Environmental Applications Term Project:

Muscovy duck suitability using Maxent modeling

Nicholas Stanley

School of Geosciences, University of South Florida

GIS 6306 Environmental Applications of GIS

Dr. Lucy Deba Enomah

April 11th, 2025

Introduction

The movement of animals has traversed scales throughout history, from individuals moving within habitats to far migrations between seasons. Movement supported by human activity, such as trade, has also brought animals along with them into new habitats. These introductions often result in releasing these animals into the wild, creating invasive species for the environment. Invasive species can represent a significant problem, as they compete for the same resources as native species, with little to no predators or consumers to keep their population in check. To reduce and prevent invasive species impacts on the environment, methods to track and predict invasive species locations can assist in management options for these issues.

One of these species is the Muscovy duck (*Cairina moschata*), a species native to Central and South America, commonly near agricultural and wetland areas (Woodyard & Bolen, 1984). Movements of the ducks have caused their population to be found far beyond their native range, resulting in population growth in locations such as Hillsborough County, Florida. Ducks are common throughout the city, especially in active suburban areas like college campuses (Downs et al., 2017). The increase in the duck population raises concern for the native wildlife. Not only do the ducks create an increase in competition for resources, but they also have the potential to be disease carriers that could affect other species (Elbestawy et al., 2019).

Understanding the potential dangers of an introduced Muscovy duck population in places like Hillsborough requires proactive engagement to protect our ecosystems. While they can be easily seen in neighborhoods, the wider extent to which the ducks are located has yet to be examined. This study aims to bring more understanding of Muscovy duck distribution in Hillsborough County with the use of suitability modeling. Suitability modeling helps to locate spatial patterns of species location but also understand how individual environmental factors contribute to their occurrence.

Methods

For this study, I used the machine-learning software Maxent (Phillips et al.), made for predicting wildlife locations probabilities over a study area. Maxent works from past occurrence data of the species and environmental factor layers. Environmental layers are weighed amongst each other and within them to assemble the best prediction combinations based on the species data. The result of using this software helps to locate suitable locations in the study area, as well as determine which factors impact the probability of an area to be suitable and by how much. With both suitable locations and important environmental factors, prediction of Muscovy habitats can being more understanding to their distribution in Hillsborough County.

Hillsborough County (fig. 1) provides a diverse area of environments and land cover types. With a dense urban center surrounding the city of Tampa that moves to a gradient of suburban and rural areas, Muscovy ducks have a range of areas to occupy. The urban and suburban sprawl extends from the county's western half to the center, followed by agricultural and preserved lands extending into the east. Continued suburban development has grown fast in recent years, extending into the rural areas in southern and eastern parts of the county. Water is an integral component of

Hillsborough's landscape, not just limited to the Hillsborough River and Tampa Bay but also with the lakes and retention ponds studded throughout the county.

