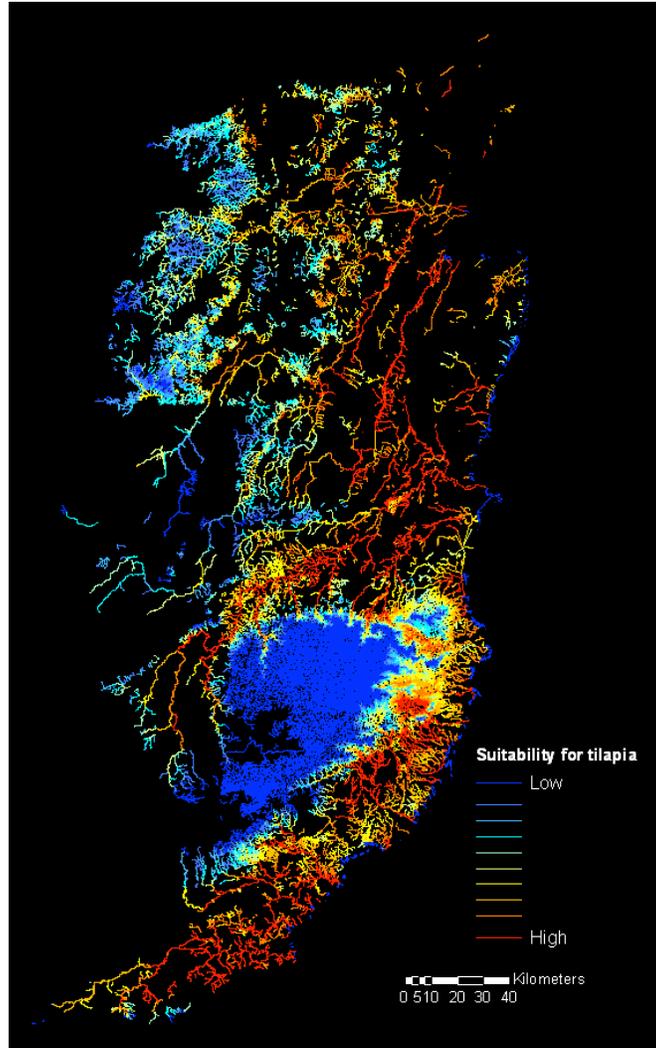


Predicted Spatial Success of African Tilapia in the Domestic and International Watersheds of Belize

Technical Summary of Model Results



17 April 2007

Peter C. Esselman, M.Sc.
School of Natural Resources and Environment
University of Michigan
Ann Arbor, MI 48109
esselman@umich.edu



Abstract

The increased availability of online geographic and species data combined with powerful software for predictive modeling create excellent opportunities for generation of meaningful predictions of species geographic distributions. Invasive species are a substantial threat to biodiversity in aquatic ecosystems, and accurate predictive models of non-indigenous species could assist with efforts to control or stop the spread of these species. The goal of this study was to create and evaluate predictive models of the Nile tilapia (*Oreochromis niloticus*), which is currently invading aquatic ecosystems of Central America. Environmental data representing local and watershed predictors of terrain, geology, soils, ecosystems, and land use were combined with tilapia occurrence data to model habitat suitability for tilapias in the watersheds that drain to the coast of Belize. Maximum entropy (MaxEnt) models were trained on watersheds that have been invaded for at least 5 years, and the results were projected onto data for watersheds that either have not been invaded or have been invaded for less than 5 years. The resulting model predicted tilapias to be present in all of the larger main stem rivers and many coastal plains tributaries within the study area. Indicators of model performance suggest that the model performs very strongly, showing excellent results for each of the three evaluation criteria used. Of the watersheds that have no known or only recently established populations of tilapia, the model predicts that the coastal plain mainstem and large tributary streams of Rio Sarstoon, Temash River, Rio Grande, Monkey River, some of the larger creeks that drain to Placencia Lagoon, and S. Stann Creek are particularly vulnerable to successful colonization. Model outputs predict that smaller rivers draining to Port Honduras (Golden Stream, Deep River, Middle River) are less vulnerable. This pattern of potential tilapia presence suggests that tilapias are most likely to occur in the areas that have also been documented to have the highest richness of fishes, pointing to a potential conservation conflict should tilapias prove to have negative consequences within the study ecosystems. The prediction map is a valuable starting point for conservation planning, and a useful tool for hypothesis generation and formulation of focused research questions. Furthermore, the environmental datasets, when combined with existing point occurrence data of native species, have excellent potential for expanded application to modeling native species and, eventually, the creation of predicted richness maps (by combining all native species predictions) with much potential for use in aquatic conservation planning.

Introduction

Free online specimen and geospatial databases are a tremendous resource to assist biodiversity conservation. Among the most promising potential applications of online knowledge bases is the prediction of habitat suitability for native and non-indigenous species. When applied to non-indigenous species, predictive models allow conservation practitioners to both anticipate species invasions in yet-to-be-invaded areas, and to build expectations about already-invaded areas with little data. Empowered with the ability to visualize current invasions, and to anticipate future invasion potential, practitioners can then assess invasive species threats and formulate prevention and control strategies for susceptible habitats in a timely cost-effective manner.

Aquatic ecosystems are especially vulnerable to disruption by invasive species compared to terrestrial and marine systems (Sala et al. 2000). This is of concern because they also harbor substantial biological diversity and provide many services to human communities. Thus, preventing biological invasions in these systems can be of critical conservation importance, and predictive models can help accomplish this. However, invasion prediction in aquatic (versus terrestrial) environments poses special challenges. In terrestrial environments ecological predictions can be made across wide swaths of landscape that encompass relatively indiscrete environmental gradients. By contrast, when making predictions in aquatic ecosystems, where

the water's edge is a hard barrier to dispersal and where there is often strong upstream to downstream directionality, the best models should consider attributes pertaining not only to the local conditions near a habitat, but all of the integrated conditions in the watershed upstream of a habitat. Thus, the aquatic modeling approach is necessarily different than those commonly used in terrestrial systems. Fortunately, many strong environmental predictors of aquatic species distributions can be easily assessed with geospatial data from earth images and GIS, and these, in turn, can be applied to prediction of habitat suitability in discrete water bodies across large landscape areas (e.g., Joy and Death 2004).

