

Precipitation levels predict the distribution of an invasive apple snail (*Pomacea maculata*) throughout the Southeastern United States

Abstract

With an invasive apple snail species dispersing rapidly throughout the southeastern United States, it is crucial to understand the extent of its range in order to manage aquatic ecosystems and to infer the success of other species that select for it, namely the endangered Florida snail kite population. *Pomacea maculata*, a large freshwater snail native to South America, was introduced to southeast Florida in the early 1990's. Since then, these amphibious snails have spread northwards and altered aquatic landscapes throughout the coastal southeastern US with their voracious feeding on aquatic plants and other snail species. The random forest model was employed to discover that the potential range for *P. maculata* is a wide margin of the coastal Southeast. Through forward feature selection, it was determined that the snail is limited to these areas because of a high sensitivity to precipitation levels. This model exemplifies how machine learning methods are ideal for modeling species distributions since they are relatively easy to implement and adept at capturing complex spatial dependencies.

Keywords: invasive species, apple snail, species distribution modeling, machine learning

Introduction

The island apple snail, *Pomacea maculata* (formerly known as *Pomacea insularum*), is native to South America and is thought to have been introduced to the Southeastern United States in the early 1990's through the pet trade (Rawlings, Hayes, Cowie, & Collins, 2007). Since then, the amphibious snails have spread across eight southeastern states in the USA and altered aquatic landscapes with their voracious feeding habits (Carlsson, Bronmark, & Hansson, 2004). Aside from its impacts on the environment itself, the *Pomacea* genus is capable of harboring toxins that are hazardous to wildlife higher up on the food chain (Rawlings, et al., 2007). More specifically, laboratory studies have showed that *P. maculata* can transfer the neurotoxin linked to the frequently-lethal neurological disease Avian Vacuolar Myelinopathy (AVM) (Robertson, 2012).

The spreading of AVM is of heightened concern since the federally-listed endangered Everglade snail kite selects highly for the apple snail. The island apple snail is much more fecund than the native *Pomacea paludosa*; it can produce clutch sizes up to 100 times larger than that of its native congener (Byers, McDowell, Dodd, Haynie, Pintor, & Wilde, 2013). Hence with the invasive snails becoming increasingly more available to the kites than the native ones, the risk of AVM in the already-threatened kite population continues to rise.

Initial research suggested that the introduction of *P. maculata* led to diminished foraging success for the Everglade snail kite, where it was observed that juvenile kites were dropping a greater proportion of the invasive snails due to their larger size (Cattau, Martin, & Kitchens, 2010). However, more recent studies have shown that snail kite bill size and body mass have been increasing substantially since the island apple snail was introduced, leading to a greater foraging and nesting success of the juvenile kites (Cautta, Fletcher, Kimball, Miller, & Kitchens, 2018). In turn, the distribution of the kites has expanded northwards through Florida, closely following the spread of the non-native snail populations (Cautta, Fletcher, Reichert, & Kitchens, 2016).

Thus the objective of this paper is to build a predictive model that captures the distribution of the invasive island apple snail population once it has reached its equilibrium and to identify which bioclimatic factors are influential of its success. With the snail's ecological niche realized, one can take preventative measures and manage for the aquatic areas where the snails are likely to inhabit, as well as infer the distribution and success of the endangered Everglade snail kite population. This species distribution was modeled using machine learning methods since the algorithms are adept at capturing complex spatial dependencies within the data and relatively easy to implement.

Methods

The distribution data for *P. maculata* was gathered from the USGS Nonindigenous Aquatic Species database. There were a total of 725 observations reported by citizens which included the coordinates of the sighting. Since pseudo-absences can enhance the accuracy of machine learning predictions, an equal number of background points was generated to supplement the presence-only dataset, as shown in Figure 1 (Barbet-Massin, Jiguet, H  l  ne Albert, & Thuiller, 2012). Seventeen linearly independent bioclimatic variables obtained from WorldClim along with the longitude and latitude of the observations constituted the 19 predictors considered in the analysis. The machine learning methods employed predicted a binary categorical response of whether the snail was present or absent at the given coordinates.

