Predictive Occurrence Modeling for Three Rare Plants Within the Croatan National Forest

Heather Moonier

Submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the requirements for the
Degree of Masters of Fisheries, Wildlife, and Conservation Biology

Raleigh, North Carolina

March 13, 2023

Under the Advisement of:
Dr. Lara Pacifici, Chair
Dr. Jodi Forrester
Dr. Christopher Moorman

Abstract

Rare plants are valuable indicators of areas of biological significance, and their persistence over time serves as a measure of ecological health and good land stewardship. Knowledge of the extent of rare plant distributions and the viability of rare plant populations is valuable information for assessing the risk of extirpation, future research, and management decisions. I constructed predictive occurrence models for three rare plants, *Platanthera integra*, *Pinguicula pumila*, and *Asclepias pedicellata*, using the Presence-only Prediction (MaxEnt) Tool in ArcGIS Pro 2.9. Explanatory variables used to model rare plant distributions included a LiDAR-based digital elevation model (DEM) of Croatan National Forest, soil series, and natural community data. Models were trained with Elemental Occurrence data obtained from the North Carolina Natural Heritage Program. The tool generated significant models with an area under the receiver-operator curve (AUC)>0.80 for all three species. The model for *Asclepias pedicellata* accurately predicted two new populations. However, observers failed to find individuals at the predicted sites surveyed for *Platanthera integra* and *Pinguicula pumila*. This may have been due to the timing of surveys or the accuracy of the models. The results of this study demonstrate that the Presence-only Prediction (MaxEnt) Tool may prove to be a valuable predictor of undiscovered populations of rare species, especially once the environmental filters and life history traits driving the distribution patterns of a rare focal species are better understood.

Key Words: Biodiversity, Atlantic Coastal Plain, Croatan National Forest, Carolina Bays, elemental occurrence, rare plants, Presence-only Prediction, MaxEnt, GIS, spatial analysis

Introduction

There is broad consensus across disciplines that natural communities with a diverse assemblage of functional traits are more resilient against disturbance, disease, and exotic species invasions (Eklöf, 2006; Leitão Rafael, 2016; Lohbeck, 2016; Tilman, 2014; Van der Plas, 2019). In general, the relative contribution rare plant species contribute to the ecological function of a community is poorly understood. Despite their relatively low abundance within a community, rare species have the potential to play an influential role in the ecological function of biologically diverse natural communities. Studies have shown that some rare plant species have a disproportionate effect on the persistence of other species, many having a mutualistic relationship with other rare flora or fauna. Thus, can be considered a keystone species of an ecological community (Jain, et al., 2013; Leitão Rafael, 2016). Depending on the species, their persistence can significantly influence ecological functions, such as increased resistance to secondary extinctions, nutrient cycling, and exotic species invasions (Lyons, 2001).

Furthermore, as valuable indicators of areas of biological significance, the persistence of rare plant communities over time serves as a measure of ecological health and good land stewardship. Conversely, rare plant declines may serve as an early warning sign of increased environmental stress on an ecosystem. Because rare plants are typically restricted to a narrow set of environmental conditions, it is also widely recognized that this constraint makes them more vulnerable to extirpation (Leitão Rafael, 2016). Although protected areas are a vital tool in preventing these losses, the IUCN estimates that only 15% of land is protected in national parks or nature reserves globally. Consequently, recreation is the greatest threat to rare plants in the United States, followed by land use change, invasive species, and shifts in fire regime (Hernández-Yáñez, 2016).

Studying rare plants to further our understanding of their natural history and response to environmental changes poses unique obstacles. Because they typically occur in patchy distributions, sometimes consisting of only a few individuals occurring sparsely over a broad geographic range, obtaining an adequate number of individuals for replicate sites for ecological studies can be a challenge. Nevertheless, knowledge of the extent of their distribution on the ground and the viability of those populations is essential for assessing the risk of extirpation and future research and management decisions. State Natural Heritage Programs collaborate with government agencies and non-profit conservation organizations across the United States and Canada to maintain known occurrences, or Element Occurrences (EO), of rare species on the NatureServe Network. 'Elements' can be plants, animals, or natural communities currently listed or species of special concern. EO data contains spatial information, status, characteristics, and the condition of those populations. They are used to weigh the ecological significance of natural areas and evaluate the potential environmental impacts of proposed conservation projects and development.

Effective management for conserving rare plant communities is dependent on accurate information about the spatial distribution of species present and the mosaic of natural communities across a landscape. In recent years, with the advancement of high-resolution LiDAR, more precise surface measurements of elevation, canopy cover, and vertical forest structure are making highly accurate characterizations across large spatial scales and remote areas more discernible to researchers (Farrell, 2013; Rebelo, 2010). Digital elevation models (DEM) can be derived from LiDAR point-cloud data and are compatible with many other data sources making them a versatile tool for analyzing complex datasets (Farrell, 2013).

The ability to accurately predict spatial distributions of undiscovered populations of rare species presents a well-documented challenge (Buechling, 201; Farrell, 2013; Leitão Rafael, 2016; Rebelo, 2010). The difficulties come largely from the limitations in data. More specifically, the number of presence observations of a rare species is often small and outdated. Data for possible explanatory variables can be challenging to acquire. Additionally, most predictive occurrence models require absence data, which are assumed to represent areas unsuitable for a focal species to persist (Buechling, 2011). Verifying absence locations can be problematic for species that are difficult to detect, for species with patchy distributions occurring over a broad geographic range, and for disturbance-dependent species that remain dormant through temporal fluctuations in environmental conditions, all of which may result in 'false absences' (Gogol-Prokurat, 2011; Rebelo, 2010;). These challenges may result in biases in predicting species distributions (Kiedrzyński, 2016; Zhang, 2020).

Technical advances in LiDAR-based DEMs and the emergence of more sophisticated machine learning for spatial data analysis provide powerful tools for predicting species distributions (Cho, 2021; Zhang, 2020). Maximum Entropy (MaxEnt) is a machine learning approach known for its predictive ability to perform better than other presence-only modeling approaches (Baldwin, 2009; Cho, 2021) in generating useful models of correlation between explanatory variables and low sample sizes of known presence with a lack of absence data (Fitzgibbon, 2022). Thus, the explanatory variables identified as significant from MaxEnt models can offer more insight into the environmental filters driving distribution patterns of species or natural communities (Cho, 2021). One example is the Presence-only Prediction (MaxEnt) Tool released with the ArcGIS Pro Version 2.9 update in November 2021. To my knowledge, this is the first time the Presence-only Prediction (MaxEnt) Tool in ArcGIS Pro 2.9 has been used to predict the presence of rare plant species.