This study integrates geospatial data with field data to demonstrate an effective approach for utilizing GIS and the type of data that exist in online knowledge bases—georeferenced presence-only records—to create predictive models that anticipate species invasions into aquatic habitats. This type of modeling is often referred to as “ecological niche modeling”, because the models attempt to define the “ecological niche” of the organisms being modeled—the combinations of all relevant ecological variables (including biotic interactions) under which a species or population can persist.

This effort focused on the most widely distributed aquatic invasive fish species in Mesoamerica—African tilapias (Canonico et al. 2005). Several species of tilapias have been introduced widely outside of their native ranges for purposes of vegetation control, aquaculture, and capture fisheries (Courtenay 1997). Tilapias are a major cause of conservation concern as shown by studies reporting them as the cause of local extinctions of native species (Twongo 1995, Goudswaard et al. 2002), predation on eggs and young of other fishes (Arthington et al. 1994), eutrophication (Starling et al. 2002), de-vegetation of extensive areas of lake bottom, introduction of non-indigenous parasites to other fishes (McCrary et al. 2001), and food web alterations (Taylor et al. 1984). Predictive models of habitat suitability for tilapias can assist with the development of conservation plans focused on minimizing their impacts.

The approach demonstrated here can be used to develop models for any riverine ecosystem where a species' success or failure is largely determined by the physical, chemical, and positional aspects of aquatic habitats. For instance, habitat suitability modeling can also be extremely useful for other conservation applications, such as the prediction of native species distributions, which when added together can yield important conservation indicators such as native species richness. Expanded applications of this study are addressed in the Discussion.

Methods

Study Area

This research was carried out in the watersheds that drain to the Belizean coast. This includes 16 major watersheds—bounded by Rio Hondo in the north and Rio Sartsoon in the south—and numerous smaller watersheds (Figure 1). This area encompasses both the low-elevation limestone-based watersheds of the Yucatan Peninsula, and the more mountainous watersheds of southern Belize and Guatemala that originate in variable geologies and from elevations greater than 1000 m. The northern part of the study area is characterized by spring fed streams and large meandering rivers that cross an extensive coastal plain with many lagoons. The southern part of the study area is characterized by high-gradient surface and spring fed streams flowing to large meandering rivers that traverse a short distance across the coastal plain to the sea. The topographic and geologic differences between the northern and southern parts of the study area make it an excellent test site for development of predictive models that will be able to generalize to mountainous landscapes of the Caribbean slope of Central America and to the low elevation areas of Yucatan Peninsula.

Only streams and rivers were considered in this study, though field efforts are currently underway (2007 field season) to sample lagoon and wetland habitats. In preparation for this effort, all stream lines ($n = 36,436$) and water bodies ($n = 3,966$) in the study area were hand digitized from scanned and georectified 1:50,000 topographic maps. The hand-digitized streamlines were then used to condition a 30 m Shuttle Radar Topography Mission digital elevation model, which was in turn used to create a flow direction grid that was used to calculate watershed variables that match the stream line layer.

Description of modeling approach

A modeling approach was chosen that was (1) quantitatively rigorous; (2) required presence-only data, not presence and absence; (3) available to the public for free; and (4) user friendly. The latter two criteria are important so that the methods reported here can be easily replicated by other users. The approach chosen is called maximum entropy or MaxEnt. MaxEnt is a mathematical approach to predicting an unknown probability distribution based on the principle that the estimated distribution must agree with everything that is known about its occurrence and be subject to no unfounded constraints. The approach estimates the most uniform distribution (e.g., the distribution with “maximum entropy”) across a defined area subject to the constraints imposed by information available about environmental conditions at the locations where a species is known to occur and at random background points (Phillips et al. 2004, Phillips et al. 2006).

In mathematical terms the final MaxEnt probability distribution maximizes the product of the probabilities of the sample locations, and takes the form:

$$P(x) = \exp(c_1 * f_1(x) + c_2 * f_2(x) + c_3 * f_3(x) \dots) / Z$$

Here c_1, c_2, \dots are constants (“weights”), f_1, f_2, \dots are the environmental features, and Z is a scaling constant that ensures that P sums to 1 over all cells in the study area. The algorithm that is implemented by this program is guaranteed to converge on the values of c_1, c_2, \dots , and Z , that give the optimum distribution P (Phillips et al. 2006).

In model training, MaxEnt uses a mathematical process that iteratively adjusts the weights associated with each environmental variable as it is presented to the model to maximize the likelihood that the occurrence data used to train the model are correctly predicted. By adjusting these weights many times (e.g., 500), the MaxEnt algorithm converges on the optimum probability distribution—the best solution to the problem given the data available. The output of a MaxEnt model is a continuous surface of values ranging between 0 and 100, with higher values indicating a higher suitability of that area for the target species.

Several recent studies have compared MaxEnt to other modeling techniques, including another popular presence-only technique called Genetic Algorithm for Ruleset Production (GARP). MaxEnt has consistently performed highly in modeling tests, often outperforming GARP (Phillips et al. 2004, Hernandez et al. 2006, Phillips et al. 2006, Pearson et al. 2007), and displaying good predictive power even with extremely low species occurrence sample sizes (less than 10)(Hernandez et al. 2006, Pearson et al. 2007). This makes MaxEnt a well-suited method for use in data poor areas, or for prediction of rare species.

MaxEnt software was downloaded from the internet¹ and used to develop predictive models. Inputs to the MaxEnt software include georeferenced point data of places where a species is known to occur, and raster (pixel-based) data layers of the different environmental attributes (climate, topography, etc.) hypothesized as important to freshwater fish distributions. Data points for tilapias were only collected from within those watersheds with a confirmed presence of tilapias for more than 5 years (an assumed time period sufficient for tilapias to expand their population and ranges into suitable habitats within the study watersheds). The watersheds that met this criterion, and thus were included in the model training process, were (from N to S) Rio Hondo, New River, Central River, Belize River, Sibun River, North Stann Creek, and Moho River (Figure 1). The results from models developed for these 7 watersheds were then projected onto the environmental conditions in the remaining watersheds to yield the final predictive map of tilapias in the area.