K-fold Cross Validation

In order to account for the spatial dependence between data points, a spatial k-fold cross-validation was implemented. Cross-validation is used to estimate the predictive power of a machine learning model. Each data point is assigned to a fold and the model is built on the data

PRECIPITATION LEVELS PREDICT THE DISTRIBUTION OF AN INVASIVE APPLE SNAIL (*POMACEA MACULATA*) THROUGHOUT THE SOUTHEASTERN UNITED STATES

contained in $k-1$ of the folds. An average misclassification rate can then be obtained from how well the model predicts the response of the observations in the left out fold.

Rather than having the observations randomly assigned to k folds, as is typical, the 1450 observations were first sorted into approximately 100 localized blocks, as shown in Figure 2. Then with the observations clustered geographically, these blocks were randomly assigned to $k=10$ folds for the analysis. This way the methods were guaranteed to predict the presence of snails in a previously unseen geographic area.

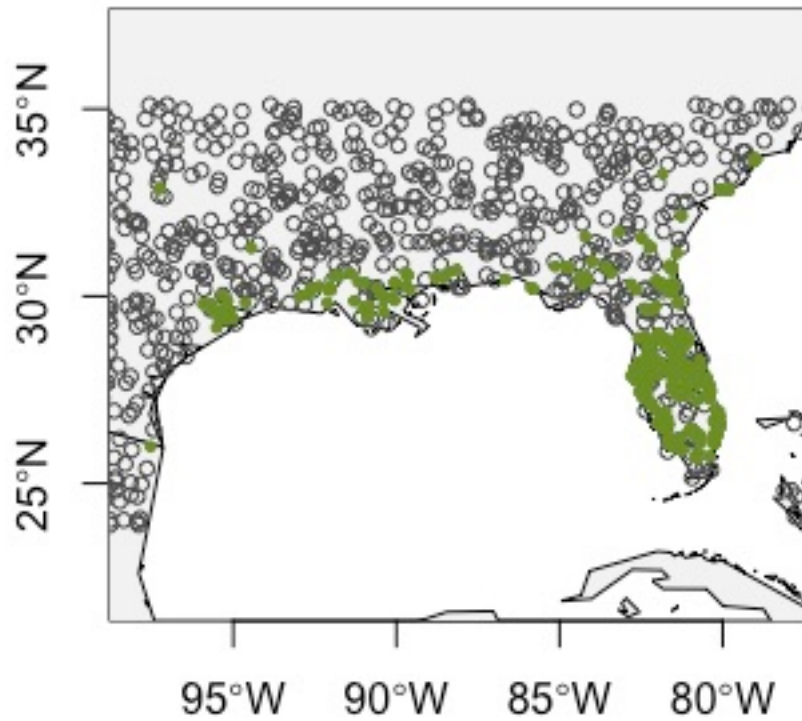


Figure 1. Presence and generated pseudo-absence points of *P. maculata*

Machine Learning

Firstly, forward feature selection was implemented with a logistic regression to determine which variables had the highest predictive power. This was calculated by ordering the predictors from highest to lowest explanatory power in each fold. A glm was then fit for each of the predictors, beginning with the null model and then iteratively adding in the next best predictor. The validation error rate per number of predictors was then averaged across folds.

This method indicated that including the best 11 out of the 19 predictors yielded the most accurate model with a misclassification rate of 13.6%. The best predictor, precipitation of the warmest quarter (Bio18), accounted for 48.9% of the variability in the distribution of the apple snail. The other strong predictor, the mean temperature of the wettest quarter (Bio8), explained an additional 6.4% of the variability. The remaining selected predictors individually added <1% increase in R^2 value. Using the best 11 predictors increased the R^2 value from 48.9% to 58.2%.