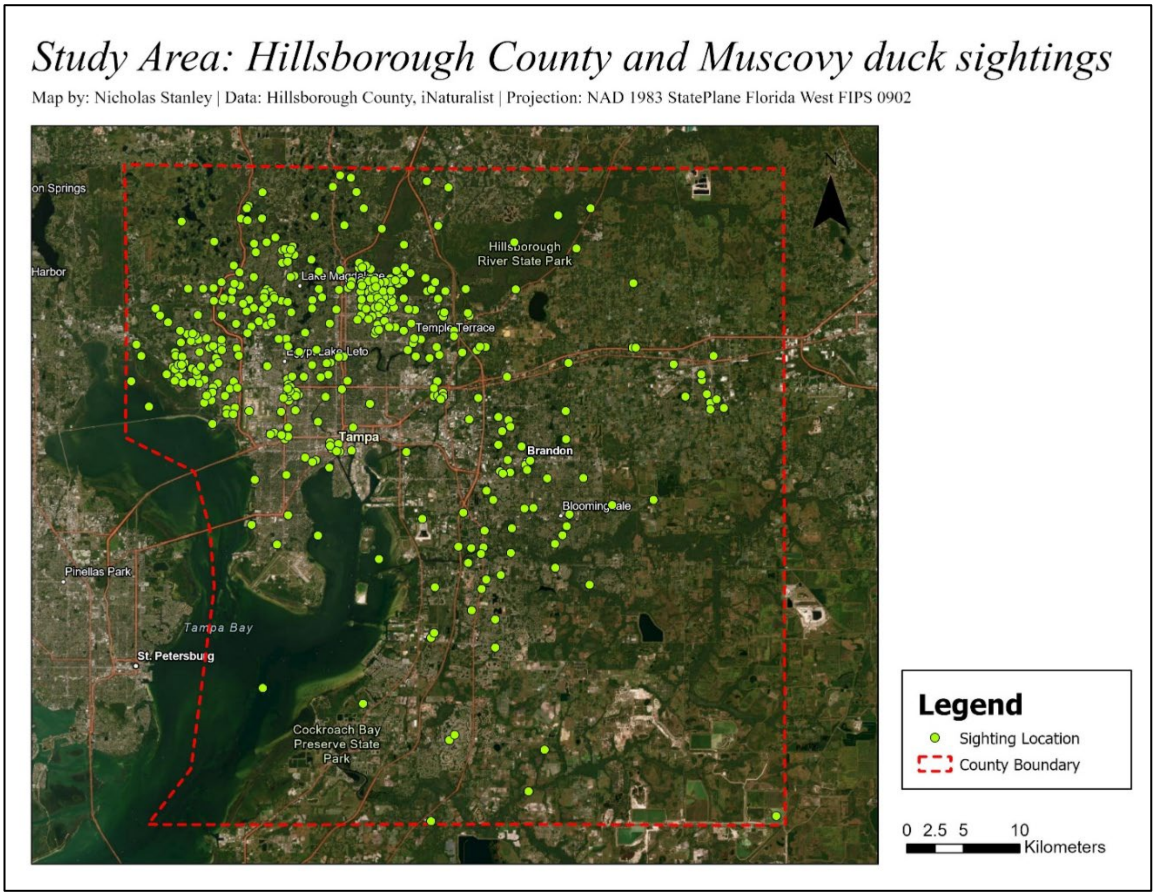


Figure 1 – Study Area: Hillsborough County and Muscovy duck sightings

We can also start looking at the occurrence data of the Muscovy duck (fig. 1). Residents report duck occurrences through iNaturalist, a website where users can report species sightings. The recorded points come from the start of 2020 to the end of 2024, providing a five-year period of sightings to work with. Originally, 1526 observations were reported in Hillsborough County during this period. To help reduce the over-sighting of occurrence points for a single area, past sighting locations were processed through ArcGIS Pro's (Esri, 2023) Delete Identical tool to remove random points within 250 meters of each other. This process leaves at least a 250-meter distance between each point to reduce nearby clustering while maintaining areas with higher sighting rates. This process resulted in 402 remaining occurrences to be used in the analysis.

In addition to the occurrence data, Maxent uses environmental layers for its predictions. These layers reflect various aspects of the environment to evaluate suitable areas of duck locations. Each layer is processed into raster files of the same cell size and grid, allowing Maxent to overlay and process the data. To determine which environmental years to use, I took inspiration from Schaaf et al.'s study of using Maxent to assess Muscovy duck habitats in Argentina (2018). In their study, six of the twenty-three environmental variables scored a 2% or higher contribution to the model. These layers include distance to waterlines, mean temperature of coldest quarter, temperature seasonality, temperature range, distance to waterbodies, and slope. My model used equivalent datasets for the study area, with the addition of elevation, agricultural cover, and woody wetland cover. Hillsborough County's range of low-lying coastal areas and inland hills potentially impact suitability through how elevation may represent different parts of the landscape. Additionally, agricultural and woody wetland land cover was included as individual layers, as they serve as both areas of cover and food sources for the ducks (Woodyard & Bolen, 1984). Preemptively, a correlation matrix (fig. 2) was created using Band Collection Statistics in ArcGIS Pro (Esri, 2023). I aimed for correlation values less than |0.8|, for which all reported below the value for any combination of environmental layers.