Environmental attributes

A total of 30 variables were prepared as individual raster layers for possible inclusion in the model. There were two primary data sources used to get base data layers, which were then subject to various resampling and processing operations (described below): The Nature Conservancy's (TNC) Selva Maya Ecoregional dataset² and the Inter-American Biodiversity Information Network-Development Grant Facility's (IABIN-DGF) 30 m hydrologic derivatives³. The former dataset was used as a source for temperature, precipitation, elevation, geology, soils, ecosystems, roads, and settlements data, and the latter was used for slope and flow direction (Table 1). The flow direction grid used was derived from a 30 m digital elevation model with the hand-digitized streamlines generated for this project burned into the grid to constrain the flow direction to the location of the stream channel used here.

Both local variables representing only the conditions occurring underneath each stream unit, and catchment variables representing average or cumulative conditions upstream of a given location in the stream network were used (Table 1). Four variable preparation processes were performed in ArcGIS 9.0® (ESRI Corporation) to derive the data layers: (1) clipping and resampling of raster grids to attain an equal grid extent and cell size across all layers; (2) calculation of positional metrics for each stream line in the stream and river

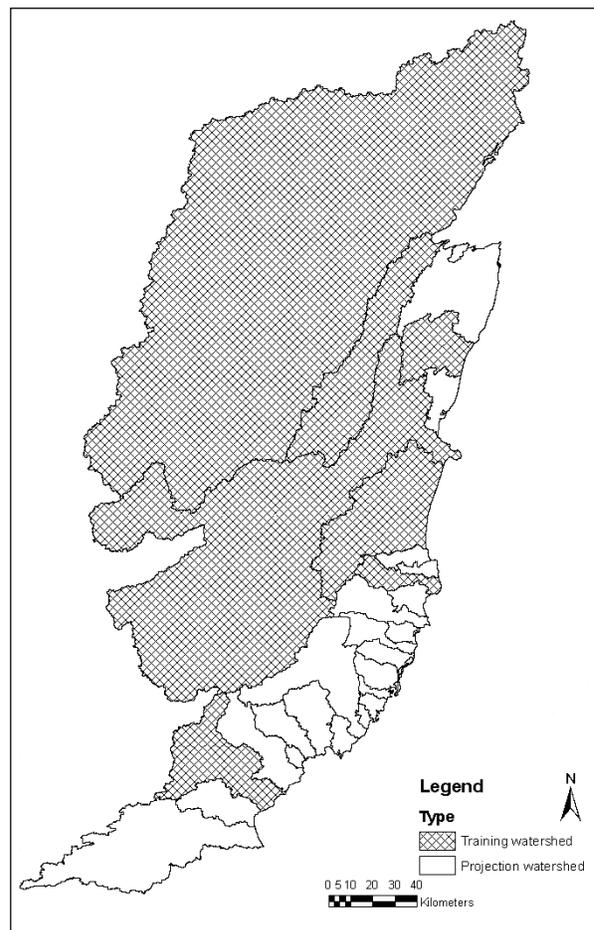


Figure 1. The study area extended from southern Mexico (Rio Hondo) to eastern Guatemala (Sarstoon River), and included all of Belize. Seven watersheds were used to train models (cross hatch fill), which were in turn used to project results to the other watersheds (white fill).

¹ <http://www.cs.princeton.edu/~schapire/MaxEnt/>

² <http://www.selvamaya.org>

³ http://edcintl.cr.usgs.gov/iabin_datadownload.html

coverage, which were then converted to individual rasters for each variable; (3) calculation of weighted flow accumulation to represent catchment proportions of geologies, soils, ecosystems, and upstream averages of catchment precipitation, temperature, slope and elevation; and (4) application of a river mask to exclude only those pixels that lay on top of designated stream or river segments.

Clipping and resampling occurred to ensure that all environmental layers had the same extent and cell size. Elevation, precipitation, and temperature base layers were resampled to 30 m from 60 m, 1000 m, and 1000 m respectively, using cubic convolution as the resampling technique.

Positional metrics are variables that measure the distance from the center of any given stream segment to some feature of interest. In this case each stream segment was attributed with a measure of linear distance to the Caribbean Sea, the nearest human settlement, and the nearest perennial water body. The size of the nearest water body was also calculated. After these calculations were made, the vector stream layer was converted to an individual raster for each of these four fields.

Weighted flow accumulation relies heavily on the IABIN-DGF 30 m resolution flow direction grid and a weight grid to calculate the values that flow into the next downstream cell. The values in all the upstream weight grid cells are added to calculate the value in each downstream cell of a weighted flow accumulation grid. In this way, the states of different variables in the upstream watershed can be calculated and incorporated into the models. There are several ways that weighted flow accumulation was used. One way was to calculate how many cells above a given pixel in the watershed contained some class state, such as a type of geology. The weight grids used for this purpose were a binary grids (1 or 0 values only) that had a 1 in cells representing locations with the class state present and a 0 where the class was absent. The weighted flow accumulation grid in this case was a calculation how many cells with a 1 flowed into each downstream pixel. This in turn was divided (“normalized”) by an unweighted flow accumulation grid that contained the number of pixels upstream of each pixel to yield the proportion of the watershed in that given class state. Another way that weighted flow accumulation was used was to calculate the sum total of cells in grids containing continuous values (that were not limited to 0 or 1, but had many possible values). The accumulated values of such variables (e.g., such as elevation, precipitation, temperature, etc.) were also divided by an unweighted flow accumulation grid to yield average values of each variable in the watershed above each pixel (rather than proportions). All catchment variables listed in Table 1 were prepared by these two processes.