I used the Presence-only Prediction (MaxEnt) Tool in ArcGIS Pro 2.9 to predict occurrences of three rare plants, *Platanthera integra*, *Pinguicula pumila*, *Asclepias pedicellata* occurring in Croatan National Forest. These three species were selected because they are state-listed, and several EOs within the national forest have not been reassessed since the early 2000s. One EO for *Asclepias pedicellata* had not been visited since 1989 despite that stand being actively managed under the Croatan National Forest Management Plan. The objectives of this study were to 1) reassess elemental occurrence data for three rare plants within the Croatan National Forest and 2) use geospatial analysis to identify potential areas where undocumented populations of the three focal plant species may occur. Based on observation notes from experts in the EO data received from the North Carolina Natural Heritage Program, and the assumption that *Pinguicula pumila* and *Asclepias pedicellata* are thought to occur within the pond pine woodlands/ wet pine flatwoods ecotone along the sandy rims of Carolina bays, my hypotheses were that 1) elevation would be the strongest environmental factor that can be used to predict the spatial distributions of these three species, and 2) that the pond pine woodlands/wet pine flatwoods ecotone would be an additional explanatory variable to predict the spatial distribution of *Pinguicula pumila* and *Asclepias pedicellata*. Other explanatory variables considered were soil series and natural communities. The data collected from this study serves as baseline presence data that can be used for future studies and monitoring programs for *Platanthera integra*, *Pinguicula pumila*, and *Asclepias pedicellata* within Croatan National Forest.

Methods

Study Area

Croatan National Forest is located within the North American Coastal Plain (NACP), a region with exceptional flora and fauna richness (5470 vascular plants, 120 amphibians, 230 reptiles, 204 mammals) and high levels of endemism (29.7% vascular plants, 36.7% amphibians, 34.8% reptiles, 34.3% mammals) (Noss, 2015). Given that the NACP has >1,500 endemic vascular plants and <30% of its historical range remains, in 2015 the NACP was recognized as the world's 36th biodiversity hotspot by the Critical Ecosystem Partnership Fund.

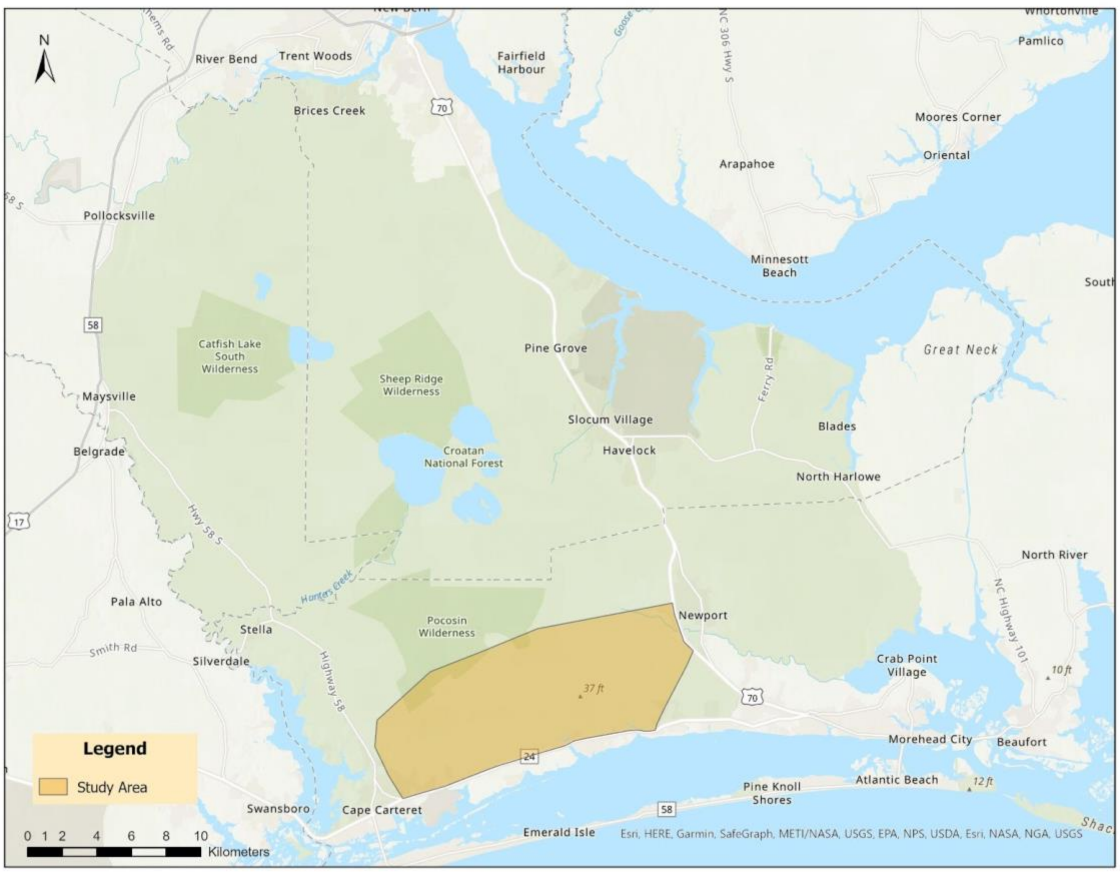

Fig. 1 Map of the study area within Croatan National Forest

The Croatan National Forest spans approximately 160,000 acres and is bordered by the Bogue Sound and two tidal rivers, the Neuse and White Oak. Defined by water and fire, the low topography of Croatan National Forest is characterized by pocosins, lime sinks, and Carolina bays, all of which foster a unique array of endemic and endangered flora and fauna. The study area contained a mosaic of natural communities, among the most notable are the variety of pine savannas and wet pine flatwoods. Today, less than 6% of these woodlands that once stretched across some 90 million acres of the extent of the NACP remains. Early accounts of the southeastern US, as described by the American botanist William Bartram in 1791, were long expanses of longleaf pine (*Pinus palustris*) or pond pine (*Pinus serotina*) woodlands with an open midstory and an understory comprised of grasses interspersed with an infinite variety of herbaceous plants. Frequent fire-return intervals (<3-yr), driven by climate and anthropogenic activity, maintained open savanna-woodland ecosystems across the NACP. When fire ignitions are limited, these biologically complex communities shift to closed-canopy mixed hardwood-pine forests, significantly reducing the flora and fauna diversity (Rother, 2022; White, 2016).

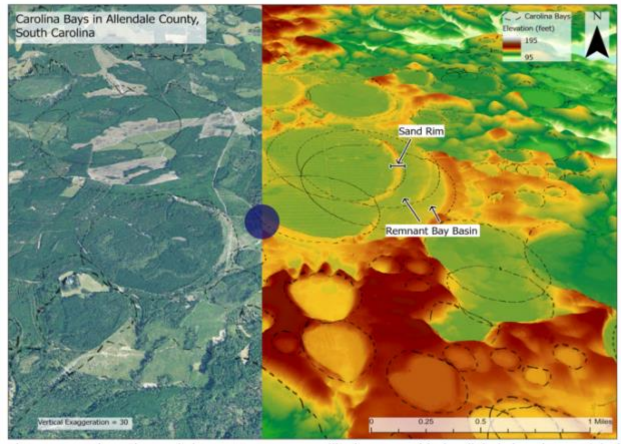

Fig. 2 Satellite imagery with the digital elevation model of Carolina bays. Source: SCDNR Geological Survey

The focal area for this study was approximately 28,000 acres located at the southernmost extent of the national forest and is frequented by hunters and recreationists. Within the area are several elliptical depressions known as Carolina Bays (Fig 2). Though there is no consensus on how these natural phenomena formed, the widely accepted hypothesis is from southwesterly winds on ponded water during the late Pleistocene. These depressions are relatively shallow, less than 3m deep, and vary in size ranging from 9m to almost 5km in diameter. What makes these depressions particularly perplexing is the long axes uniformly oriented in a northwest-southeast direction running perpendicular to the shoreline. The sandy rims along their southeast and northeast edges provide a novel set of environmental conditions that support a charismatic assemblage of flora, including Venus flytrap (*Dionaea muscipula),* sundews (*Drosera spp.),* pitcher plants (*Sarracenia spp.*), butterworts (*Pinguicula spp.*), and a variety of species in the family Orchidaceae.