Row Labels	distance_wtrbds	slope	woody_wetlands	cultivated_crops	dem	annual_temp_range	temp_seasonality	coldest_quarter_avg	distance_to_streams
distance_wtrbds	1	0.00795	0.27007	0.08803	0.32581	0.27379	-0.07935	-0.14425	-0.254
slope	0.00795	1	0.04905	-0.0307	0.1416	0.03025	-0.05679	0.05493	-0.13508
woody_wetlands	0.27007	0.04905	1	-0.0613	0.06276	0.11787	0.01728	-0.31105	-0.3028
cultivated_crops	0.08803	-0.0307	-0.0613	1	0.1656	0.08988	-0.15357	0.06275	-0.13372
dem	0.32581	0.1416	0.06276	0.1656	1	0.66764	-0.34301	-0.10358	-0.33025
annual_temp_range	0.27379	0.03025	0.11787	0.08988	0.66764	1	-0.11633	-0.37079	-0.01559
temp_seasonality	-0.07935	-0.05679	0.01728	-0.15357	-0.34301	-0.11633	1	-0.67463	0.1275
coldest_quarter_avg	-0.14425	0.05493	-0.31105	0.06275	-0.10358	-0.37079	-0.67463	1	-0.03941
distance_to_streams	-0.254	-0.13508	-0.3028	-0.13372	-0.33025	-0.01559	0.1275	-0.03941	1

Figure 2 – Environmental layers correlation matrix

Data was collected and processed for all environmental layers into the same grid cell size and dimensions. The data was projected to NAD 1983 StatePlane Florida West FIPS 0902, resampled to 30 meters, and clipped using the Hillsborough County boundary (2024) to easily maintain coherency. The 30-meter cell size was chosen based on the LULC data (NLCD, 2023) originally being at this size. I wanted to limit alteration to its grid cells and dimensions as they were the only qualitative data source we worked with. After the resampling and clipping to maintain consistency, the resulting cell size is 29.991 by 29.992 meters. While not exactly true to the 30-meter cell size chosen earlier, it was best to settle at this size as clipping with maintaining extent allowed the environment to be in the same cell size and dimensions for Maxent. Any further resampling may cause over-processing of the data that could further distance itself from real-world conditions.

For the individual layers themselves, data was compiled from different sources for Maxent. The bioclimatic factors, mean temperature of coldest quarter, temperature seasonality, temperature range, and temperature seasonality, were compiled from 30 second resolution data from WorldClim (2020). Elevation data in meters was retrieved from the USGS (2022), which was then calculated for slope. Distance to waterlines was calculated using euclidean distance of stream location data (Hillsborough County, 2023). Agricultural and woody wetland areas were calculated from land cover data

(NLCD, 2023), each calculated using a focal cell average of either cover type within a 100-to-400-meter annulus. Using the annulus would also account for nearby area of the same cover type for ducks to roam in. All calculations and processing of the environmental layers were done in ArcGIS Pro (Esri, 2023), then exported to ASCII raster files for use in Maxent.

Most of the MaxEnt settings were kept default, with a logistic output used for this study. The logistic output was chosen as it is more suitable for impermanent locations, such as sightings from moving ducks. In addition, 75% of the past occurrence was used for training the model, while 25% was set aside for testing afterwards. Jackknife test and response curve graphs were also produced to evaluate the performance of individual layers. The jackknife test will show how much individual layers contribute to prediction, providing a sense of which environmental factors lean more towards preference by the ducks. The individual response curves will show which values within a layer lend more towards prediction. An ROC graph will also be used to test the model's ability to predict duck suitability by comparing the rate of which true positives are collected to false positives. Looking at the AUC value of the graph will help determine model performance, as a value of 1 representing best fit, 0.5 representing random chance, and 0 representing inverse fit.

Results

The resulting Maxent model (fig. 3) shows a range of probabilities for Muscovy duck suitable areas. The model located the higher suitable areas towards the northern neighborhoods of Tampa, center-west of the county. While some spatial connections exist, these highly suitable areas often exist in pockets. In contrast, the mid to lower suitable areas are more present, as the transition from the northern Tampa neighborhoods into the suburban and rural areas of the wider county shows this change. The further from the center of the county, the lower the suitability is likely to be. There also appear to be some low-probability areas with sightings. However, it is limited compared to the higher points count in more suitable areas.