A mask limits the analysis extent to only those pixels that match pixels that have values within a mask layer. Because this analysis focused on the suitability of streams and rivers for tilapias, only pixels with streams and river present on them were considered. To create a mask that would limit each layer to only those pixels with stream or river lines on them, the vector stream line layer was converted to a raster and applied to all data layers by setting a mask and using either Raster Calculator or the Environments settings to apply that mask.

Tilapia point occurrence data

Tilapia point occurrence data were collected using three separate methods applied within the 7 training watersheds: backpack electrofishing, boat electrofishing, and interviews with fishermen. Backpack electrofishing involves the use of a pulsed electrical current coming from a pair of 12 volt motorcycle batteries to stun and collect fish, which are enumerated and returned to the stream alive. Backpack electrofishing (Smith-Root model 12B) was used in wadable mountain

streams only, where depths were frequently less than waist height. Boat electrofishing also uses a pulsed electrical current, but the current originates from a 5000 watt generator. Fishes are netted with long dip nets from the bow of the boat, placed in a live well, then enumerated and released. Boat electrofishing (Smith-Root GPP 5.0) was used to sample fishes in non-wadable habitats of large deep rivers in the coastal plain. Electrofishing sites were selected by a systematic sample every 20 km along the river channel with a random starting point in the first 5000 river meters. Additionally, 38 interviews were conducted with fishermen about the presence of tilapias in their fishing grounds and voucher specimens were collected to confirm the identity of the species they were catching. Maps were used to help the fishermen identify the areas where they catch tilapia. For each river segment where tilapia was reported as present, a point location was placed in the center of that segment.

Table 1. Environmental variables prepared for entry into MaxEnt models of tilapias (or other species). Variables in bold represent those that were selected for entry into the model after using PCA to eliminate redundant variables (see Model Development below).

| Variable | Min. Value | Mean Value | Max. Value |
|---|--------------|--------------|---------------|
| Average annual catchment air temperature (° C) | 20.0 | 22.9 | 26.0 |
| Average annual catchment rainfall (mm) | 795 | 1519 | 2357 |
| Average catchment elevation (m) | 0.42 | 334 | 1047 |
| Average catchment slope (percent) | 0.00 | 8.93 | 36.03 |
| Average local annual air temperature (° C) | 20.0 | 22.6 | 26.0 |
| Average local annual rainfall (mm) | 794 | 1493 | 2391 |
| Local elevation (m) | 0 | 277.8 | 1019 |
| Local slope (percent) | 0.00 | 7.59 | 99 |
| Catchment area (km ²) | 0.01 | 682 | 164728 |
| Catchment geology proportions of: | | | |
| Alluvium | 0 | 0.10 | 1 |
| Andesite | 0 | 0.43 | 1 |
| Mudstones and shales | 0 | 0.43 | 1 |
| Catchment soil proportions of: | | | |
| Fluvial Gleysols and Vertisols | 0 | 0.38 | 1 |
| Leptosol | 0 | 0.35 | 1 |
| Catchment terrestrial ecosystem proportions of: | | | |
| Caribbean lowland season swamp forest | 0 | 0.02 | 1 |
| Central American Atlantic season and evergreen lowland forest | 0 | 0.19 | 1 |
| Meso American waterlogged savanna | 0 | 0.02 | 1 |
| Peten lowland alluvial seasonal forest | 0 | 0.01 | 1 |
| Peten seasonal evergreen forest on karstic hills | 0 | 0.30 | 1 |
| North Meso American premontane wet forest | 0 | 0.19 | 1 |
| Lowland pine forest | 0 | 0.12 | 1 |
| Catchment landuse proportions of: | | | |
| Agriculture | 0 | 0.09 | 1 |
| Urban | 0 | 0.00 | 1 |
| Reach length (m) | 1.90 | 821 | 35771 |
| Straight line distance to the sea (m) | 13636 | 85562 | 182776 |
| Straight line distance to next perennial lake (m) | 0 | 32442 | 118926 |
| Surface area of nearest lake (km²) | 0.52 | 6.13 | 56.78 |
| Straight line distance to nearest human settlement (m) | 69 | 8282 | 31398 |
| Proportion of catchment with roads | 0 | 0.01 | 0.51 |

Model development

The model development process involved data reduction, training of the model, evaluation of the model, and projection of final predictions to the watersheds not used in model training.

Data reduction was necessary to eliminate redundant variables (e.g., variables carrying roughly the same information) and to yield models that are easier to interpret. The goal was to retain no more than 10 variables that characterized the range of different environmental characteristics at all sites sampled in the watersheds (whether or not tilapias were captured there). Principal Components Analysis (PCA) was used to assist with the variable reduction process. PCA is a statistical approach that can assist with the identification of intercorrelation between many variables within a multivariate dataset like the environmental dataset presented here. The approach allows for an assessment of redundancy between variables (e.g., if 3 variables correlated strongly to the first axis, they carry the same information and can probably be reduced to only one). PCA was performed using PC-Ord software® (McCune and Mefford 1999). This process led to the selection of 8 variables to include in the final model (Table 1, bold).

After variables were selected, the MaxEnt model was parameterized. The model was run in 'samples with data' mode for 500 iterations, using 10,000 randomly selected background points, a convergence threshold of 0.00001, and a regularization factor of 1. Twenty-five percent of the tilapia presence points were withheld from the data set and used to test the accuracy of the model. The jackknife option was selected to assist with the interpretation of variable importance to the model outcome. The jackknife operation automated by the MaxEnt software runs the model over again *without* each environmental variable, and then with each variable *alone*. Running the model without the variable reveals which contains the most information that is not present in other variables. Running the model with each variable alone shows which appear to have the most useful information by themselves.