Focal Species

The yellow fringeless orchid (*Platanthera integra*), dwarf butterwort (*Pinguicula pumila*), and savannah milkweed (*Asclepias pedicellata*) are three federally listed plants thought to occur within the pond pine woodlands/ wet pine flatwoods ecotone (Fig.3) associated with Carolina Bays in hydric soils. As small, herbaceous plants, the likelihood of an


Fig. 3 Photograph of the ecotone between pond pine woodlands(right) and wet pine flatwoods(left)

observer spotting one of these plants during a survey is greatest within one year of prescribed burning.

Fig. 4 and 5 Photograph of *Platanthera integra* observed in Croatan National Forest

P. integra occurs in wet pine savannas and meadows. They grow from 0.2-0.75m in height and bloom from July into early September. The inflorescence is a densely packed raceme consisting of small golden-yellow flowers less than 1cm. wide with a small spur present. The raceme averages 5-7cm in height. *P. integra* has an elongated wavy-margined lip lacking fringe, distinguishing it from the co-occurring *Platanthera cristata* (LeGrand, 2023). The American bumblebee (*Bombus pensylvanicus)* is the only documented pollinator of *P. integra* (NatureServe).

The smallest of the butterwort family in the region, *P. pumila*, is an insectivorous plant occurring in wet pine flatwoods. The basal rosette is 2.5-3.8cm in diameter. The leaves are elliptic and involute, curling slightly inward. *P. pumila* has a short flowering period, blooming from April to May. The corolla is white to pale violet with a spur at the base. The inflorescence is borne on erect leafless scapes reaching 15-18cm in height (LeGrand, 2023). Results from a recent study suggest *P. pumila* may be pollinated by various Hymenopterans and Dipterans in the family Syrphidae (Bradford, 2020).

Fig. 6 and 7 Photograph of *Pinguicula pumilla* observed in Croatan National Forest

A. pedicellata occurs in pine savannas and wet pine flatwoods. They grow from 15-30cm in height and bloom from May to July. The leaves are narrowly elliptic, about 3-4cm in length, and opposite along the stem. The greenish-yellow inflorescence is erect, pedicellate, and borne in loose terminal clusters with at 10-15 individual flowers. Unlike other milkweeds, the petals are not reflexed (LeGrand, 2023). There are no known pollinators of this species; however, it may be a host species for the Monarch, queen, and soldier butterflies (NatureServe).

Fig. 8 Photograph of *Asclepias pedicellata* observed in Croatan National Forest

Elemental Occurrence datasets

EO data for *P. integra*, *P. pumila*, *A. pedicellata*, wet pine flatwoods, pond pine woodlands, pine savanna, and pine scrub oak natural communities within Croatan National Forest were obtained as shape (shp.) files from the North Carolina Natural Heritage Program.

Environmental datasets

The United States Forest Service provided the high-resolution, LiDAR-based digital elevation model of Croatan National Forest and shp. file data of completed prescribed burns within the Croatan National Forest for the 2022 fiscal year.

The Web Soil Survey (WSS) contains soil series data collected from the National Cooperative Soil Survey. The data are maintained by the USDA Natural Resources Conservation Service (NRCS). I obtained the soil series shp. file for the extent of Croatan National Forest was obtained from the Web Soil Survey.

Presence-only Prediction (MaxEnt) model development and validation

Maximum Entropy is a machine learning approach known for its predictive ability (Baldwin, 2009; Cho, 2021) to generate models of correlation between explanatory variables and presence locations of a focal species or natural communities despite the constraint of low sample sizes and lack of absence data (Fitzgibbon, 2022). The Presence-only Prediction (MaxEnt) Tool in ArcGIS Pro uses a maximum entropy approach with known occurrence points and explanatory variables in the form of raster to estimate the probability of presence across a study area. The tool assigns background points to all non-presence areas; however, it is important to note that background points are not considered pseudo-absences in the model. They are used to compare the explanatory variables between presence locations and the study area to estimate the probability of presence. In the output features, the Presence-only Prediction (MaxEnt) Tool provides percentages of the four classification results; presence points correctly identified as presence, presence points misclassified as a low chance of presence, background points classified as potential presence, and background points that have a low chance of presence, to help researchers evaluate the validity of the model.

A caveat of MaxEnt is that it is prone to overfitting, resulting in predicted distributions clustering around known location points (Baldwin, 2009). To counteract this, the Presence-only Prediction (MaxEnt) Tool applies regularization in a way where all explanatory variables added to the model share a limited coefficient budget. As coefficients are reduced to satisfy the budget, explanatory variables with low coefficients are reduced to zero and removed from the model. With fewer explanatory variables considered, the model is less susceptible to overfitting. Thus, following the principle of parsimony, the simplest explanation of a phenomenon is usually the best (Phillips, 2006).

The area under the receiver-operator curve (AUC) is used to check the ability of the model to distinguish between positive and false-negative classes. AUC indicates the probability that a randomly chosen presence point is ranked higher than a randomly chosen background point (Cho, 2021). The closer an AUC value is to 1, the more reliable the estimate of the probability of presence is from the MaxEnt model. An AUC value of 0.50 is essentially a random guess that a point is either presence or absence. An AUC >0.70 is often considered acceptable, whereas an AUC >0.80 would be considered an excellent model.

Procedure

The Presence-only Prediction (MaxEnt) Tool requires presence data to be in points and all explanatory variables to be in raster. Because the EO data for the three focal species, natural communities, and soil series were received as polygon data all files needed to be converted to their respective file type prior to running the analysis.

I used the Create Random Points Tool to convert the EO data polygons for *P. integra*, *P. pumila*, and *A. pedicellata* to randomized points. First, a new field titled 'Individuals' was added to the attribute table. The average number of individual plants observed, documented in each EO record, was added as a numerical value to the 'Individuals' field. This field was then used to set the number of points that was generated within the constraint of each polygon. Minimum Allowance Distance controls how close together randomized points can be. For *P. integra*, and *A. pedicellata* the Minimum Allowance Distance was set to 5m and *P. pumila* was set to 3cm.

Because *P. pumila*, and *A. pedicellata* are thought to occur within the pond pine woodlands and wet pine flatwoods ecotone, I constructed a new layer representing the ecotone. I used the Buffer Tool to add a 16m buffer around the pond pine woodlands and wet pine flatwoods polygons. Then, I applied the Intersect Tool to the pond pine woodlands buffer and wet pine flatwoods buffer to create a new polygon representing the approximate spatial distribution of the ecotone.

Once the ecotone layer was constructed, I converted the soil series layer, each natural community layer (wet pine flatwoods, pond pine woodlands, pine savanna, and pine scrub oak), and the newly constructed ecotone layer to raster using the Polygon to Raster Tool. Cell Assignment was set to Center, and Cell Size was set to the LiDAR-based digital elevation model of Croatan National Forest. The Extract by Mask Tool was used last to reduce each raster to the extent of the Study Area.