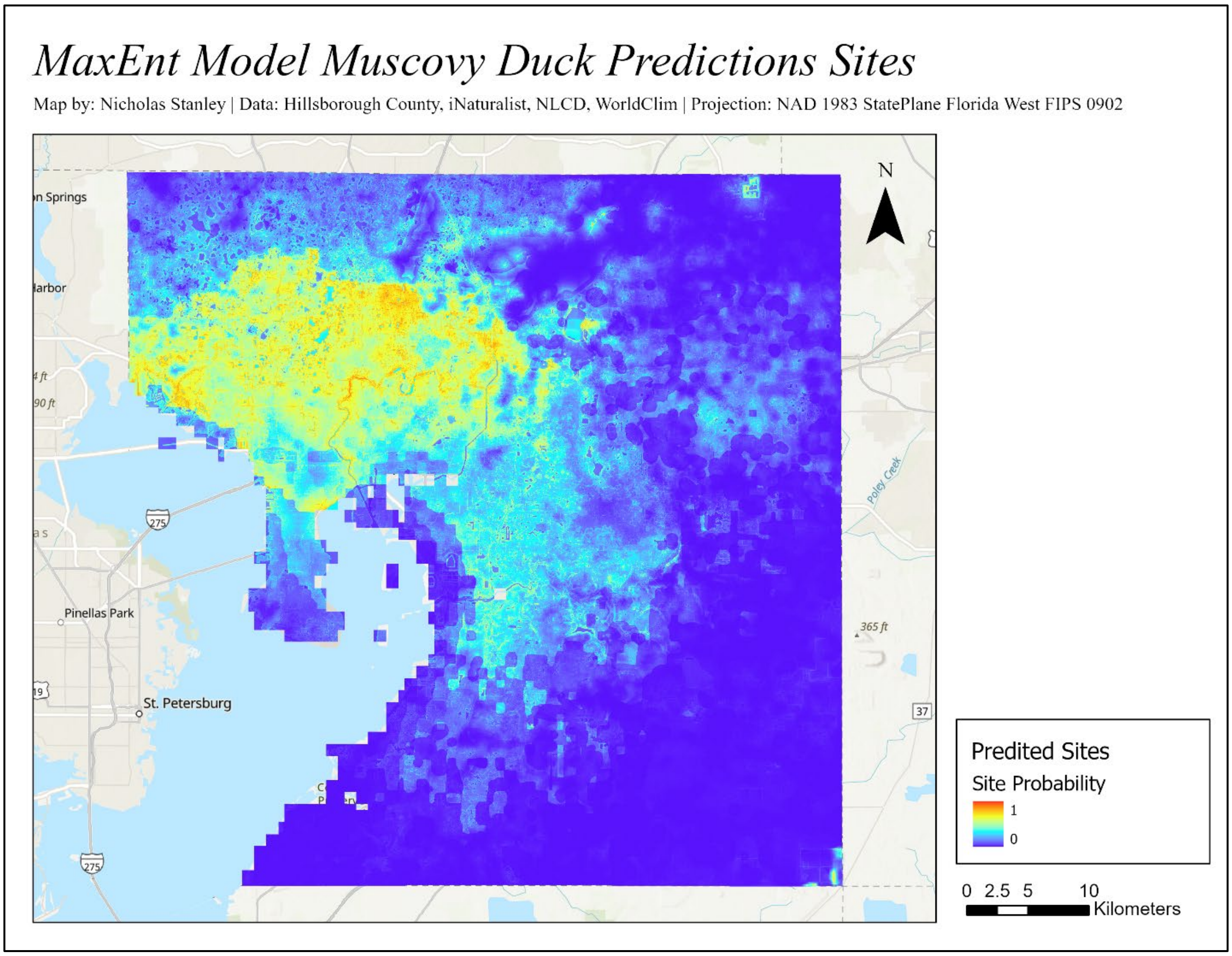


Figure 3 – Maxent model Muscovy duck prediction sites

The AUC graphs (fig. 4) have given insight into the accuracy of the model's training and testing data. The threshold for a highly accurate graph is typically an AUC of 0.9. The training data has an AUC of 0.891, while the test data has a value of 0.818. Both AUC values are also above the random prediction value of 0.5, meaning the model does better than randomly predicting suitability values but is not wholly considered to be highly accurate.