Three tests were used to evaluate the performance of the model: (1) the area under the curve of the receiver operating characteristic plot (ROC); (2) test prediction success; and (3) a one-tailed binomial test. All three of these outputs are automated by the MaxEnt software. A ROC plot is created by plotting the fraction of true positive predictions (sensitivity) against the fraction of false-positive predictions across all available decision thresholds (a threshold is a point along the curve above which you assume the species is present, and below which it is assumed not present or of unknown presence). A ROC curve that maximizes sensitivity at low values of the false-positive fraction is considered a good model and can be quantified by calculating the area under the curve (AUC; Fielding and Bell 1997). The AUC is considered a measure of the model's overall performance and usually has values ranging from 0.5 to 1.0, where a score of 0.5 implies that the predicted probability distribution does not discriminate any better than a random probability distribution, and 1 indicates that the model can discriminate between true and false positive occurrences perfectly. The other two metrics of model performance require the selection of a threshold value. Here, the minimum training presence value was used as the threshold. The minimum training presence threshold uses the lowest predicted suitability value of all the pixels occurring underneath the training samples (where we knew the fish to be present), and can be interpreted as the value that is at least as suitable as those where the species has been recorded present. Once a threshold is selected two additional performance metrics are applicable. Test prediction success is the percentage of sites in the test dataset that were successfully predicted as present by the model at the given threshold. A one-tailed binomial test determines whether a model predicts the test localities significantly better than random. If the P-value is greater than 0.05 then the model failed to significantly predict better than random using a 95% confidence interval.

To display results, raster predictions were converted back to vector by averaging pixels underneath each stream segment (using zonal statistics in ArcGIS Spatial Analyst). Stream segments were then classified by the threshold value resulting in two categories (“predicted present” and “unknown presence”). Because MaxEnt uses presence-only data, it is impossible to infer true absence from the outputs. However, under the assumption that sampling was intensive enough to document tilapia presence to the edge of its range within a watershed, for planning and hypothesis generation, I present values below the presence threshold as “Assumed absent” here.

Results

Oreochromis niloticus, the Nile tilapia, was documented from a total of 67 different localities within the 7 training watersheds (Figure 2). These point data were used in conjunction with 8 variables selected from the environmental variable pool to create MaxEnt models of habitat suitability for tilapias. The 8 variables retained in the model included linear distance to the sea, local elevation, local temperature, catchment proportion of the Peten seasonal evergreen forest on karstic hills ecosystem, catchment proportion of agricultural land use, catchment proportion covered with roads, distance to nearest human settlement, and size of nearest water body. The models trained on data from tilapia invaded watersheds were projected to environmental layers of all the watersheds in the study area.

The minimum training presence threshold (at cumulative value = 2.659) was used to classify all river cells as to whether the habitat was predicted to be suitable for tilapias (e.g., tilapias present) or not appropriate (assumed absent). Twenty-eight percent of all rivers in the study area (8256 km of 29858 km) were predicted to be suitable for *O. niloticus*, especially those rivers located within lower elevation areas of the coastal plains near main stem rivers (Figure 3).

Model performance measures indicated that the model predicted tilapia presence with very high accuracy that was very significantly different from what would be expected from random predictions. A test AUC of 0.95 is a very high value that suggests that 95% of the time a random selection of values from pixels with tilapia known to be present will have a

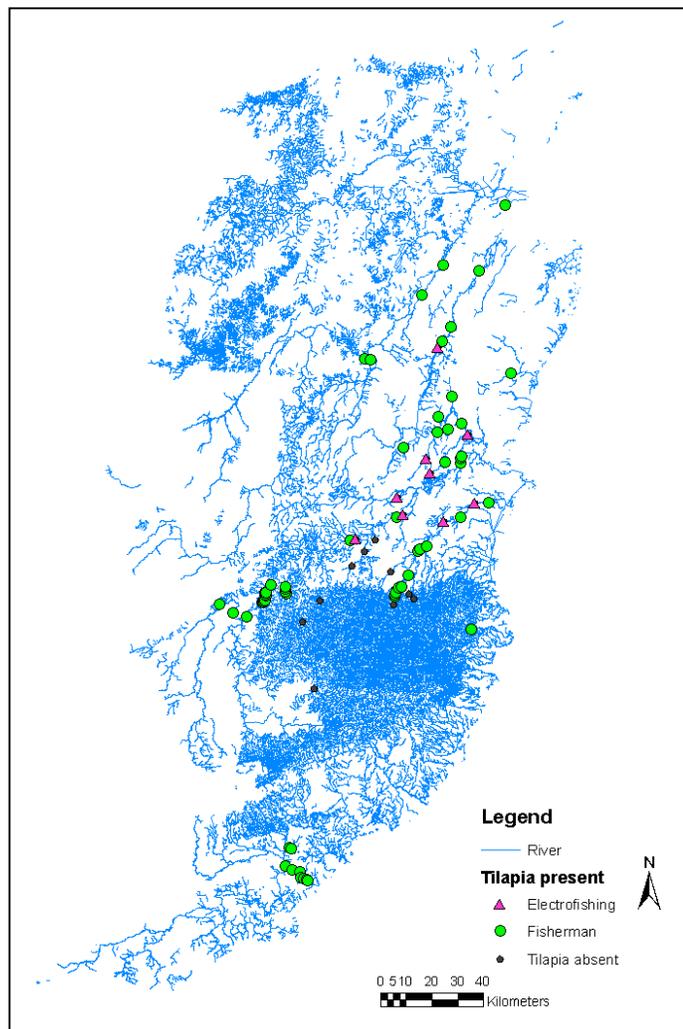


Figure 2. Tilapias were collected at 67 localities within the 7 training watersheds using electrofishing (pink triangles) and fisherman surveys (green dots). Electrofishing sites where no tilapias were collected are represented in black (these sites not included in model development).