I ran training models for each species using each explanatory variable in isolation first. Then, I paired explanatory variables with an AUC >0.70 to determine if a more robust model could be generated. To reduce sampling bias, Spatial Thinning was applied to each analysis. I set the Minimum Nearest Neighbor setting to 5m for *P. integra* and *A. pedicellata* and 1m for *P. pumila*. The Random Sampling Scheme was set to Random, and the Number of Groups was set to the default value of 3. Because the natural community and soil type layers were categorical, I applied a linear transformation for the Explanatory Variable Expansion (basis function) for each analysis.

The Output Prediction Raster provides the predicted probability of presence for each species within the extent of the study area. I converted the Output Prediction Raster to polygons using the Raster to Polygon Tool to get a more concise breakdown of the probability of presence that was generated by the most robust model for each species. Considering the likelihood of observing each species is greatest within one year of disturbance, I clipped the Output Prediction polygon layer for each species to the extent of the 2022 Completed Prescribed Burns layer using the Clip Tool. By performing this step, I was able to identify all locations within the Study Area where the predicted probability of presence was above 80% and the likelihood of observing each species was high. I then added the clipped Output Prediction polygon layer for each species to a new map layer with the original EO data for each species. I exported these maps to ArcGIS Online to be accessed using the ArcGIS Field Maps Application during field surveys.

Reassessing Elemental Occurrences (EO) and Surveying Predicated Sites

During the summer of 2022, I visited sites with a probability value of 0.80-1.00 two-three times each during the months indicated as the optimum survey windows by the Vascular Plants of North Carolina website (Table 1). Each EO dated prior to 2018 was also visited for reassessment at least two times.

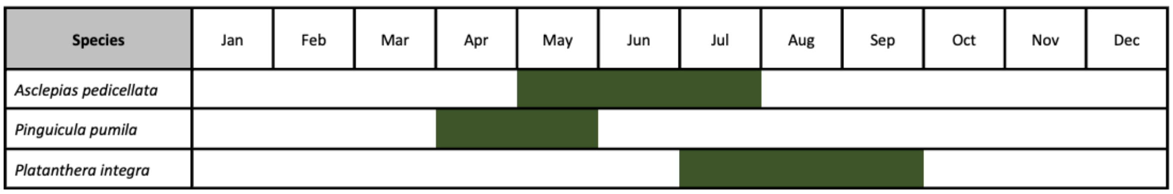

Table 1. Federally listed vascular plant in North Carolina and their optimum survey times. Source: Vascular Plants of North Carolina

Because the US Forest Service has currently only proposed protocols for rare plant surveys, I used the NC Department of Transportation Protocols. Prior to conducting surveys, permission to access the forest was granted by Andrew Walker, the USFS Botanist for Croatan National Forest. Known populations that had been surveyed in 2018 or later that were excluded from the EO reassessment plan were visited first to gain a search image of each species and its habitat. Andrew also accompanied me as the 2^{nd} observer for most surveys conducted within the study area.

During site visits, one to two observers started from the center of the polygon keeping a distance of at least 2.5m apart and walked towards opposite ends of the polygon. Observers walked the entire project study area and visually searched for each species' habitat. When an individual was observed, a GPS point was marked on the map in the ArcGIS Field Maps App. A counter was used to track the total number of individuals observed. For each EO reassessment, phenology, site characteristics, associate species, and descriptions of the condition of the population, if it was observed, was documented on the Rare Plant Form provided by the North Carolina Natural Heritage Program. Photographs of individuals were also taken as reference.

Results

The Presence-only Prediction (MaxEnt) Tool provided satisfactory results based on the seven explanatory variables. The most robust model for *P. integra* with an AUC of 0.8988 was generated by combining the LiDAR-based digital elevation model (DEM) of the Croatan National Forest and soil series (Fig. 9). In the training portion of this model, 100% of the known presence points were correctly identified as presence. The relative importance of environmental filters based on the regression coefficients indicated that elevation and soil series contributed most to the model when used in isolation. The environmental variable with the highest gain, thus providing the most useful information when used in isolation, was elevation with an AUC of

0.7658. Because there were no presence points within any of the natural communities, those models resulted in a failed analysis (Table 2). During surveys, only 2 individuals were observed at one of the two EOs reassessed. Observers failed to detect new populations when visiting selected sites with a probability value of 0.80-1.00.

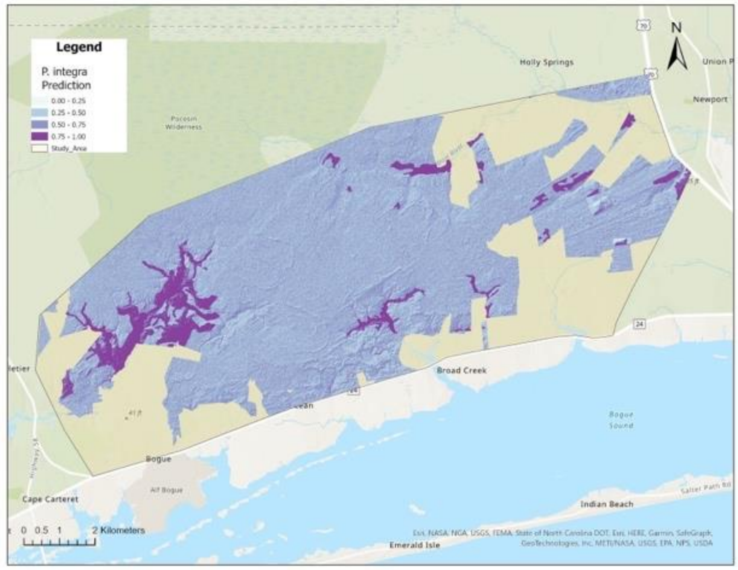

Fig 9. Output Prediction Raster for *Platathera integra*

The most robust model for *Pinguicula pumila* with an AUC of 0.9014 was generated by combining the LiDAR-based digital elevation model (DEM) of Croatan National Forest, soil series, and the pond pine woodlands and wet pine flatwoods ecotone (Fig. 10). In the training portion of this model, 86.45% of the known presence points were correctly identified as presence, and 13.55% of known presence points were misclassified as a low chance of presence. The relative importance of environmental filters based on the regression coefficients indicated that the pond pine woodlands/ wet pine flatwoods ecotone contributed most to the model when used in isolation. The environmental variable with the highest gain, thus providing the most useful information when used in isolation, was pond pine woodlands/ wet pine flatwoods ecotone with an AUC of 0.8561 (Table 3). During surveys, individuals were observed at five of the ten EOs reassessed. Observers failed to find new populations when visiting selected sites with a probability value of 0.80-1.00.