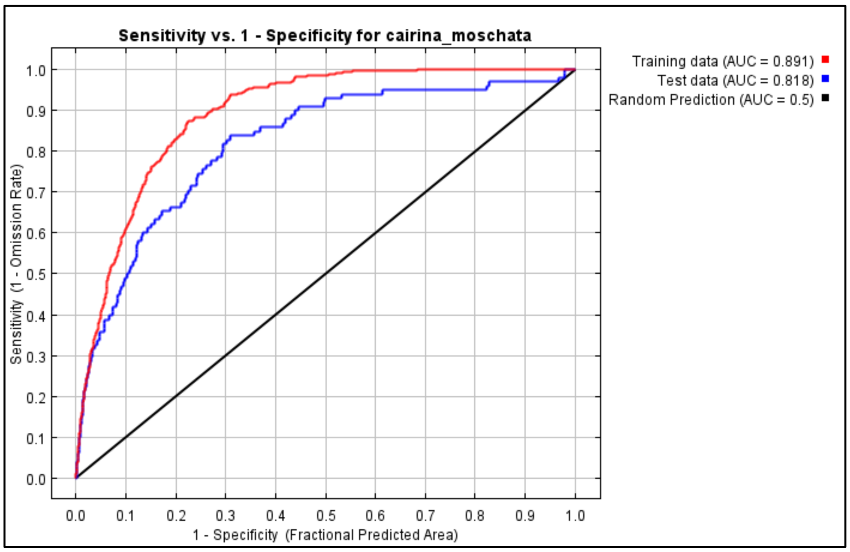


Figure 4 – AUC graphs for training and testing data

The response curves help to see how each environmental layer contributes to predictions. The graphs in Fig. 5 show how each of the different values contributes. They all have various shapes in the graphs, representing different values and how they affect probability. For instance, the dem response curve shows higher duck suitability at areas at sea level or 69 meters above it, with elevations in between being less suitable. Several of the response curves straddle at or around 0.5 probability, with small parts of the curve having a high probability. The two most linear curves, percentage of cultivated crops and distance to streams have slight differences in probability, with a spike in suitability towards the left side of the graph.

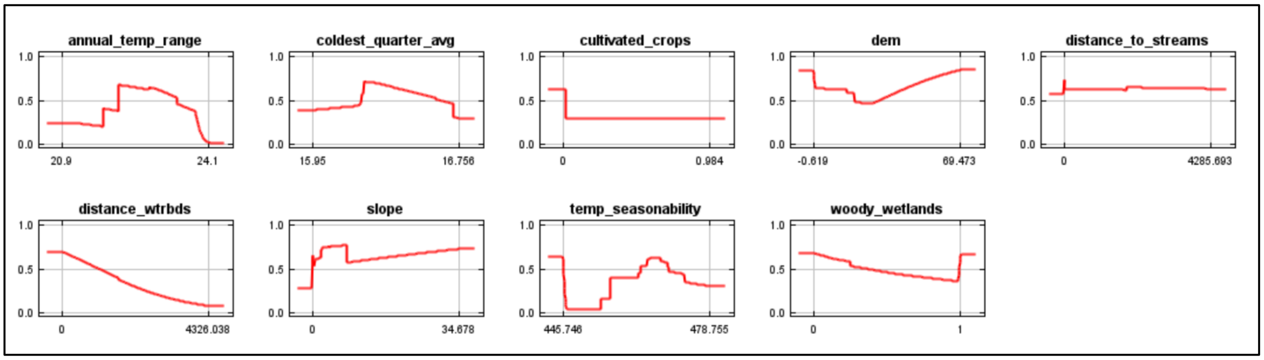
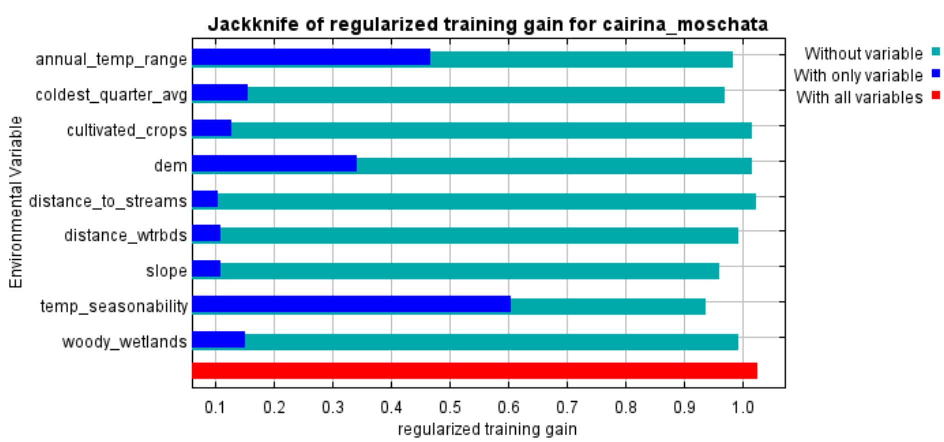
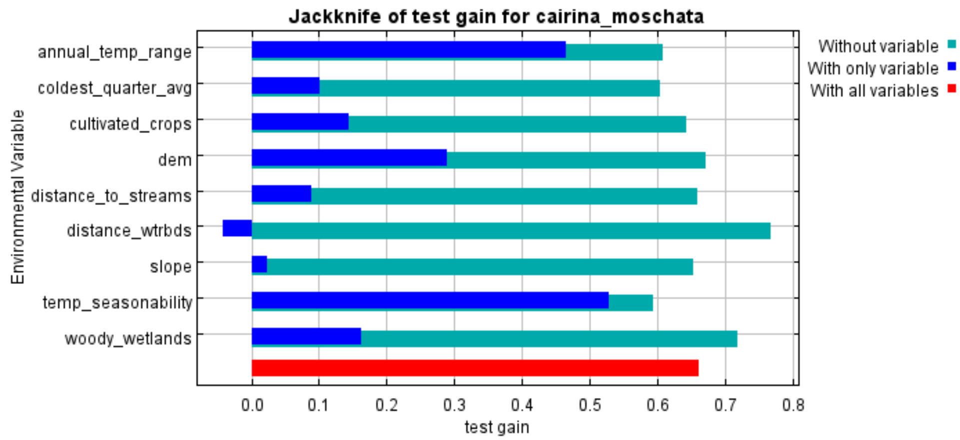


