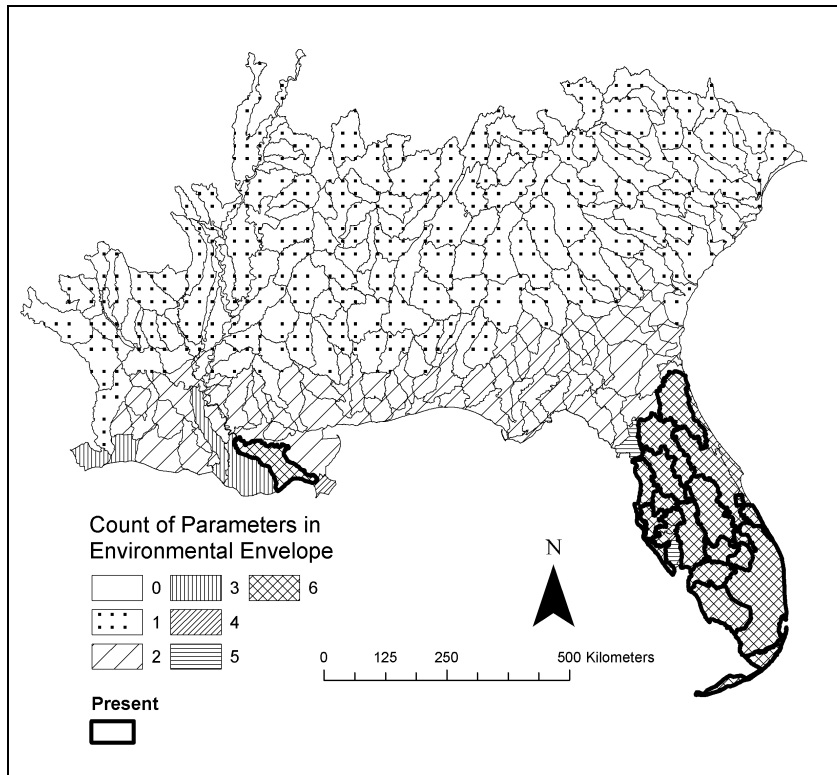
Figure 3. Envelope model of the cane toad showing the number of parameters in each 6-digit HUC that could contain the species.

by the ArcGIS program for all of the locations that the species is present. These minimum and maximum values together become the "envelope" in which the species can survive. For instance, if a species exists in only three counties and the temperature in county A is 45° F, county B is 40° F, and county C is 43° F, then the temperature envelope is 40 to 45° F. We would then compare the temperature for other counties to see if they fell within the range of potential habitats. The model can include several different environmental layers to determine suitable habitats. The output of the model informs how many of the input variables lie within the environmental envelope of the species.

We conducted an EEM analysis on the cane toad to compare results generated from the regression model (Figure 3). We used environmental data retrieved from the Daymet website (www.daymet.org) that was originally at 1 km^2 resolution for the dependant variables. We used zonal statistics in ArcGIS's Spatial Analyst (ESRI 2004) to summarize the data for the 6-digit HUCs. Variables included minimum radiation, minimum temperature, mean temperature, maximum temperature, maximum humidity, and growing degree days. We used the same cane toad data from the regression model. The resulting map showed that as distance increases from the peninsula of Florida, there are less environmental variables that fall within the cane toad's environmental envelope. This trend supports the regression model that showed the cane toad did not have much more suitable habitat than what is already occupied.

GIS on the Web
Common issues confronting GIS users today include software and data availability and user friendliness. GIS software is often expensive, making it difficult for many people to obtain. Another subset of would-be GIS users have access to software, but do not have the time required to learn to effectively and efficiently use the software. These issues are changing with the advances in GIS technology. Many of the functions that are found in proprietary software can also be found on the internet. Much of the species distribution data used in the examples in this paper were found and downloaded from the internet. Many websites, such as NIISS are encouraging an environment of data sharing. The NIISS website includes an interface to upload data and a GIS interface to view data graphically, create models, and print and save final map products. The technology is very

sophisticated and is open to the general public. This may be the direction GIS software is heading toward, reducing dependence on desktop GIS software in the future.