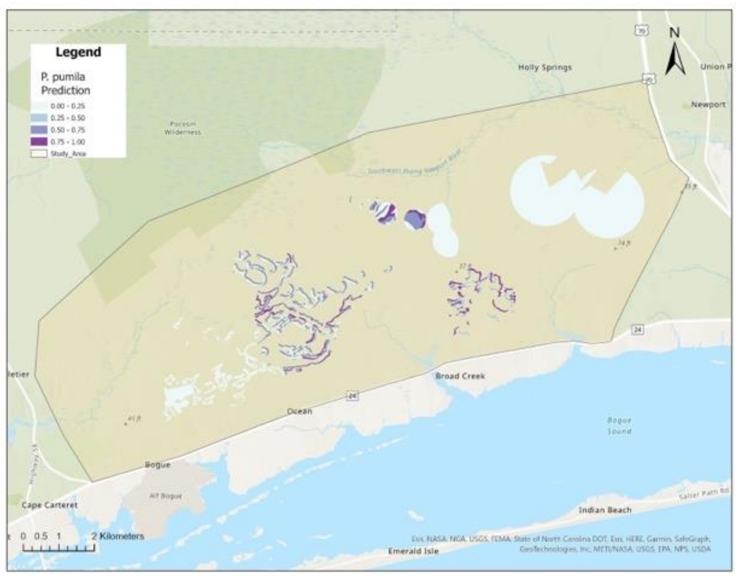

Fig 10. Output Prediction Raster for *Pinguicula pumila*

The most robust model for *Asclepias pedicellata* with an AUC of 0.8573 was generated by combining the LiDAR-based digital elevation model (DEM) of the Croatan National Forest, soil series, the wet pine flatwoods dataset, and the ecotone dataset (Fig.11). In the training portion of this model, 90.32% of the known presence points were correctly identified as presence, and 9.68% of known presence points were misclassified as a low chance of presence. The re lative importance of environmental filters based on the regression coefficients indicated that soil series and the wet pine flatwoods natural community contributed most to the model when used in isolation. The environmental variable with the highest gain, thus providing the most useful information when used in isolation, was soil series with an AUC of 0.8044 (Table 4). During surveys, individuals were observed at seven of the nine EO reassessed. Visiting selected sites with a probability value of 0.80-1.00 resulted in two new populations. One other population of *A. pedicellata* was discovered haphazardly while traveling between EO sites.

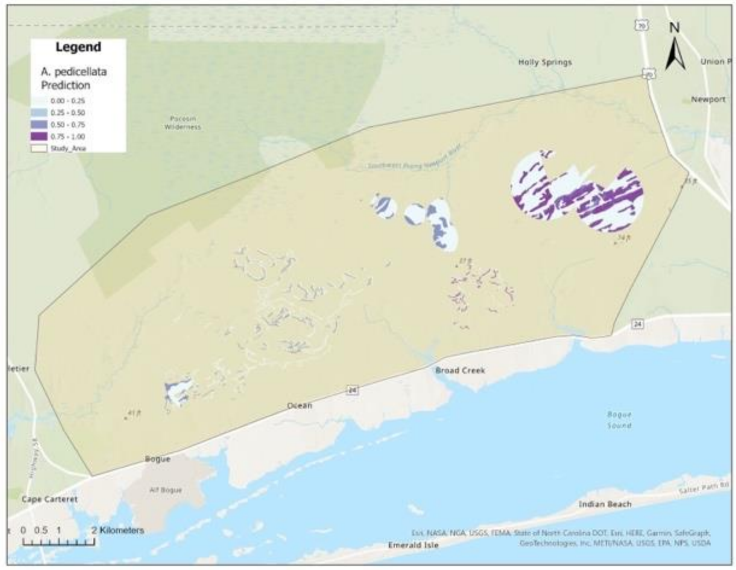

Fig 11. Output Prediction Raster for *Asclepias pedicellata*

Discussion:

The goal of this study was to explore the performance of the Presence-only Prediction (MaxEnt) Tool released with the ArcGIS Pro Version 2.9 update in November 2021 to predict occurrences of three rare plants, *P. integra*, *P. pumila*, and *A. pedicellata* occurring within Croatan National Forest. Based on observation notes from experts in the EO data I received from the North Carolina Natural Heritage Program, going with the assumption that these three species are thought to occur within the pond pine woodlands/ wet pine flatwoods ecotone found along the sandy rims of Carolina bays, my hypotheses were that 1) elevation would be the strongest environmental factor that can be used to predict the spatial distributions of these three species, and 2) that the pond pine woodlands/wet pine flatwoods ecotone would be an additional explanatory variable to predict the spatial distributions of *P. pumila* and *A. pedicellata*. To my knowledge, this is the first time the Presence-only Prediction (MaxEnt) Tool has been used to predict occurrences of rare plant species.

My hypotheses were only partially supported. Although elevation was used in the models for all three species, it was only significant for predicting the occurrences of *P. integra*. Additionally, the pond pine woodlands/ wet pine flatwoods ecotone was only a significant explanatory variable for *P. pumila*. Soil series was a significant explanatory variable in predicting occurrences for all three species, with all three species showing an association with the Murville, Rains, and Pantego soil series.

The results from surveys suggest other environmental filters are acting on these populations that were not considered, leading to the three considerations that arise in scientific studies: sampling bias, data limitations, and limitations in the analysis.

Failure to find populations of *P. integra* and *P. pumila* may have been attributed to observer bias. Surveying small-statured plants in dense vegetation is challenging, and it is possible that observers overlooked plants during surveys. Improper survey timing may have also been a factor, as the timing of flowering was more temperamental than expected from the information retrieved from the Vascular Plants of North Carolina Website. *Pinguicula pumila* is exceptionally difficult to observe when not in flower. Furthermore, species in the family Orchidaceae are known for being unpredictable as they are able to remain dormant through unfavorable temporal fluctuations in environmental conditions and do not always flower as expected.

Limitations in the data used is a second consideration. This became evident when reassessing EOs for *A. pedicellata* and *P. pumila*. Two EOs that were surveyed during the optimal flowering period appeared to have the appropriate environmental conditions to provide habitat for both species. However, survey efforts resulted in failure to find on both occasions the site was visited. Survey notes indicate that canopy cover and time since fire may be worth considering. Further analysis of the soil properties and composition of the Murville, Rains, and Pantego soil series would have also been worth exploring and may have resulted in more accurate predictions.

Another observation was that nearly all EOs exist in close proximity to roads. The possibility that the extent of the pond pine woodlands/ wet pine flatwoods ecotone, and the three focal species only occur around a road seems unlikely. My assumption is that the EO data are primarily based on areas accessible to biologists rather than areas determined by spatial analysis.

The Presence-only Prediction (MaxEnt) Tool requires presence data to be in points. Although points were randomized for each EO within the extent of the original polygon, exact locations were unknown. After observing these plants in the field, I am hesitant to say the randomized points reflect the true distribution of the population on the ground. As a result, the accuracy of known presence and the effect of the explanatory variables used in the model is likely skewed. There were also significant gaps in the soil series dataset, which may have further weakened the analysis.

There is also the issue with overfitting. Although the tool applies regularization to the model to account for this issue, many of the predicted sites for *Pinguicula pumila* were clustered around known presence points. Hence, overfitting of the model could have also occurred. Further adjustments to the Nearest Neighbor settings may have generated different results.

Finally, there are limitations with the Presence-only Prediction (MaxEnt) Tool. Initially, I intended to weigh natural communities based on their viability rating; however, because no absence points were included in the data set, natural communities had to be analyzed as a categorical variable rather than a quantitively variable. Another limitation is in instances where all explanatory variables added to the model are not represented within a grid where a present point is located results in those points being omitted from the model, which reduced sample size in an already limited dataset. In some training models, the omission rate was as high as 0.473. Additionally, the model would fail to run if no points occurred within grids where all explanatory variables were represented.

Conclusions:

The ability to accurately predict occurrences of undiscovered populations, namely rare species, presents one of the greatest challenges in ecology (Buechling, 201; Farrell, 2013; Leitão Rafael, 2016; Rebelo, 2010). It requires a sound understanding of a species' natural history and how environmental filters drive spatial distributions across gradients (Keddy, 2022; Wisz, 2013). Thus, knowledge of the extent of their distribution on the ground and the viability of those populations is essential information for assessing the risk of extirpation and future research and management decisions. With the advancements in remote sensing and more sophisticated machine learning capabilities for spatial data analysis, the ability to predict spatial distributions of rare plant species may continue to improve with time and continued research.

