

SEARCHING FOR THE WHITE FRINGELESS ORCHID, *PLATANThERA INTEGRILABIA*,
IN WESTERN NORTH CAROLINA

A thesis presented to the faculty of the Graduate School of Western Carolina
University in partial fulfillment of the requirements for the degree of Master of Science in
Biology

By

Maria “Mitzi” Harding

Director: Dr. Katherine Mathews
Professor of Biology
Biology Department

Committee Members: Dr. Beverly Collins, Biology
Dr. Joseph Pechman, Biology

December 2023

ACKNOWLEDGMENTS

Thank you to my adviser, Dr. Katherine Mathews, and committee members Dr. Beverly Collins, and Dr. Joe Pechman for your assistance and guidance. Thank you to Western Carolina University for funding support through the Summer Research Assistantship program. Much credit is due to Wesley Knapp for giving me the idea for this research project. I offer my sincerest gratitude to my husband, Josh, and daughters Naomi and Hazel for all of their support, sacrifice, and patience in helping me achieve this goal. I would like to dedicate my thesis to my friend and mentor, Dr. Tasha LaDoux, for inspiring my love for botanizing and rare plant conservation.

TABLE OF CONTENTS

LIST OF TABLES	iv
LIST OF FIGURES	v
ABSTRACT.....	vi
CHAPTER 1: INTRODUCTION.....	1
CHAPTER 2: METHODS.....	8
Initial habitat characterization.....	8
Species Distribution Model.....	9
Study Area.....	11
Presence and Absence Points	12
Environmental Predictor Variables	14
Running the model and cross validation.....	17
Model output analysis.....	19
Targeted Surveys.....	21
CHAPTER 3: RESULTS.....	24
Habitat characterization of known occurrences	24
Model output	25
Individual models	25
Ensemble model	28
Predicted Habitat in WNC.....	31
Targeted Surveys.....	32
CHAPTER 5: DISCUSSION.....	37
REFERENCES	45
APPENDIX A: R CODE	52
APPENDIX B: TARGETED SURVEY LOCATIONS	64

LIST OF TABLES

Table 1. List of associated species observed growing with <i>P. integrilabia</i> during initial site visits in 2022.	24
Table 2. Model performance metrics for the individual models and final weighted ensemble model.....	26
Table 3. Estimated relative importance of predictor variables in each of three habitat suitability models which were averaged to create the final ensemble output.....	27
Table 4. Proportion of the entire study area (%) and the proportion of the study area falling within western North Carolina (WNC) predicted in each habitat suitability index (HSI) bin from.	29
Table 5. Total area of each physiographic region included in the model study area and the proportion (%) of each physiographic region predicted to be suitable habitat (HSI >0.4).....	30
Table B1. Results of targeted surveys completed in August and September 2023.	64

LIST OF FIGURES

Figure 1. <i>Platanthera integrilabia</i> inflorescence.....	1
Figure 2. Element occurrences of <i>Platanthera integrilabia</i>	2
Figure 3. Infructescence of <i>Platanthera integrilabia</i> , at Starr Mountain, TN.....	4
Figure 4. Location of 8 element occurrences of <i>P. integrilabia</i> visited in 2022 for initial habitat observations..	9
Figure 5. Study area and presence-pseudoabsence data used to train the range-wide ensemble SDM for <i>P. integrilabia</i>	11
Figure 6. Habitat Suitability Index (HSI) predictions for <i>P. integrilabia</i> from three modeling algorithms, a) generalized additive model (GAM), b) generalized boosted model (GBM), and c) maximum entropy model (MaxEnt), and the final d) weighted ensemble model output.	25
Figure 7. Probability surface output for range-wide ensemble species distribution model.....	28
Figure 8. Predicted probabilities (habitat suitability index) for presence and absence points used to train the <i>Platanthera integrilabia</i> ensemble model output (n=654).	29
Figure 9. Predicted suitable habitat and targeted surveys for <i>P. integrilabia</i> within western North Carolina (habitat suitability index ≥ 0.4)	31
Figure 10. Conservation status of ensemble SDM predicted suitable habitat for <i>P. integrilabia</i> in western North Carolina (HSI>0.4).....	32
Figure 11. Average predicted habitat suitability (HSI) scores for 25 survey areas compared to the observed high, medium, or low quality habitat ranks based on observed characteristics during targeted surveys.	33
Figure 12. Survey site in Cherokee County with a predicted habitat suitability score of 0.96 being actively destroyed during targeted surveys.....	34
Figure 13. Wesley Knapp confirms new occurrence of the critically endangered <i>Sagittaria fasciculata</i> discovered during targeted surveys.	36