CONCLUSION
Advances in GIS technology have made it become a useful tool for land managers and academics alike. It is widely used as a tool to perform basic functions such as displaying data and more complex functions like creating and displaying SEM models. As we look to our computers today and continue to look to the future of GIS technologies, GIS is a tool that should, and could be used by many scientists and resource managers alike.

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

BEAUMONT, L. J., A. J. PITMAN, M. PAULSEN, AND L. HUGHES. 2007. Where will species go? Incorporating new advances in climate modeling into projections of species distributions. In G. C. Biology, editor. doi.10.1111/j.1365-2486.2007.01357.x.

BREIMAN, L., J. H. FRIEDMAN, R. A. OLSHEN, AND C. G. STONE. 1984. Classification and regression trees. Wadsworth International Group, Belmont, California, USA.

BUSBY, J. R. 1991. BIOCLIM - A bioclimate analysis and prediction system. Pages 64-68 in C. R. M. a. M. P. Austin, editor. Nature conservation: cost effective biological surveys and data analysis. CSIRO, Melbourne.

CARPENTER, G., A. N. GILLISON, AND J. WINTER. 1993. Domain - a flexible modeling procedure for mapping potential distributions of plants and animals. Biodiversity and Conservation 2:667-680.

DE'ATH, G. 2007. Boosted trees for ecological modeling and prediction. Ecology 88:243-251.

ELITH, J., C. H. GRAHAM, R. P. ANDERSON, M. DUDIK, 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.

ESRI. 2004. ArcGIS 9.1. *in* ESRI, Redlands, CA.

FRIEDMAN, J. H. 2001. Greedy function approximation: a gradient boosting machine. Annals of Statistics 29:1189-1232.

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.

HEIKKINEN, R. K., M. LUOTO, M. B. ARAUJO, R. VIRKKALA, W. THUILLER, AND M. T. SYKES. 2006. Methods and uncertainties in bioclimatic envelope modelling under climate change. Progress in Physical Geography 30:751-777.

HIJMANS, R. J., AND C. H. GRAHAM. 2006. The ability of climate envelope models to predict the effect of climate change on species distributions. Global Change Biology 12:2272-2281.

HIRZEL, A. H., J. HAUSSER, D. CHESSEL, AND N. PERRIN. 2002. Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? Ecology 83:2027-2036.

JARNEVICH, C. S., D. T. BARNETT, T. J. STOHLGREN, AND J. KARTESZ. 2007. A simple framework for an invasive species early warning system for counties. Frontiers in Ecology and the Environment in review.

MCCULLAGH, P., AND J. A. NELDER. 1989. Generalized linear models, 2nd edition. Chapman and Hall, London; New York.

PEARSON, R. G., AND T. P. DAWSON. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global Ecology and Biogeography 12:361-371.

PETERSON, A. T. 2006. Uses and requirements of ecological niche models and related distributional models. Biodiversity Informatics 3:59-72.

PHILLIPS, S. J., R. P. ANDERSON, AND R. E. SCHAPIRE. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling 190:231-259.

SOBERON, J., AND A. T. PETERSON. 2005. Interpretation of models of fundamental ecological niche and species' distributional areas. Biodiversity Informatics 2:1-10.

SSI. 2004. SYSTAT 11.0. San Jose, California, USA.

STOHLGREN, T. J., D. BARNETT, C. FLATHER, P. FULLER, B. PETERJOHN, J. KARTESZ, AND L. L. MASTER. 2006. Species richness and patterns of invasion in plants, birds, and fishes in the United States. Biological Invasions 8:427-447.

STOHLGREN, T. J., AND J. L. SCHNASE. 2006. Risk analysis for biological hazards: What we need to know about invasive species. Risk Analysis 26:163-173.