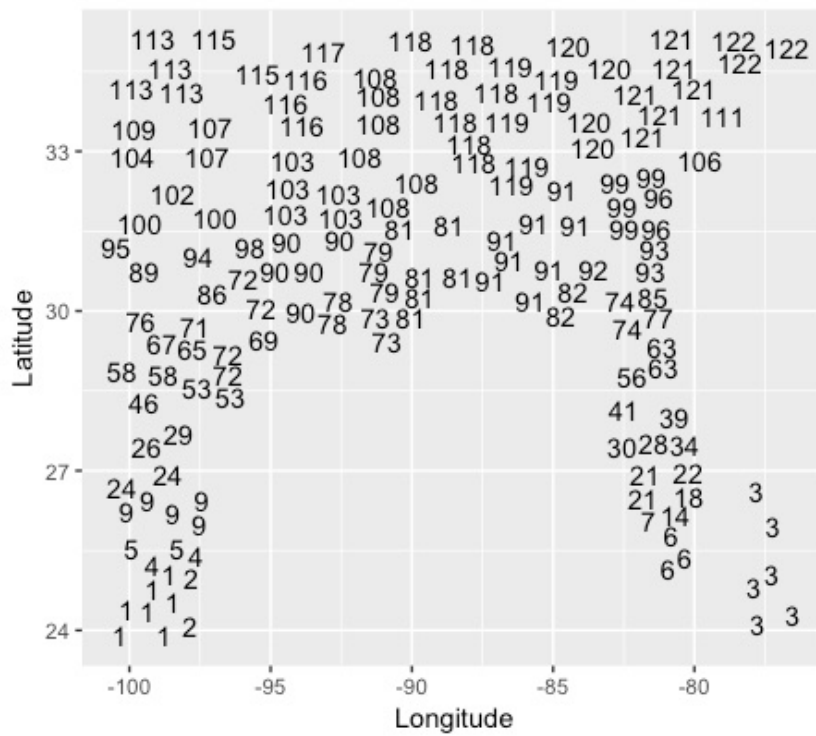


Figure 2. Observations sorted into localized blocks for k-fold cross-validation

Furthermore, a generalized additive model (GAM) and a random forest were tested for best fit. A GAM is slightly more flexible than the logistic regression considered in the forward feature selection since the low-order polynomials fit to localized areas of the curve are adept at capturing non-linearities while still controlling for the complexity of the model. Three GAMs were considered with natural splines placed on the best 3, 5, and 8 predictors. This resulted in misclassification rates (AUC) of 14.5% (0.865), 14.4% (0.867), and 18.7% (0.819), respectively,

PRECIPITATION LEVELS PREDICT THE DISTRIBUTION OF AN INVASIVE APPLE SNAIL (*POMACEA MACULATA*) THROUGHOUT THE SOUTHEASTERN UNITED STATES

indicating that the increased flexibility reduced the accuracy of predicting the distribution of the island apple snail.

Lastly, the random forest was fitted to the data. Each individual tree is shallow with only a few internal splits and is decorrelated from the other trees by only utilizing a fraction of the predictors. Averaging the results across the trees reduces the overall variance of the model. The random forest model had a misclassification rate of 12.0% and AUC of 0.885. Hence the random forest was best at classifying which areas are habitable for the island apple snail.

Results

Through the random forest, it was discovered that the island apple snail's ecological niche is a large margin of the coastal Southeast, as seen in Figure 3. As shown, the state of Florida offers optimal habitat for the snails, along with the coastal regions of the states between Texas and North Carolina.