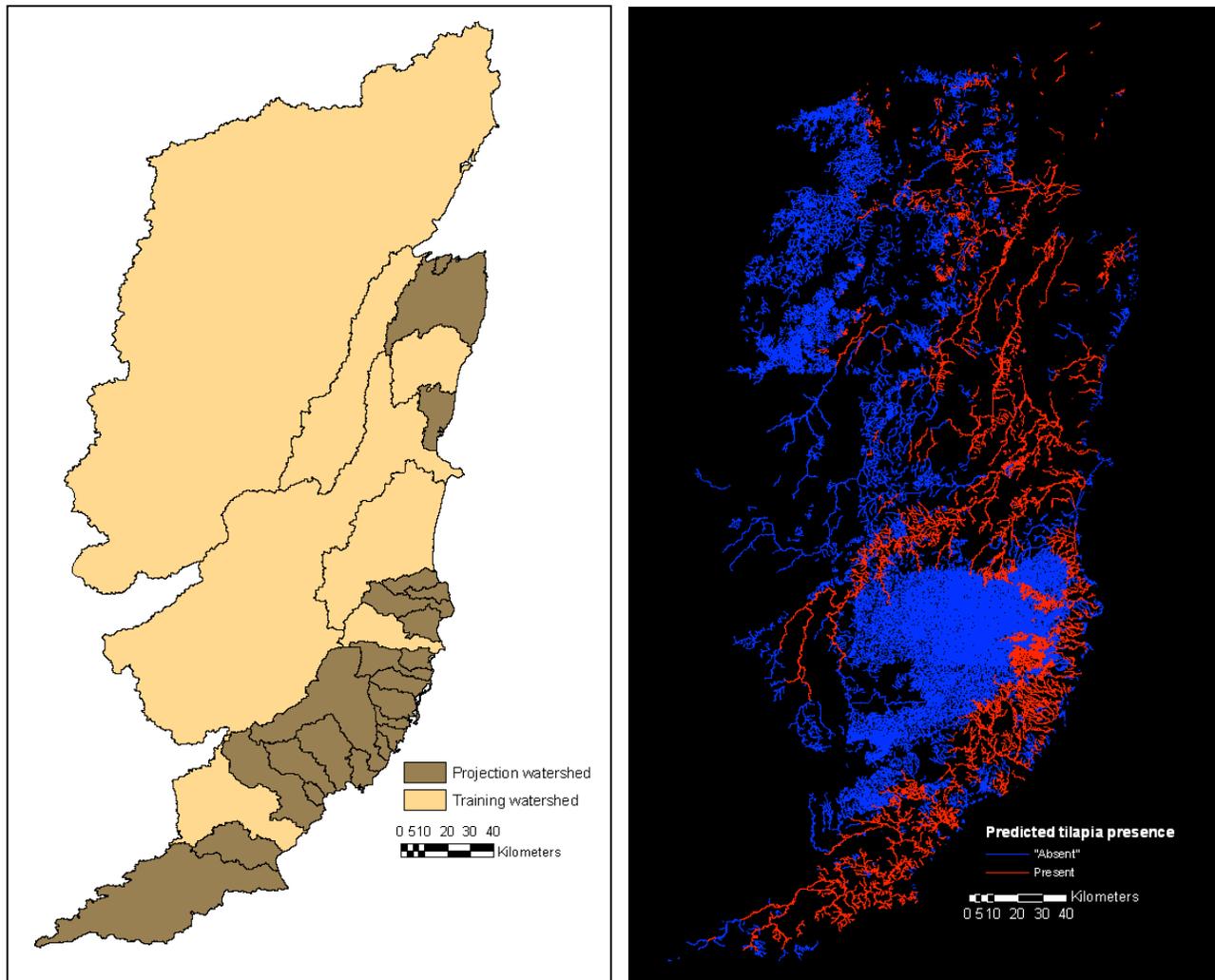


Figure 3. Training and projection watersheds used in the model (left). Predictions within projection watersheds represent an estimate of vulnerability to tilapia colonization. (Right) Predicted results from the MaxEnt model of tilapia after a threshold (minimum training presence) was applied. Red lines indicate that the habitat was predicted to be suitable for tilapias (predicted presence), and blue lines represent assumed absence (“Absence”).

better predicted suitability than a random selection of values from pixels with no known tilapia presence. This is further reinforced by an overall prediction success rate of 100% of test values, and a highly significant ($P < 0.0001$) binomial test result indicating that the model predicted test results very significantly better than random.

Results of the jackknife test of variable importance suggested that the variable that appears to have the most useful information by itself is local elevation, followed by catchment proportion of the Peten seasonal evergreen forest on karstic hills ecosystem, and catchment proportion of agriculture (Figure 4, blue bars). Local temperature had the least useful information by itself. The environmental variable that appeared to have the most information that was not present in the other variables was catchment proportion of Peten seasonal evergreen forest on karstic hills ecosystem, followed by local elevation and distance to the nearest human settlement (Figure 4, green bars). The variable that had the least information that was not present in the others was local temperature.

Discussion

The model of habitat suitability for tilapias in Belize represents the first of its kind for tilapias in the region and perhaps the world (based on review of the published literature). The results of the model can be interpreted as displaying those habitats that are suitable for tilapias, and those for which suitability is unknown or unlikely. It is important to note the need to use “unknown suitability” instead of a more absolute designation of “unsuitable”. This is the result of using presence-only data rather than presence and absence data. Were true absence data available for modeling, then the conclusions could be more assertive. Nevertheless, having a prediction of those habitats that are suitable allows for the conclusion that, if tilapias were to occupy all habitats suitable to them, that they would be present in all of the river segments represented as red in Figure 3.

From these results, it can be hypothesized that the coastal plain reaches and some larger mountain tributaries of the river systems that have not yet been invaded by tilapias are vulnerable to invasion. The majority of these rivers are those flowing from west to east out of the Maya Mountains, which have no known tilapia populations (except for Monkey River). These results support the idea that tilapia prevention measures and education are warranted in these un-invaded watersheds, especially if they are found to have exemplary or unique species assemblages or ecological communities.