ABSTRACT

SEARCHING FOR THE WHITE FRINGELESS ORCHID, *PLATANATHERA INTEGRILABIA*, IN WESTERN NORTH CAROLINA

Maria “Mitzi” Harding

Western Carolina University (December 2023)

Director: Dr. Katherine Mathews

Rare species conservation has taken on an urgency in the face of unprecedented environmental degradation and biodiversity loss, and tools such as species distribution modeling are needed to increase the efficiency and efficacy of conservation efforts. *Platanthera integrilabia* (White Fringeless Orchid) is a rare orchid endemic to southern Appalachia and the surrounding area occupying a relatively narrow niche: non-alluvial sphagnum wetlands. The species was historically documented in North Carolina but has not been observed in the state since 1992. I hypothesized that some undiscovered populations may persist in western North Carolina in under-surveyed areas of suitable habitat. I used a weighted ensemble method to develop the first range-wide species distribution model (SDM) for *P. integrilabia* and performed targeted surveys based on model results. The SDM predicted approximately 6% of the species' range to be suitable habitat, including approximately 3% (788km²) of the study area within western North Carolina (AUC-PR=0.991). Surveys did not yield any new occurrences of *P. integrilabia*, but seven of the 25 sites contained vestigial patches of suitable habitat, one of which is recommended for incorporation into the reintroduction efforts currently underway in Henderson County, NC. The results of this study do not rule out the possibility that *P. integrilabia* may still persist in North Carolina. Results do indicate that patches of suitable habitat still remain and can

be predicted using a species distribution model. I also provide a framework for future iterations of the model intended to enhance the accuracy of predictions to more precisely direct future targeted surveys for *P. integrilabia*, and to aid in the conservation and restoration of this threatened orchid in its historical range.

CHAPTER 1: INTRODUCTION

Effective rare species conservation planning depends on reliable and comprehensive knowledge of the distribution and abundance of the species of concern. The only way to directly protect *in situ* populations is to know where they exist. The very characteristics qualifying a species as rare (narrow geographic ranges, habitat specificity, low population numbers) can lead to low detectability preventing thorough documentation of occurrences and adequate protection of populations (Lomba et al. 2010). However, the habitat specificity of rare species can offer an opportunity to target searches in areas containing the restricted niches the species inhabit. Species distribution models can be used to direct targeted searches for new populations of rare species by analyzing the unique values of environmental parameters correlated with known occurrences to then predict areas of suitable habitat. The use of species distribution modeling



Figure 1. *Platanthera integrilabia* inflorescence. (photo: M. Harding)

(SDM) to predict suitable habitat and target area searches has often led to the discovery of previously undocumented occurrences of rare species (McCune et al. 2020, Lawson et al. 2022, McCune 2016, Lomba et al. 2010).

The White Fringeless Orchid, *Platanthera integrilabia* (Correll) Luer, is a rare terrestrial orchid (Fig. 1) occurring within the Appalachian Plateaus, Coastal Plain, Piedmont, Blue Ridge, Valley and Ridge, and Interior Low Plateaus physiographic regions. It occurs in sandy acidic soils of partially shaded sphagnum bogs and other non-alluvial

(mainly precipitation fed) wetlands, generally found near level headwaters of small streams or seepage slopes (Shea 1992). *Platanthera integrilabia* was listed under the Endangered Species Act as a federally threatened plant in 2016 (81 FR 62826). The rare orchid has a relatively broad range in the southeast United States, but is patchily distributed throughout southeast Kentucky, eastern Tennessee, northern Georgia, northern Alabama, and northern Mississippi, and is historically known to occur in western North Carolina and northern South Carolina (Fig. 2). The majority of occurrences are concentrated on the Cumberland Plateau of Tennessee. *Platanthera integrilabia* is currently considered extirpated in North Carolina, as it has not been observed in the state since 1992 (USFWS 2021). Thorough surveys of the single documented occurrence in South Carolina (Greenville County) in August 2022 by myself and NHP botanist, Dr. Samantha Tessel, yielded no flowering individuals nor basal leaves of the species or commonly associated congeners.