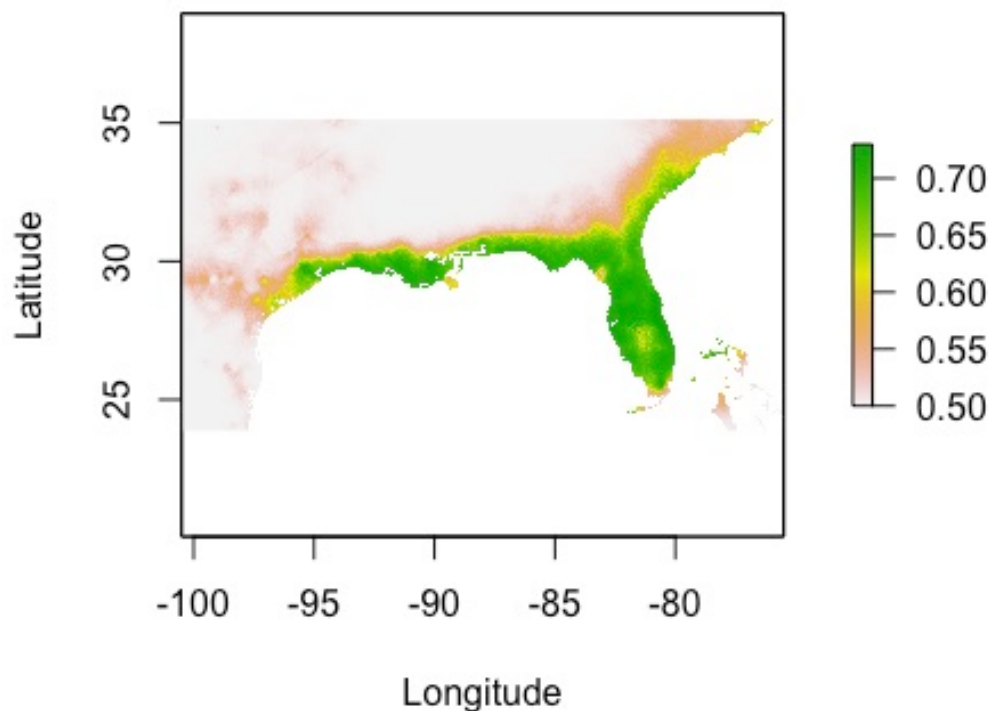


Figure 3. Projected probability of distribution of island apple snail

This trend can be largely attributed to the snail's sensitivity to the bioclimatic variable precipitation of the warmest quarter (Bio18), as shown in Figure 4. This variable indicates that the snail thrives in geographical areas with high amounts of precipitation. Furthermore, the other strong predictor, mean temperature of the wettest quarter (Bio8), illustrates that the snails have a high aversion to low temperatures in the wet season, as seen in Figure 5.

By this model, the snail populations have largely reached the extent of their projected range, but the snails will become more abundant in these optimal geographic areas. Thus it follows that there will be food resources available for the specialist predator as far north as the coastal Carolinas and as far west as coastal Texas and at much higher densities than was previously available with the native apple snail. If the northward movement and higher juvenile survival rates trend continues for the Everglade snail kite, we have reason to be hopeful for the survival of the species.

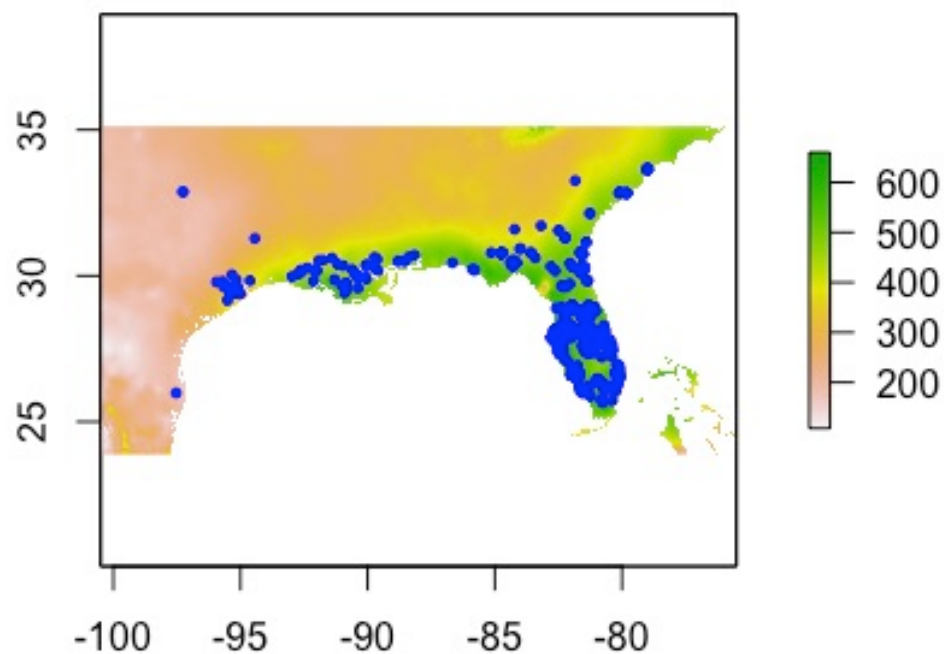


Figure 4. Precipitation of warmest quarter (in mm) variable with observed presences

PRECIPITATION LEVELS PREDICT THE DISTRIBUTION OF AN INVASIVE APPLE SNAIL (*POMACEA MACULATA*) THROUGHOUT THE SOUTHEASTERN UNITED STATES