The results of this study demonstrate that the Presence-only Prediction (MaxEnt)Tool may prove to be a valuable predictor of undiscovered populations of rare species for land managers and conservation biologists when GPS points are available for known presence of individuals and the environmental filters and life history traits driving the distribution patterns of a focal species are better understood. When the most robust explanatory variables are known, less time can be spent haphazardly searching for rare plant populations on the ground, allowing rare plant surveys to be executed with a more methodical approach. Preparing the data and running the models in the Presence-only Prediction (MaxEnt) Tool was relatively intuitive, and the Output Prediction Raster for each species was reasonably simple to interpret. This leads me to believe that this tool can be a powerful instrument for conservation biologists and managers of varying skill levels with spatial analysis, resulting in more effective management decisions in the future.

I plan to use the occurrence data collected during this study to reanalyze the prediction models to determine if more precise predictions can be made using GPS point data. I intend to visit those sites in hopes of adding additional elemental occurrence data for *P. integra, P. pumila*, and *A. pedicellata*.

Acknowledgments:

As an expression of gratitude and appreciation, I want to acknowledge that the Croatan National Forest is on the traditional territory of the Neusiok, Coree, and Lumbee tribes, who stewarded this land throughout generations. As Conservation Biologists, our goal is to study the natural world around us to understand how complex ecological communities function to apply what we learn to better conserve natural resources for future generations. Long before western scientists decided this was paramount, expert ecological knowledge of the natural world was already applied as an integral part of indigenous cultures. Regrettably, Indigenous knowledge was dismissed by our predecessors rather than embraced, leading to years of amending avoidable mistakes. We are privileged to be able to practice science and study ecology in academia. However, I think it is important to acknowledge that the modern Western scientific way is not absolute and is likely not even the best approach to conserving biodiversity.

Thank you to Dr. Lara Pacifici for your consistent guidance and encouragement as I navigated this change in career path and for allowing me to pursue a master's degree. Thank you to Dr. Christopher Moorman for pushing me beyond what I believed were my limits and igniting my interest in fire ecology, birding, and making the time to serve on my committee. A special thank you to Dr. Jodi Forrester, who was incredibly influential in the development of this project, for taking the time to serve on my committee and providing invaluable insight, wisdom, and consistent encouragement. Thank you, Dr. Alexander Krings, for teaching me about rare flora of North Carolina; your enthusiasm and compassion for plant conservation inspired me to pursue this project. And thank you for advising me to select plants near a beach if the project involves surveying in the summer in the Southeast. Thank you to Andrew Walker with the USFS for allowing me access to the Croatan National Forest and for taking time out of your busy schedule to share your passion and botanical knowledge for flora of the Atlantic Coastal Plain while accompanying me on several surveys. Your dedication to the preservation of Croatan is genuinely inspiring. Thank you to Brenda Wichmann and Michael Schafale for offering me direction for my initial proposal and providing the essential data to complete this project. Thank you to Jeff Essic for your assistance with ArcGIS Pro. A special thank you to my dog Duke for accompanying me on field surveys through wet savannas and canebreaks, you are my rock and truly tiny but mighty. Lastly, I would like to thank my friends and family for your patience and encouragement throughout this journey.

References

1) Baldwin RA. Use of Maximum Entropy Modeling in Wildlife Research. Entropy. 2009; 11(4):854-866. https://doi.org/10.3390/e11040854
2) Barbet-Massin, M., Jiguet, F., Albert, C. H., & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: How, where and how many? Methods in Ecology and Evolution, 3(2), 327-338. https://doi.org/10.1111/j.2041-210X.2011.00172.x
3) Bradford, John, and G. Rogers. "Small Butterwort (Pinguicula Pumila) in Its Natural Habitat." International Carnivorous Plant Society, Sep. 2020, https://doi.org/10.55360/cpn493.jb928 . Accessed 1 Feb. 2023.
4) Buechling, A., & Tobalske, C. (2011). Predictive habitat modeling of rare plant species in pacific northwest forests. *Western Journal of Applied Forestry*, *26*(2), 71-81. https://doi.org/10.1093/wjaf/26.2.71
5) Cho, E., Hunsaker, A. G., Jacobs, J. M., Palace, M., Sullivan, F. B., & Burakowski, E. A. (2021). Maximum entropy modeling to identify physical drivers of shallow snowpack heterogeneity using unpiloted aerial system (UAS) lidar. Journal of Hydrology (Amsterdam), 602, 126722. https://doi.org/10.1016/j.jhydrol.2021.126722
6) Eklöf, A., & Ebenman, B. (2006). Species loss and secondary extinctions in simple and complex model communities. The Journal of Animal Ecology, 75(1), 239-246. https://doi.org/10.1111/j.1365-2656.2006.01041.x
7) Farrell, S. L., Collier, B. A., Skow, K. L., Long, A. M., Campomizzi, A. J., Morrison, M. L., Hays, K. B., & Wilkins, R. N. (2013). Using LiDAR-derived vegetation metrics for high-resolution, species distribution models for conservation planning. Ecosphere (Washington, D.C), 4(3), art42-18. https://doi.org/10.1890/ES12-000352.1
8) Fitzgibbon, A.; Pisut, D., Fleisher, D. Evaluation of Maximum Entropy (Maxent) Machine Learning Model to Assess Relationships between Climate and Corn Suitability. Land 2022, 11, 1382. https://doi.org/10.3390/land11091382
9) Gamfeldt, L., Hillebrand, H., & Jonsson, P. R. (2008). Multiple functions increase the importance of biodiversity for overall ecosystem functioning. Ecology (Durham), 89(5), 1223-1231. https://doi.org/10.1890/06-2091.1
10) Godbold, J. A., & Solan, M. (2009). Relative importance of biodiversity and the abiotic environment in mediating an ecosystem process. Marine Ecology. Progress Series (Halstenbek), 396, 273-282. https://doi.org/10.3354/meps08401
11) Gogol-Prokurat, M. (2011). Predicting habitat suitability for rare plants at local spatial scales using a species distribution model. *Ecological Applications*, *21*(1), 33-47. https://doi.org/10.1890/09-1190.1
12) Hernández-Yáñez, H., Kos, J. T., Bast, M. D., Griggs, J. L., Hage, P. A., Killian, A., Loza, M. I., Whitmore, M. B., & Smith, A. B. (2016). A systematic assessment of threats affecting the rare plants of the united states. Biological Conservation, 203, 260-267. https://doi.org/10.1016/j.biocon.2016.10.009
13) Jain M, Flynn DF, Prager CM, Hart GM, Devan CM, Ahrestani FS, Palmer MI, Bunker DE, Knops JM, Jouseau CF, Naeem S. The importance of rare species: a trait-based assessment of rare species contributions to functional diversity and possible ecosystem function in tall-grass prairies. Ecol Evol. 2014 Jan;4(1):104-12. doi: 10.1002/ece3.915. Epub 2013 Dec 12. PMID: 24455165; PMCID: PMC3894892.