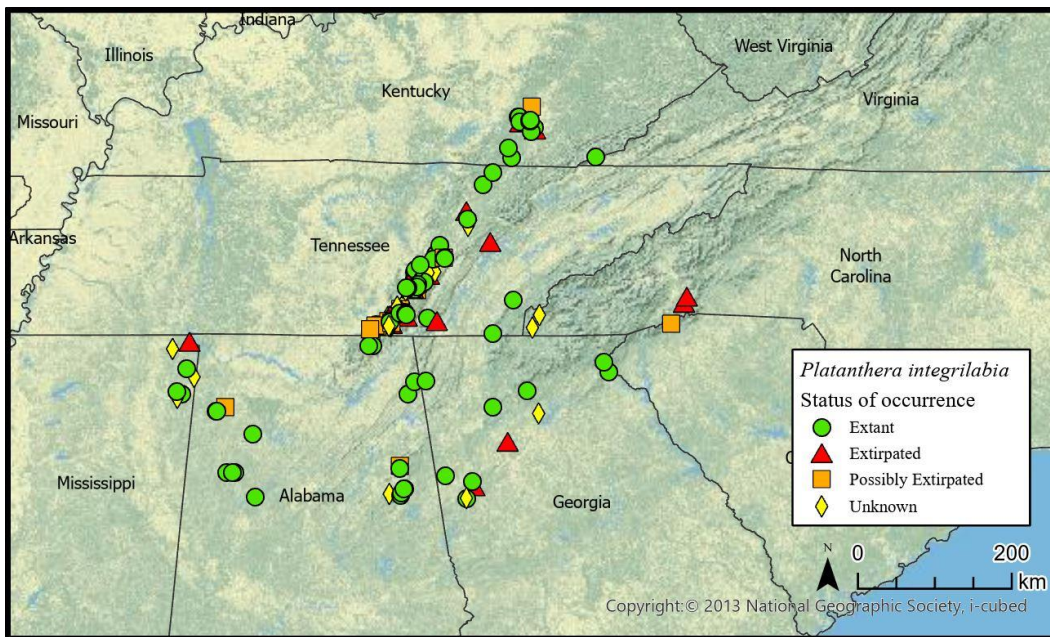


Figure 2. Range and element occurrence status of *P. integrilabia*. Possibly extirpated indicates no individuals were detected during the most recent survey, but suitable habitat remains. Unknown refers to sites where only vegetative *Platanthera* have been observed upon recent visits (species unconfirmed), or geographic locality of initial observation was too obscure to verify persistence. (ALNHP 2022, GDNR 2022, MDWFP 2022, NCDENR 2022, OKNP 2022, SCDNR 2022, TDEC 2022, USFWS 2021).

The patchy distribution of *P. integrilabia* is likely a function of the inherently rare mountain wetlands which it inhabits (Spira 2011). The wet and nearly flat headwaters within the southern Appalachian Mountains that *P. integrilabia* occupies are scarce to begin with, and land use changes due to development and agriculture are causing further decline (Murdock 1994). Historic records indicate approximately 122 known occurrences of *P. integrilabia*, only 86 of which distributed among 56 populations are now considered extant (USFWS 2023). The number of *P. integrilabia* individuals observed per occurrence ranges from 1 to just over 6000, the majority of which are reported to have low numbers of individuals. At the time of listing, 64% of known occurrences had fewer than 50 individuals ever observed, and 37% had fewer than 10 individuals (USFWS 2016). This high percentage of small and isolated populations could contribute to overall species decline due to Allee effects and the low seed viability found to be associated with small populations (USFWS 2021, Zettler and McInnis 1992). In the recent Species Status Assessment by USFWS, only 22% of 50 assessed populations were considered to have high or very high resilience (2021).