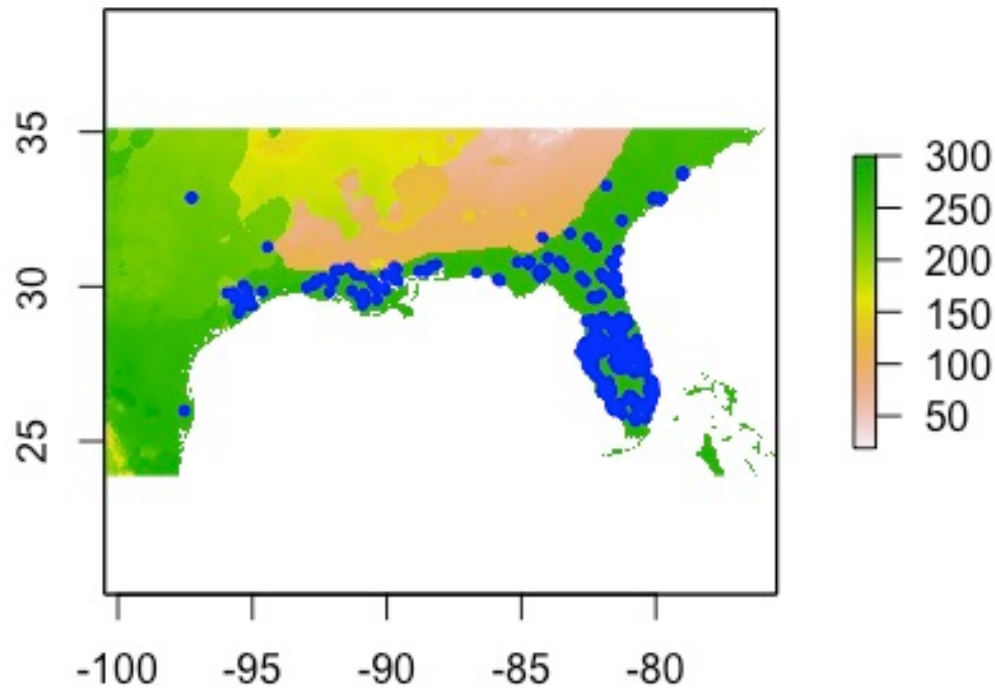


Figure 5. Mean temperature of wettest quarter (in °C *10) variable with observed presences

References

- Barbet-Massin, M., Jiguet, F., Hélène Albert, C., & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: how, where and how many?. *British Ecological Society*, 3(2), 327-338. doi: <https://doi.org/10.1111/j.2041-210X.2011.00172.x>
- Byers, J., McDowell, W., Dodd, S., Haynie, R., Pintor, L., & Wilde, S. (2013). Climate and pH Predict the Potential Range of the Invasive Apple Snail (*Pomacea insularum*) in the Southeastern United States. *PLoS ONE*, 8(2): e56812. doi:10.1371/journal.pone.0056812
- Carlsson, N., Brönmark, C., & Hansson, L. (2004). Invading herbivory: The golden apple snail alters ecosystem functioning in Asian wetlands. *Ecology*, 85(6), 1575-1580. doi: <https://doi.org/10.1890/03-3146>
- Cattau, C., Fletcher, R., Kimball, R., Miller, C., & Kitchens, W. (2018). Rapid morphological change of a top predator with the invasion of a novel prey. *Nature Ecology & Evolution*, 2(1), 108-115. doi: <https://doi.org/10.1038/s41559-017-0378-1>
- Cattau, C., Fletcher, R., Reichart, B., & Kitchens, W. (2016). Counteracting effects of a non-native prey on the demography of a native predator culminate in positive population growth. *Ecological Applications*, 26(7), 1952-1968. doi: <https://doi.org/10.1890/15-1020.1>
- Cattau, C., Martin, J., & Kitchens, W. (2010). Effects of an exotic prey species on a native specialist: Example of the snail kite. *Biological Conservation*, 143(2), 513-520. doi: <https://doi.org/10.1016/j.biocon.2009.11.022>

Rawlings, T., Hayes, K., Cowie, R., & Collins, T. (2007). The identity, distribution, and impacts of non-native apple snails in the Continental United States. *BMC Evolutionary Biology*, 7(9). doi: <https://doi.org/10.1186/1471-2148-7-97>

Robertson, S. (2012). Potential threats of the exotic apple snail *Pomacea insularum* to aquatic ecosystems in Georgia and Florida. *University of Georgia Theses and Dissertations*.