Model results suggest that potential negative influences of tilapias are likely to be constrained to coastal plain ecosystems, unless impacts

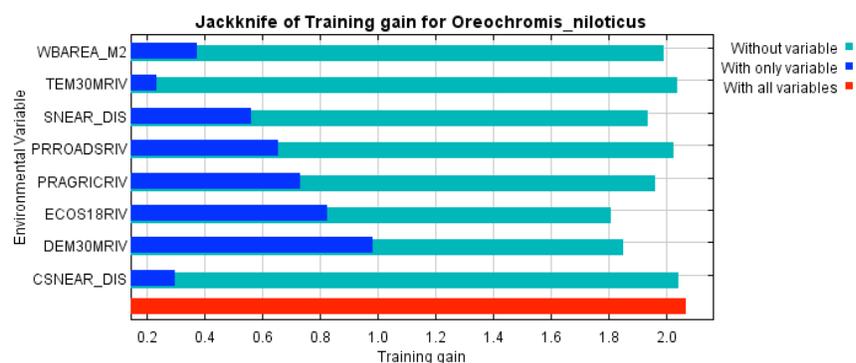


Figure 4. Results of jackknife test of variable importance. Blue bars represent the model run with only each variable listed at the left. Green bars represent the model run with all variables except the one listed at left. These can be interpreted as the variable that contributes the most to the model alone, and the one that has the most amount of information not represented by other variables respectively. WBAREA_M2 = water body area; TEM30MRIV = local temperature; SNEAR_DIS = distance to human settlement; PRROADSRIV = proportion of catchment in roads; PRAGRICRIV = proportion of catchment in agriculture; ECOS18RIV = proportion of catchment in Peten seasonal evergreen forest; DEM30MRIV = local elevation; CSNEAR DIS = distance to sea.

are transmitted upstream via interactions with migratory species that move to and from the mountains to fulfill their life cycles. Previous studies in Belize (Esselman et al. 2006) and in other locations in Mesoamerica (Lyons and Schneider 1990) have shown that native fish species richness is highest in coastal plain rivers, increasing as they near the sea. This suggests that, should the predictions of tilapia presence prove accurate in nature, the habitats that are most suitable to tilapias are also the most species rich habitats within the river network. This raises obvious conservation concerns, as tilapias have been proven to have negative consequences for native species or ecosystem processes (e.g., nutrient cycling) in many ecosystems outside of the study area (Canonico et al. 2005).

The results of the jackknife tests of variable importance point to the clear importance of local elevation, and the possible relationship between tilapia occurrence and anthropogenic activities, particularly upstream agricultural activities and distance to the nearest human settlement (Figure 4). It is difficult to say for certain whether these variables or their correlates (e.g., slope is a correlate to elevation) are actually driving apparent patterns in tilapia occurrence. The Peten seasonal evergreen forest on karstic hills ecosystem is another variable that is difficult to interpret. The ecosystem is located in the medium elevation hills of almost all of the watersheds in the study area, but do not occur in the lowlands or the highlands. This is also a plant community that is affiliated with limestone geology. It is possible that one or both of these relationships—middle elevations and/or limestone bedrock—are influencing the predicted pattern of tilapia habitat suitability.

The results of predictive models—even those with high performance against test data—should always be interpreted somewhat conservatively. More than anything, the models help us recognize possible scenarios that should be validated and elucidated through targeted research programs. One important research question is what impacts, if any, does tilapia have on fish communities and ecosystem processes within the specific ecological context of these study systems? Other questions relate to the siting of conservation activities. Questions such as, what habitats that are suitable to tilapias contain the most species richness? Which of these high diversity systems are most prone to disruption and why? Which of the rivers that are yet to be invaded, are most important to conservation? Each of these questions could influence the types of decisions that should be made to conserve native biodiversity.

The model predictions were made based on incomplete information. Not all stream types were sampled in this study. Additional

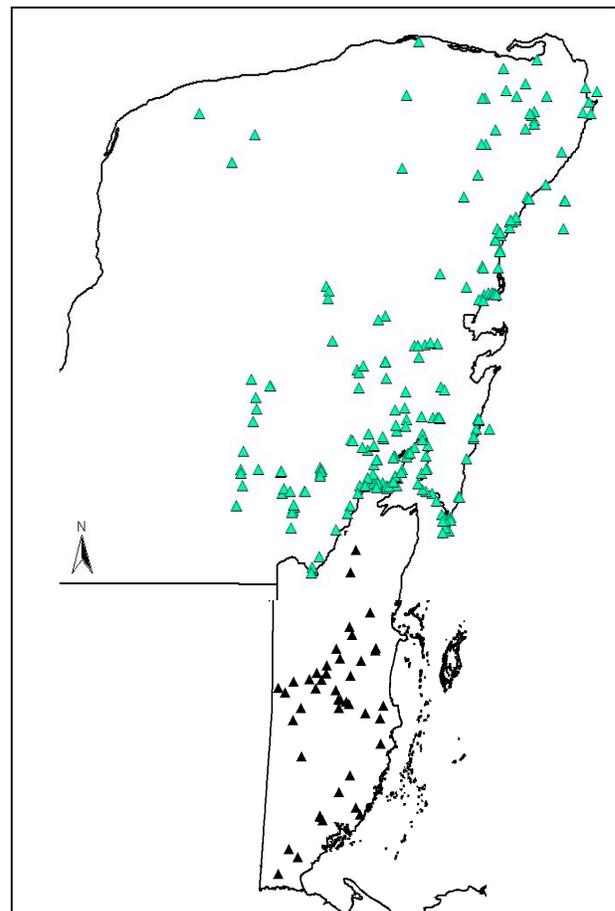


Figure 5. Sample sites for native species available from Esselman (unpublished; black dots), and Schmitter-Soto (1998). An additional 115 sites are not shown (from Greenfield and Thomerson 1997)

information from a greater diversity of habitats will only improve the accuracy of the model, which already performed well. Simply having the ability to ask informed questions reveals the utility of such models for aiding the formation of reasonable hypotheses that can help guide our conservation and research activities.

Though only one species of conservation concern is presented here, all native species for which data are available could be predicted from the same exact datasets. In practical terms, the majority of the work in this process is in preparation of the environmental data layers. Now, at least 30 different variables are available to conduct high resolution modeling of other species. Data for such an effort include samples from this study ($n = \sim 100$), samples collected from 1970 to 1979 in Belize by Greenfield and Thomerson (1997; $n = 115$), and samples from the mid-1990's in Mexico ($n = 237$; Schmitter-Soto 1998) for a total of 452 sites (Figure 5). These data points give the potential to model more than 119 fish species from the study area. Not each species is present at all of these points, but in cases where even 10 points are available, MaxEnt models have been shown to produce accurate results (Pearson et al. 2007). Thus, an important expanded application of these data is to predict habitat suitability for all native species in the study area.