Low population numbers and declining occurrences of *P. integrilabia* may be due to a number of threats, both direct and indirect. The most pressing direct threats to *P. integrilabia* include land use changes, altered hydrology, poaching, herbivory by white tailed deer, feral hog activity, and native and non-native plant encroachment (USFWS 2021). In addition, like most orchid species, coexistence with specific obligate symbionts is required for survival and reproduction. Threats to these symbionts may pose an indirect threat to *P. integrilabia* populations, further increasing the orchid's sensitivity to environmental degradation. *Platanthera integrilabia* is mycotrophic, dependent on a highly specialized mutualism with the mycorrhizal fungus symbiont, *Epulorhiza inquilinia*, for seed germination and protocorm development

(Currah et al. 1997). Flowers of *P. integrilabia* have a long nectar spur, white color, strong evening fragrance, and wide-set pollinia suggesting specialized co-adaptation with nocturnal sphingid moths for pollination (Zettler et al. 1996). Although *P. integrilabia* is self-compatible, fruit set and seed germination are significantly greater in large populations where out-crossing is more frequent (Zettler & McInnis 1992). Therefore, inbreeding depression could account for low seed germination in small populations. In addition, the hydrologic balance of the headwater wetlands where *P. integrilabia* occur could be vulnerable to increased climate variability associated with global climate change (Levison et al. 2013).



Figure 3. Inflorescence of *Platanthera integrilabia*, at Starr Mountain, TN. Long nectar spurs persist distinguishing the species from other congeners into the fall. November 2, 2022 (photo: Ian Sabo)

Platanthera integrilabia can be difficult to detect considering its brief emergence and bloom period. During winter plants persist as a pair of subterranean protocorms, one large tuber from last season's growth and one small tuber which will lead to the coming season's growth (Shea 1993). A large strap-shaped basal leaf emerges in late spring, approximately 20 cm long and 3 cm wide. Based on morphology alone, the basal leaves are nearly indistinguishable from other closely associated congeners, adding to the orchid's low detectability. However, recent advancements in DNA barcoding have allowed researchers to distinguish the species from other congeners using leaf material (Aaron Floden, pers. comm.). *Platanthera integrilabia* is

most readily distinguished from other *Platanthera* species in the field once an elongated stem, up to 60 cm tall, produces an inflorescence with few-20 flowers in late July through mid-August. After approximately six weeks, flowers senesce and develop into capsule fruits containing thousands of dust-like seeds, maturing in October (Shea 1993, USFWS 2016).

It is only during the brief late summer bloom period that *P. integrilabia* is easily distinguished from co-occurring congeners by its large white flowers with an elongated nectar spur and entire labellum. The nectar spur persists through the fruiting phenophase prolonging window of detection into early fall, but only if the individual produces an inflorescence that year (Fig. 3). Inconspicuous phenology through most of the year, low numbers of individuals per occurrence, and inconsistent flowering frequency all contribute to low detectability of populations, further warranting targeted surveys for *P. integrilabia* in suitable habitat during its bloom period.

Species distribution models aim to detect patterns of correlation between environmental predictor variables and known species occurrences to then spatially project those patterns over a given geographic area. SDMs have been used to predict habitat suitability and direct targeted surveys for rare species, often resulting in the discovery of previously undocumented populations (McCune et al. 2020, Lawson et al. 2022, McCune 2016, Lomba et al. 2010). Lomba et al. (2010) introduce the idea of the “rare species modelling paradox”. Small sample sizes, patchy or uneven distributions throughout their range, and lack of true absence locations for rare species can pose challenges to accurate predictive habitat modeling. However, the small number of occurrences across relatively small geographic ranges pose an opportunity to accurately capture and describe the restricted niches the rare species inhabit. They also point out that rare species with restricted niches that are uniquely adapted to highly fragmented habitat types are the ones

most in need of predictive habitat suitability modelling. Given *P. integrilabia*'s restricted niche, relatively large number of known occurrences throughout its range, and evidence of declining populations and habitat loss, it appears to be an excellent candidate for habitat suitability modelling.