Figure 5 – Maxent response curves

For a further look at how the different environmental layers contribute to the Maxent model, we can look at the jackknife test graphs (fig. 6). The jackknife tests help to compare how the individual environmental layers contribute to the model by themselves and from their absence. Looking at the only variable portions of the graph, temperature seasonality contributed the most to predicting suitability, as it has the highest training gain than the other layers if just used by itself. Removing temperature seasonality from the other layers has a similar impact, decreasing the gain more than if others were to be removed. Annual temperature range and elevation follow in importance in terms of single variable contributions to gain.





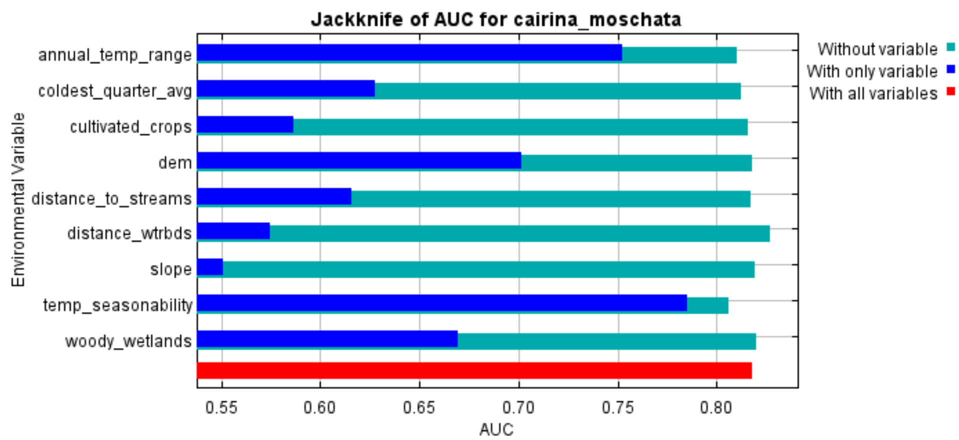


Figure 6 – Jackknife test results of the Maxent model

The importance of these three layers is representative in the other jackknife graphs. The test gain graph also shows the three layers' ability to stand independently for the test data. However, two new differences exist. When removed from the rest of the models, the layers of distance to water bodies and woody wetlands have helped to increase test gain more than other layers. Distance to water beds stands out especially, as it is the only layer to negatively predict suitability when used solely.

The last graph shows the AUC values of the model, providing insight into model accuracy. Temperature seasonality, annual temperature range, and elevation remain the most accurate layers if used independently, with temperature seasonality lowering accuracy the most if not used with the rest of the model. Model accuracy also increases with the removal of either distance to water bodies or woody wetlands. While most layers, if used by themselves, would create a low accuracy model, omission of any one environmental layer would keep the model AUC over 0.8, with some being close to the AUC of all variables being included.