Powerful conservation planning tools will result from having multiple data layers of predicted native species presence, and layers of potential threats like tilapias, or a combined threat index (e.g., Clark and Schill in prep). It is an easy operation to add raster layers in ArcGIS to yield a predicted richness surface. Overlaying such a richness surface with some index of anthropogenic stress could lead to effective ways for visualizing and planning aquatic biodiversity conservation in settings where data are scarce. In such cases, the sole option is often to proceed based on our best conservation hypotheses. This is the situation in most developing countries, in many developed countries, and certainly within the study area presented here. Fortunately, with the technology and the datasets both already available, the situation is ripe to create predictions of the entire fish community to continue to push aquatic conservation planning forward in a region that is in great need of such capacity.

Acknowledgements

Funding for this research was provided by the Inter-American Biodiversity Information Network as part of their Development Grant Facility project (IABIN-DGF). Additional funds were provided by the Belize Foundation for Research and Environment (BFREE), the University of Michigan School of Natural Resources, and the University of Michigan Office of the Vice President for Research. The project would not have been possible without the support of the Belize Forestry Department and Department of Fisheries who provided permits, and the Belize Land Information Center (Edgar Ek) and SERVIR (Emil Cherrington). Matt Cushing (USGS), Emil Cherrington (CATHALAC), Steve Schill (TNC), and Christine Geddes (Michigan Department of Natural Resources) all provided critical GIS support. This project would not have been possible without the cooperation of The Nature Conservancy in southern Mexico and Guatemala, BFREE, and El Colegio de la Frontera Sur in Mexico. Drs. Vince Abreu, J.J. Schmitter-Soto, J.D. Allan, J.S. Diana, and G.R. Smith provided important intellectual input and advice throughout the term of the project. M. Issa, A. Garbutt, Miles and Lars Lish, and Irin Combs are thanked for their hard work in the field.

Literature Cited

- Arthington, A. H., D. R. Bluhdorn and M. J. Kennard. 1994. Food resource partitioning by *Oreochromis mossambicus*, and two native fishes in a sub-tropical Australian impoundment. In L. M. Chou (ed.), The 3rd Asian Fisheries Forum. Asian Fisheries Society, Philippines.
- Canonico, G.C., A. Arthington, J.K. McCrary, and M.L. Thieme. 2005. The effects of introduced tilapias on native biodiversity. *Aquatic Conservation: Marine and Freshwater Ecosystems* 15: 463-483.
- Courtenay, W. R. 1997. Tilapias as non-indigenous species in the Americas: environmental, regulatory and legal issues. Pp. 18-33 in B. A. Costa-Pierce and J. E. Rakocy (eds.), *Tilapia Aquaculture in the Americas*, Vol. 1. World Aquaculture Society, Baton Rouge, Louisiana.
- Esselman, P.C., M.C. Freeman, and C.M. Pringle. 2006. Fish-assemblage variation between geologically defined regions and across a longitudinal gradient in the Monkey River Basin, Belize. *Journal of the North American Benthological Society* 25: 140 – 154.
- 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.
- Goudswaard, P. C., F. Witte and E. F. B. Katunzi. 2002. The tilapiine fish stock of Lake Victoria before and after the Nile perch upsurge. *Journal of Fish Biology* 60: 838-856.
- Greenfield, D.W. and J.E. Thomerson. 1997. *Fishes of the Continental Waters of Belize*. University of Florida Press, Gainesville, FL.
- Hernandez, P.A., C.H. Graham, L.L. Master, and D.L. Albert. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* 29: 773 – 785.
- Joy, M.K., and R.G. Death. 2004. Predictive modeling and spatial mapping of freshwater fish and decapod assemblages using GIS and neural networks. *Freshwater Biology* 49: 1036 – 1052.
- Lyons, J., and D.W. Schneider. 1990. Factors influencing fish distribution and community structure in a small coastal river in southwestern Costa Rica. *Hydrobiologia* 203:1–14.
- McCrary, J.K., E.P. van den Berghe, K.R. McKaye, and L.J. Lopez Perez. 2001. Tilapia cultivation: a threat to native fish species in Nicaragua. *Encuentro* 58: 3-13.
- McCune, B. and M.J. Mefford. 1999. *PC-ORD: Multivariate Analysis of Ecological Data*, Version 4. MjM Software Design, Gleneden Beach, OR.
- Pearson, R.G., C.J. Raxworthy, M. Nakamura, and A.T. Peterson. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography* 34: 102 – 117.
- Phillips, S.J., M. Dudik, and R.E. Schapire. 2004. A maximum entropy approach to species distribution modeling. Pages 655-662 in *Proceedings of the Twenty-First International Conference on Machine Learning*.

- Phillips, S.J., R.P. Anderson, and R.E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231 – 259.
- Sala, O. E., F. S. Chapin, et al. 2000. Biodiversity - global biodiversity scenarios for the year 2100. *Science* 287: 1770-1774.
- Schmitter-Soto, J.J. 1998. *Catalogo de los peces continentales de Quintana Roo*. San Cristobal de las Casas, Mexico, El Colegio de la Frontera Sur.
- Starling, F., X. Lazzaro, C. Cavalcanti and R. Moreira. 2002. Contribution of omnivorous tilapia to eutrophication of a shallow tropical reservoir: evidence from a fish kill. *Freshwater Biology* 47: 2443-2452.
- Taylor, J. N., W. Courtenay Jr and J. A. McCann. 1984. Known impacts of exotic fishes in the continental United States. Pp. 322-373 in W. R. Courtenay Jr and J. R. Stauffer (eds.), *Distribution, Biology, and Management of Exotic Fishes*. Johns Hopkins University, Baltimore.
- Twongo, T. 1995. Impact of fish species introductions on the tilapias of Lakes Victoria and Kyoga. In Hart, T.J. (ed.). *The Impact of Species Changes in African Great Lakes*. Chapman & Hall, London.