Platanthera integrilabia was historically documented in North Carolina from two sites in Henderson County and two sites in Cherokee County but has not been observed in the state since 1992 (USFWS 2021). However, the species was recently reintroduced to one historic location in North Carolina, Bat Fork Bog in Henderson County, in the winter of 2022 through a partnership between Atlanta Botanic Gardens and the US Fish and Wildlife Service. It has been suggested that some undiscovered populations may persist in western North Carolina in under-surveyed areas of suitable habitat. Thorough ground surveys on both public and private land in western North Carolina are lacking, and given the brief period of time when *P. integrilabia* is detectable, targeted surveys during the bloom period may indeed uncover undocumented occurrences (W. Knapp, personal communication 2023). Performing targeted surveys for new populations has been listed as a priority action in a USFWS draft recovery plan for *P. integrilabia* (USFWS 2023). Previous studies have demonstrated the usefulness of using species distribution models to locate new populations of *P. integrilabia* in other portions of its range (Lawson et al. 2022, Alabama; Crabtree 2014, Tennessee). However, there have been no range-wide or Blue Ridge-specific species distribution models (SDMs) developed for the species thus far.

Two SDMs have been developed for subsets of *P. integrilabia*'s range; one for the Cumberland Plateau in eastern Tennessee and one for northern Alabama. In an unpublished study, Todd Crabtree of the Tennessee Natural Heritage Program developed a simple model using only geology, topographic curvature, and slope as environmental parameters. Despite its

simple nature, this model has been used to discover many new occurrences in the Cumberland Plateau since its development in 2006. The TN model was also used to locate candidate receptor sites for a successful *P. integrilabia* translocation project in 2018 (Wooten et al. 2020). A more recent model was developed for northern Alabama using ten environmental parameters describing various climate, watershed, topography, and soil characteristics to predict suitable habitat and direct surveys. Targeted surveys of areas predicted to have suitable habitat by the SDM yielded three new occurrences of *P. integrilabia* in Alabama (Lawson et al. 2022). These studies demonstrate that species distribution modeling and targeted surveys can be an effective tool for discovering new populations of *P. integrilabia* within subsets of its geographic range. I hypothesize that by correlating environmental predictor variables with known presence points throughout the species range, we can predict where suitable habitat for *P. integrilabia* occurs in North Carolina.

The objective of this study was to develop and evaluate the performance of a range-wide species distribution model for *P. integrilabia* and to perform targeted surveys for the species in western North Carolina within predicted habitat patches. Results from this model were used to build a framework for future iterations of the model to further refine the prediction of suitable habitat. My goal is to enable the discovery of unknown populations of *P. integrilabia* in western North Carolina, and throughout its range, and to help identify suitable locations for reintroduction and translocation of the species.

CHAPTER 2: METHODS

This research project included observing *P. integrilabia* and its habitat, developing and analyzing a predictive species distribution model (SDM), and performing targeted field surveys throughout western North Carolina.

Initial habitat characterization

In August 2022, to familiarize myself with the target species and its habitat within the Blue Ridge Physiographic region, I visited eight Element Occurrences (EOs) previously documented by state Natural Heritage Programs (NHPs) (Fig. 4). Three of the eight EOs are considered extirpated: Bat Fork Bog, Laurel Branch (Henderson County, North Carolina) and Oil Camp Creek (Greenville County, South Carolina). Through herbarium collections, previous status reports, and conversations with collectors I was able to navigate to the places where the species was once documented in the extirpated locations. Extant populations were observed at Starr Mountain (McMinn County, TN), Sawmill Branch (Polk County, TN), Pine Log Mountain WMA (Bartow County, GA), Big Canoe (Pickens County, GA), and Lee Mountain (Stephens County, GA). I made on-site observations of the common topography, soil moisture, canopy cover and associated species. These site visits were important in training myself to recognize the target species, associated species and abiotic characteristics of actual suitable habitat.