Discussion

With the training data scoring an AUC of 0.891 and the testing data scoring 0.818, it presents our model as being accurate, but with room for improvement. Being below the 0.9 AUC threshold for being highly accurate does leaves consideration for what factors resulted in the different values in the jackknife tests. When comparing the contribution of our environmental layers to the previous study by Schaaf et al. (2018), there are many differences. From our tests, temperature seasonality is the highest contributor, with a 47.5% contribution (fig. 7). Schaaf et al. report that seasonality contributes only 11.8%. We see similar changes in other top contributors, with our model following contributors being the annual temperature range (18.7%), the coldest quarter temperature average (13.2%), and the slope (8%). In contrast, Schaaf et al.'s top four contributors area distance to streams (36.1%), coldest quarter temperature average (25.4), temperature seasonality (11.8%), and temperature annual range (10.3%).

Variable	Percent contribution	Permutation importance
temp_seasonability	47.5	47.2
annual_temp_range	18.7	17.4
coldest_quarter_avg	13.2	7.1
slope	8	7.9
woody_wetlands	4.2	6.7
dem	3.7	1.8
distance_wtrbds	3.2	6.5
cultivated_crops	1.3	4.8
distance_to_streams	0.2	0.6

Figure 7 – Variable contribution and permutation importance

While the order and percentage of the top contributing factors differ between Schaaf et al.'s model (2018) and ours, having three shared factors as top contributors may suggest that these are important factors to Muscovy duck suitability. I would be hesitant to name these the most important factors due to both models having different study areas, scales, number of environmental layers, kinds of environmental layers, and the percentages those layers contribute. Schaaf et al.'s top contributors have more balanced contribution percentages, while this study's model is dominated by the 47.5% contribution by temperature seasonality. Additionally, the highest contributing variable from their model is the distance to streams at 36.1%, while this version is the lowest at 0.2%, creating different models based on prediction influence.

Another important difference is how MaxEnt was used. Schaaf et al. (2018) used a batch modeling process where results were pulled from 100 model iterations. This study only used one rendition of the model, creating an imbalance in the certainty of model results. Schaaf et al.'s model can provide more certainty to their results, as they ran the model multiple times and incorporated the different results into one analysis. I did not have the opportunity to run a batch model, as only a single model was used to derive the analysis. Differences between using or not using a batch model can create significant gaps regarding certainty. One hundred renditions of a model that scored an AUC above 0.9 will be more trustworthy than one that scored at 0.8 AUC with only one rendition, even for different study areas.

In addition to the differences in layer contributions, the inclusion of woody wetlands and cultivated crops only contributed to smaller percentages. These layers were added to reflect potential land cover usage by the ducks. Lower contributions from both might suggest differences in preferred land cover between Muscovy ducks in their native habitats compared to those who are invasive. While known to occupy suburban areas (Downs et al., 2017), further investigation into land cover favorability as an invasive species may help to better tailor the model to the areas they prefer.

The choice of data and processing may influence the result, as a slight mishandling of the data or choice of environmental factors may impact the final output. Additionally, the occurrence data may not have been the most accurate representation of Muscovy duck locations as humans report them. While crowd-sourced data does allow for opportunities to work with a vast collection of duck sightings, these sightings are limited to the locations where people are more likely to be active. Fewer sightings in the eastern and southern parts of the county might not result from fewer ducks in these areas but potentially due to underreporting from the people who live there. Differences in reporting may influence the model's evaluation of the environmental layers as being more favorable to values that constitute urban and suburban landscapes rather than rural landscapes.

Discussion of the modifiable area unit problem (MAUP) is also worth consideration. One of the limitations of using Maxent is that all environmental layers need to be processed using the exact grid dimensions and cell size. While this does help to make sure the data is aligned for use, work from the operator is required to modify the data to fit these requirements. The bioclimatic data from WorldClim (2020) had a resolution of 30 seconds, or about 500 meters, once projected to the StatePlane projection. This size is much greater than the 30-meter goal established earlier, creating more generalization within the model even when resampled to a smaller cell size. Employing a method to smooth out more generalized data to incorporate Euclidean values and focal statistics better can help bring more accuracy to the model's output.

While parts of the study need more work to provide better accuracy to the model, it can help develop a framework for future Muscovy duck studies. Seeing some differences in how the modeling of native populations of ducks may differ from invasive species might require reconsidering what landscapes they rely on. Changing the current environmental layers in this model or including other factors may provide further insight into predictions for duck suitability. By operating on a county scale, different municipalities and stakeholders can use a model like this study to better direct their efforts towards managing invasive species.

Conclusion

The presence of invasive species like the Muscovy duck has become an increasingly present issue in our environment. Efforts to manage invasive populations are active, but remaining individuals still have the potential to inflict damage on local habitats and native species. Modeling suitable areas of invasive species is a step to better understand which locations they might be affecting and where efforts should be placed to address the issue.