14) Keddy, Paul A., and Daniel C. Laughlin. A Framework for Community Ecology: Species Pools, Filters, and Traits. Cambridge University Press, 2022.
15) Kiedrzyński, M., Zielińska, K. M., Rewicz, A., and Kiedrzyńska, E. (2017), Habitat and spatial thinning improve the Maxent models performed with incomplete data, *J. Geophys. Res. Biogeosci.*, 122, 1359– 1370, doi:10.1002/2016JG003629.
16) LeGrand, H., B. Sorrie, and T. Howard. 2023. Vascular Plants of North Carolina [Internet]. Raleigh (NC): North Carolina Biodiversity Project and North Carolina State Parks. Available from https://auth1.dpr.ncparks.gov/flora/index.php.
17) Leitão Rafael P., Zuanon Jansen, Villéger Sébastien, Williams Stephen E., Baraloto Christopher, Fortunel Claire, Mendonça Fernando P. and Mouillot David (2016.) Rare species contribute disproportionately to the functional structure of species assemblagesProc. R. Soc. B.2832016008420160084. http://doi.org/10.1098/rspb.2016.0084
18) Lohbeck, M., Bongers, F., Martinez-Ramos, M., & Poorter, L. (2016). The importance of biodiversity and dominance for multiple ecosystem functions in a human-modified tropical landscape. Ecology (Durham), 97(10), 2772-2779. https://doi.org/10.1002/ecy.1499
19) Lyons, K. G., & Schwartz, M. W. (2001). Rare species loss alters ecosystem function - invasion resistance. Ecology Letters, 4(4), 358-365. https://doi.org/10.1046/j.1461-0248.2001.00235.x
20) Ma, X., Mahecha, M. D., Migliavacca, M., van der Plas, F., et al. (2019). Inferring plant functional diversity from space: The potential of sentinel-2. *Remote Sensing of Environment*, *233*, 111368. https://doi.org/10.1016/j.rse.2019.111368
21) Noss, R. F., Platt, W. J., Sorrie, B. A., Weakley, A. S., Means, D. B., Costanza, J., & Peet, R. K. (2015). How global biodiversity hotspots may go unrecognized: Lessons from the north american coastal plain. *Diversity & Distributions*, *21*(2), 236-244. https://doi.org/10.1111/ddi.12278
22) Phillips, J. D., & East Carolina University. (1997). A short history of a flat place; three centuries of geomorphic change in the croatan national forest. Annals of the Association of American Geographers, 87(2), 197-216. https://doi.org/10.1111/0004-5608.872050
23) Phillips, Steven J. , Robert P. Anderson, Robert E. Schapire. 2006. "Maximum entropy modeling of species geographic distributions." Ecological Modelling, 190: 231-259. PD
24) Rebelo, H., & Jones, G. (2010). Ground validation of presence-only modelling with rare species: A case study on barbastelles barbastella barbastellus (chiroptera: Vespertilionidae). The Journal of Applied Ecology, 47(2), 410-420. https://doi.org/10.1111/j.1365-2664.2009.01765.x
25) Rother, M. T., Patterson, T. W., Knapp, P. A., Mitchell, T. J., & Allen, N. (2022). A tree-ring record of historical fire activity in a piedmont longleaf pine (pinus palustris mill.) woodland in north carolina, USA. Fire Ecology, 18(1), 34. https://doi.org/10.1186/s42408-022-00161-4
26) Tilman, D., Isbell, F., & Cowles, J. M. (2014). Biodiversity and ecosystem functioning. Annual Review of Ecology, Evolution, and Systematics, 45(1), 471-493. https://doi.org/10.1146/annurev-ecolsys-120213-091917
27) Tilman, D., Knops, J., Wedin, D., Reich, P., Ritchie, M., & Siemann, E. (1997). The influence of functional diversity and composition on ecosystem processes. Science

(American Association for the Advancement of Science), 277(5330), 1300-1302. https://doi.org/10.1126/science.277.5330.1300
28) Tilman, D., Clark, M., Williams, D. et al. (2017). Future threats to biodiversity and pathways to their prevention. Nature 546, 73–81 https://doi.org/10.1038/nature22900
29) Tilman D., J. Knops, D. Wedin, P. Reich, M. Ritchie, E. Siemann (1997). The influence of functional diversity and composition on ecosystem processes. Science, 277 (5330) (1997), pp. 1300-1302
30) Van der Plas, F. (2019). Biodiversity and ecosystem functioning in naturally assembled communities. *Biological Reviews of the Cambridge Philosophical Society, 94*(4), 1220-1245. https://doi.org/10.1111/brv.12499
31) White, C. R., & Harley, G. L. (2016). Historical fire in longleaf pine (pinus palustris) forests of south mississippi and its relation to land use and climate. Ecosphere (Washington, D.C), 7(11), n/a. https://doi.org/10.1002/ecs2.1458
32) White, P. S. (2013). Derivation of the extrinsic values of biological diversity from its intrinsic value and of both from the first principles of evolution: Derivation of extrinsic and intrinsic values of biological diversity. Conservation Biology, 27(6), 1279-1285. https://doi.org/10.1111/cobi.12125
33) Wisz MS, Pottier J, Kissling WD, et al.(2012). The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. Biol Rev Camb Philos Soc. 2013 Feb;88(1):15-30. doi: 10.1111/j.1469-185X.2012.00235.x. Epub 2012 Jun 12. PMID: 22686347; PMCID: PMC3561684.
34) Zhang, C., Chen, Y., Xu, B., Xue, Y., & Ren, Y. (2020). Improving prediction of rare species' distribution from community data. Scientific Reports, 10(1), 12230-12230. https://doi.org/10.1038/s41598-020-69157-x

Appendix:

Data Description:

202201T1SMulti001.shp (Source: NC Natural Heritage Program)-polygon data of elemental occurrence data for *Platanthera integra*, *Pinguicula pumila*, and *Asclepias pedicellata*, wet pine flatwoods, pond pine woodlands, pine savanna, and pine scrub oak natural communities within Croatan National Forest. PCS: NAD 1983 StatePlane North Carolina FIPS 3200 (Meters)

Hill_Raw[File Geodatabase]>Croatan[Raster Dataset] (Source: USFS)- Digital Elevation Model of the extent of Croatan National Forest, North Carolina. PCS: GRS_1980_Lambert_Conformal_Conic

CroatanBoundary.shp- polygon data of Croatan National Forest. PCS: NAD 1983 StatePlane North Carolina FIPS 3200 (US Feet)

FY22AccomplishedCroatan\Croatan_FY22_Accomplished_Burns04052022.shp (Source:USFS)-polygon data of completed prescribed burns within Croatan National Forest. PCS: NAD 1983 StatePlane North Carolina FIPS 3200 (US Feet)

PNV_Croatan.shp (Source:USDA WebSoil Survey)- polygon data of soil series within Croatan National Forest. PCS: NAD 1983 StatePlane North Carolina FIPS 3200 (US Feet)

Service_Roads.shp (Source:USFS)- line data of roads within Croatan National Forest. PCS: NAD 1983 StatePlane North Carolina FIPS 3200 (US Feet)

Tables:

Table 2. Environmental variables considered in a predictive occurrence model for *P. integra*