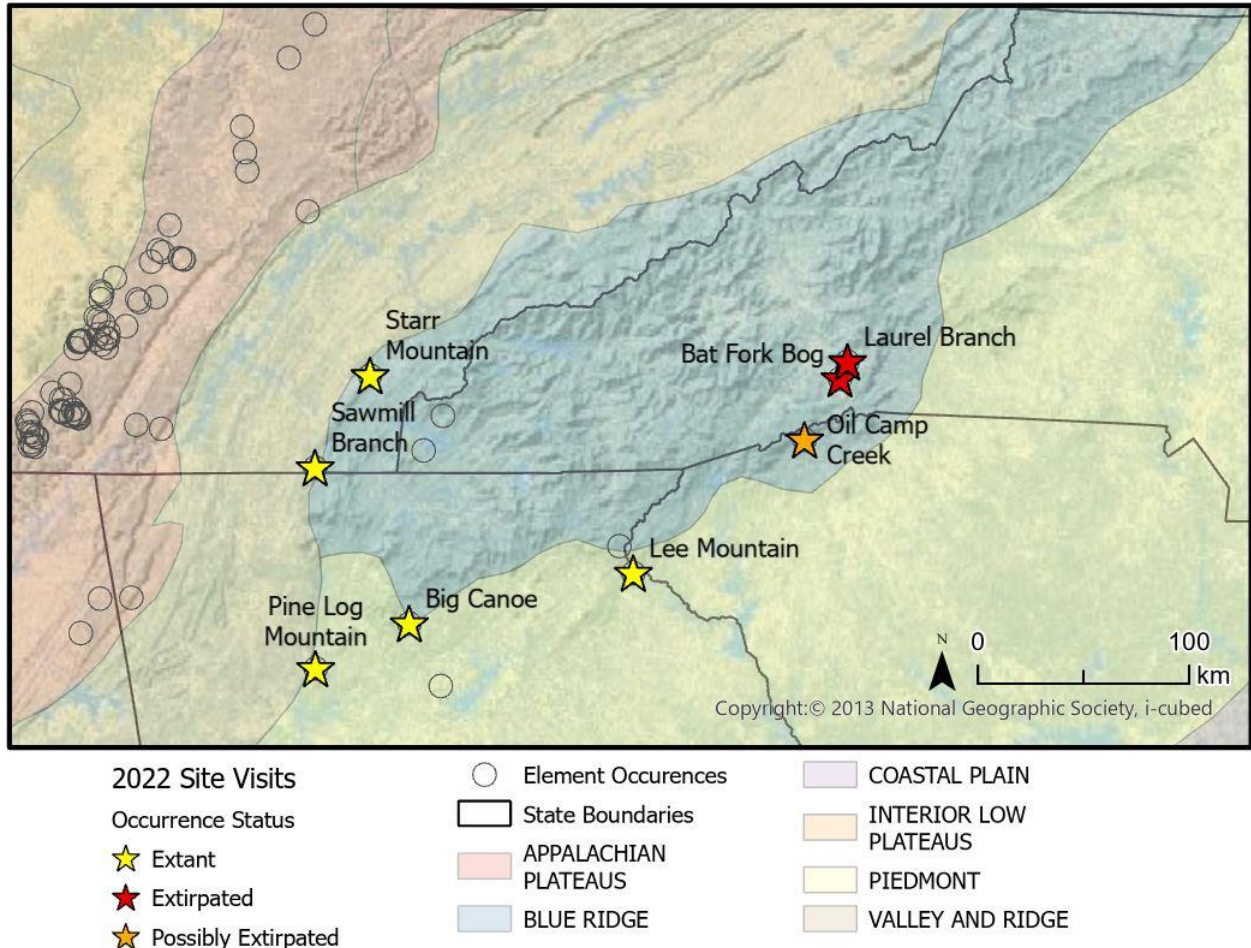


Figure 4. Location of eight element occurrences of *P. integrilabia* visited in 2022 for initial habitat observations. Extant populations were observed at Starr Mountain (McMinn County, TN), Sawmill Branch (Polk County, TN), Pine Log Mountain WMA (Bartow County, GA), Big Canoe (Pickens County, GA), and Lee Mountain (Stephens County, GA). Three of the 8 EOs are considered extirpated: Bat Fork Bog, Laurel Branch (Henderson County, North Carolina) and Oil Camp Creek (Greenville County, South Carolina).

Species Distribution Model

I developed a range wide species distribution model for *P. integrilabia* evaluating correlations between environmental predictor variables and occurrence point locations to predict habitat suitability throughout the range of the species. The final model output is available for download at <https://doi.org/10.5281/zenodo.10280442> (Harding 2023). Many different

algorithms have been evaluated for use in species distribution modeling, each with its own respective strengths and limitations dependent on sample size, study area, absence data, and species prevalence (Elith et al. 2006). Ensemble models combine multiple algorithms into one weighted output and have been shown to be more robust and accurate in predicting species distributions compared to using a single algorithm (Marmion et al. 2009, Breiner et al. 2015). In a recent review of the predictive performance of 21 common species distribution modeling approaches across 225 datasets of varying spatial extent and sample size, ensemble models consistently yielded the most accurate results (Valavi et al. 2022).