Maxent provides users with flexibility and control over how to model invasive species suitability with accurate results. The ability to predict which environmental factors help facilitate occurrence data provides insight into the potential activity of the studied spices and the composition of their occupying landscape. While there are always areas to improve in how to create and use these models, having the means to further understanding of species suitability can help decision making to protect the environment.

References

Downs, J., Loraamm, R., Anderson, J. H., Perry, J., & Bullock, J. (2017). Habitat Use and Behaviours of Introduced Muscovy Ducks (Cairina moschata) in Urban and Suburban Environments (Cairina moschata) in Urban and Suburban Environments Suburban Sustainability, 5(1). https://doi.org/https://doi.org/10.5038/2164-0866.5.1.1028

Elbestawy, A. R., Ellakany, H. F., El-Hamid, H. S. A., Zedan, R. E., Gado, A. R., Sedeik, M. E., El-Hack, M. E. A., Saadeldin, I. M., Alowaimer, A. N., Ba-Awadh, H. A., & Swelum, A. A. (2019). Muscovy ducks infected with velogenic Newcastle disease virus (genotype VIId) act as carriers to infect in-contact chickens. Poultry Science, 98, 4441-4448. https://doi.org/http://dx.doi.org/10.3382/ps/pez276

Esri (2023). *ArcGIS Pro* (3.2.0). Esri. https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview

Hillsborough County. (2025). *County Boundary* (February 27, 2025) [Shapefile dataset]. https://gis2017-01-10t133755357z-hillsborough.opendata.arcgis.com/datasets/d9f095e8dde14ba79fb8136e71c1db26_0/explore?location=27.909016%2C-82.352855%2C10.90

Hillsborough County. (2023). *Streams* (November 13, 2023) [Shapefile dataset]. https://gis2017-01-10t133755357z-hillsborough.opendata.arcgis.com/datasets/a16bc3d23a1c44259b71ff2f35b15565_0/explore

National Land Cover Database. (2023). *Annual_NLCD_H25V17_LndCov_2023_CU_C1V0* [Raster dataset]. United States Geological Survey. https://earthexplorer.usgs.gov/scene/metadata/full/67096b68d90167f6/78353/

National Land Cover Database. (2023). *Annual_NLCD_H25V18_LndCov_2023_CU_C1V0* [Raster dataset]. United States Geological Survey. https://earthexplorer.usgs.gov/scene/metadata/full/67096b68d90167f6/78587/

National Land Cover Database. (2023). *Annual_NLCD_H26V17_LndCov_2023_CU_C1V0* [Raster dataset]. United States Geological Survey. https://earthexplorer.usgs.gov/scene/metadata/full/67096b68d90167f6/81863/

National Land Cover Database. (2023). *Annual_NLCD_H26V18_LndCov_2023_CU_C1V0* [Raster dataset]. United States Geological Survey. https://earthexplorer.usgs.gov/scene/metadata/full/67096b68d90167f6/82097/

Phillips, S. J., Dudík, M., & Schapire, R. E. Maxent software for modeling species niches and distributions (Version 3.4.1). http://biodiversityinformatics.amnh.org/open_source/maxent/

Schaaf, A. A., Gomez, D., Tallei, E., Rivera, L. O., Politi, N., & Cuyckens, G. A. E. (2018). Assessing distribution and conservation potential for the Muscovy duck

(Cairina moschata) in Argentina. Neotropical Biology and Conservation, 13(3), 183-191. https://doi.org/10.4013/nbc.2018.133.01

United States Geological Survey. (2022). *USGS 1 Arc Second n28w083 20221103* (November 11, 2022) [Raster dataset]. https://www.sciencebase.gov/catalog/item/63649c97d34ebe4425075148

United States Geological Survey. (2022). *USGS 1 Arc Second n29w083 20221103* (November 11, 2022) [Raster dataset]. https://www.sciencebase.gov/catalog/item/63649c97d34ebe4425075148

Woodyard, E. R., & Bolen, E. G. (1984). Ecolgoical Studies of Muscovy Ducks in Mexico. The Southwestern Naturalist, 29(4), 453-461. https://doi.org/https://doi.org/10.2307/3670998

WorldClim. (2020). *Bioclimatic variables* (elev 30s) [Raster dataset]. https://www.worldclim.org/data/worldclim21.html