Explanatory Variables	Points included in the model (4)	Omission rate	AUC	Classification results	Regression Coefficients
Lidar Only	4	0.0000	0.7658	100% correctly classified 0% misclassified	LIDAR_AOI_PROJECTRASTER= -0.3149
CroatanSoils_AOI ONLY	4	0.0000	0.6799	100% correctly classified 0% misclassified	CroatanSoils_AOI =0.0102
WetPineFW_AOI ONLY	failed			% correctly classified % misclassified	N/A
PondPineWL_AOI ONLY	failed			% correctly classified % misclassified	N/A
PineSav_AOI ONLY	failed			% correctly classified % misclassified	N/A
WPFP_PPWL_Ecotone_AOI ONLY	failed			% correctly classified % misclassified	N/A
Lidar + CroatanSoils_AOI	4	0.0000	0.8988	100% correctly classified 0% misclassified	LIDAR_AOI_PROJECTRASTER=-0.5241 CroatanSoils_AOI 49=1.2282 CroatanSoils_AOI 50=1.0025

Table 3: Environmental variables considered in a predictive occurrence model for *P. pumila*

Explanatory Variables	Points included in the model (3445)	Omission rate	AUC	Classification results	Regression Coefficients
Lidar Only	failed			% correctly classified % misclassified	N/A
CroatanSoils_AOI ONLY	failed			% correctly classified % misclassified	N/A
WetPineFW_AOI ONLY	1049	0.0000	0.7745	100% correctly classified 0% misclassified	WetPineFW_AOI 0=5.4285 WetPineFW_AOI 4=5.1552

PondPineWL_AOI ONLY	1349	0.0348	0.7270	96.52% correctly classified 3.48% misclassified	PondPineWL_AOI 3=-1.6513 PondPineWL_AOI 4=0.000
PineSav_AOI ONLY	797	0.0326	0.8380	96.74% correctly classified 3.26% misclassified	PineSav_AOI4=4.2067
WPFP_PPWL_Ecotone_AOI ONLY	781	0.0000	0.8561	100% correctly classified % misclassified	WPFP_PPWL_Ecotone 3= 5.2668 WPFP_PPWL_Ecotone 4= 5.1251 WPFP_PPWL_Ecotone 6= 5.3547
Lidar + PineSav_AOI + CroatanSoils_AO	797	0.1355	0.9271	86.45% correctly 13.55% misclassified	LIDAR_AOI_PROJECTRASTER= -0.0077 CroatanSoils_AOI 24= 1.3222 CroatanSoils_AOI 26= 2.7833 PineSav_AOI 4= 3.999
Lidar + CroatanSoils_AOI + WPFP_PPWL_Ecotone_AOI	781	0.0807	0.9014	91.93% correctly classified 8.07% misclassified	CroatanSoils_AOI 24= 1.6606 CroatanSoils_AOI 54= 2.1177 WPFP_PPWL_Ecotone 3=4.9555 WPFP_PPWL_Ecotone 4=4.8686 WPFP_PPWL_Ecotone 6= 5.4258

Table 4. Environmental variables considered in a predictive occurrence model for *A. pedicellata*

Explanatory Variables	**Points included in the model (325)**	**Omission rate**	**AUC**	**Classification results**	**Regression Coefficients**
Lidar Only	325	0.0000	0.5000	100% correctly classified 0% misclassified	N/A
CroatanSoils_AOI ONLY	40	0.1228	0.8044	87.72% correctly classified 12.28% misclassified	CroatanSoils_AOI 24= 1.2064 CroatanSoils_AOI 26= 3.3984 CroatanSoils_AOI 44= 3.6000 CroatanSoils_AOI 50= 3.4522 CroatanSoils_AOI 52= 3.0516 CroatanSoils_AOI 54= 2.8868
WetPineFW_AOI ONLY	148	0.4730	0.7975	52.7% correctly classified 47.3% misclassified	WetPineFW_AOI 0= 3.7085 WetPineFW_AOI 2= 0.5057 WetPineFW_AOI 4=1.2275
PondPineWL_AOI ONLY	66	0.0000	0.5246	100% correctly classified 0% misclassified	PondPineWL_AOI 3=-0.1357 PondPineWL_AOI 4=0.000
PineSav_AOI ONLY	13	0.0000	0.5848	100% correctly classified 0% misclassified	PineSav_AOI 4=0.5008
WPFP_PPWL_Ecotone_AOI ONLY	45	0.0889	0.6399	91.11% correctly classified 8.89% misclassified	WPFP_PPWL_Ecotone 1=0.5501 WPFP_PPWL_Ecotone 3=1.2736 WPFP_PPWL_Ecotone 4=-0.0803 WPFP_PPWL_Ecotone 5=1.0398 WPFP_PPWL_Ecotone 6=1.0399

Lidar+ WetPineFW_AOI	148	0.4730	0.7975	52.7% correctly classified 47.3% misclassified	WetPineFW_AOI 0= 3.7085 WetPineFW_AOI 2= 0.5057 WetPineFW_AOI 4=1.2275
Lidar+ PondPineWL_AOI	66	0.0000	0.5246	100% correctly classified 0% misclassified	LIDAR_AOI_PROJECTRASTER= -0.0028 PondPineWL_AOI 3= -0.1361 PondPineWL_AOI 4= 0.0000
Lidar+ PineSav_AOI	13	0.0769	0.6781	% correctly classified % misclassified	LIDAR_AOI_PROJECTRASTER= -0.4105 PineSav_AOI= 0.5110
Lidar+ CroatanSoils_AOI	40	0.1228	0.8044	% correctly classified % misclassified	CroatanSoils_AOI 24= 1.2064 CroatanSoils_AOI 26= 3.3984 CroatanSoils_AOI 44= 3.6000 CroatanSoils_AOI 50= 3.4522 CroatanSoils_AOI 52= 3.0516 CroatanSoils_AOI 54= 2.8868
Lidar + WetPineFW_AOI + WPFP_PPWL_Ecotone_AOI	31	0.0323	0.7059	96.77% correctly classified 3.23% misclassified	LIDAR_AOI_PROJECTRASTER= -0.1636 WetPineFW_AOI 0= 1.5396 WetPineFW_AOI 2= 1.3893 WPFP_PPWL_Ecotone 1=0.000 WPFP_PPWL_Ecotone 3= 2.1267 WPFP_PPWL_Ecotone, 6)= 0.3693
Lidar + WetPineFW_AOI + CroatanSoils_AOI	148	0.4730	0.8464	52.7% correctly classified 47.3% misclassified	LIDAR_AOI_PROJECTRASTER= -0.0127 CroatanSoils_AOI 24= -0.0694 CroatanSoils_AOI 26= 0.5863 CroatanSoils_AOI 52= 2.4358 CroatanSoils_AOI 54=1.5823 WetPineFW_AOI 0= 3.9123 WetPineFW_AOI 2=0.9968 WetPineFW_AOI 4=1.3222
Lidar + WetPineFW_AOI + CroatanSoils_AOI + WPFP_PPWL_Ecotone_AOI	31	0.0968	0.8573	90.32% correctly classified 9.68% misclassified	LIDAR_AOI_PROJECTRASTER= -0.1794 CroatanSoils_AOI 24= 0.2104 CroatanSoils_AOI 26= 2.8985 WetPineFW_AOI 0= 1.1555 WetPineFW_AOI 2=1.3283 WPFP_PPWL_Ecotone 1= 0.000 WPFP_PPWL_Ecotone 3= 2.1644 WPFP_PPWL_Ecotone 6=0.2641