Ramirez-Reyes et al. (2021) recently published a comparison of four modeling approaches using occurrence data for four at-risk species. They compared three individual algorithms and an ensemble model which combined the three individual algorithms using a weighted average based on their relative performance. For all four species, the ensemble model consistently showed reduced uncertainty and improved predictive accuracy compared to the individual models based on cross-validation results. The author aimed to provide an easily repeatable modelling approach to benefit rare species conservation through optimizing surveys, identifying sites for relocation or reintroduction, and supporting habitat conservation planning. Considering this alignment with my own research goals, I chose to apply the same ensemble modelling technique as the author, with a few modifications, by adapting the provided R code to the *P. integrilabia* presence data and predictor variables. The ensemble model combines three commonly used algorithms: a generalized additive model (GAM; Hastie and Tibshirani 1986), a generalized boosted model (GBM; Friedman et al. 2000), and a maximum entropy model (MaxEnt; Phillips et al. 2006). GAM is a multiple regression algorithm, while GBM and MaxEnt are machine-learning methods. I processed occurrence point data, prepared environmental

predictor variable rasters, and generated pseudoabsence points using ArcPro 3.1.0 (ESRI 2022). Model fitting, projection, and cross-validation was done using RStudio (R Core Team 2021).

Study Area

I delineated an arbitrary modelling extent representing *P. integrilabia*'s recorded historic range throughout southern Appalachia. The 201,634 km² study area includes northeast Mississippi, northern Alabama, northern Georgia, northwest South Carolina, western North Carolina, southwest Virginia, southeast Kentucky, and eastern Tennessee (Fig. 5). The portion of the study area falling within western North Carolina is 26,619 km².

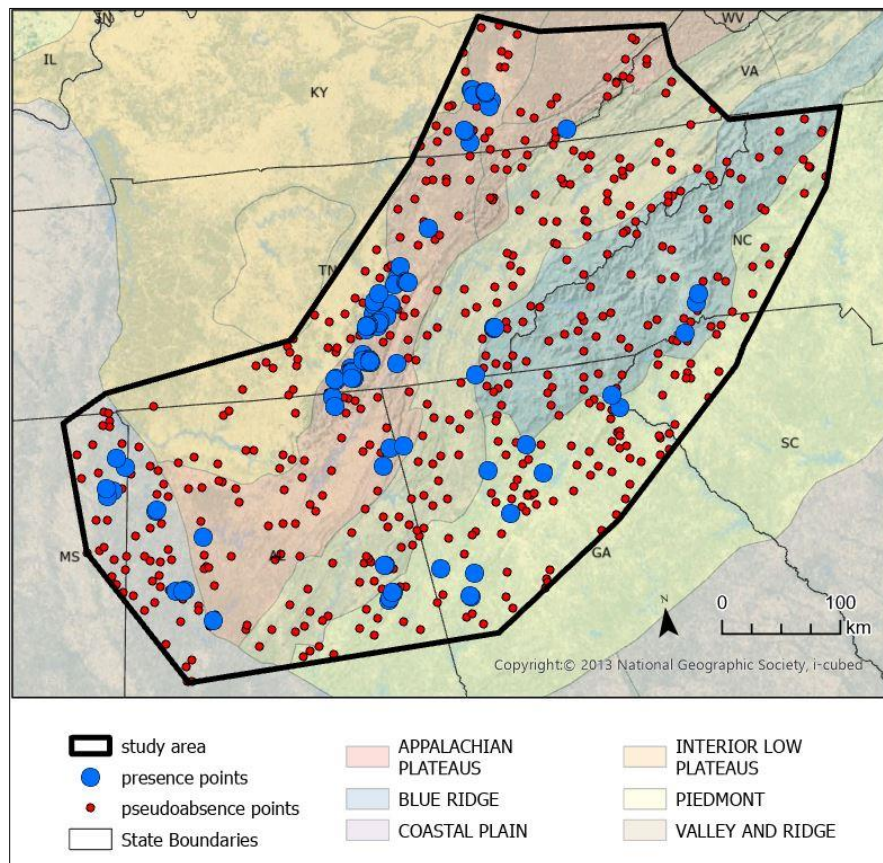


Figure 5. Study area and presence-pseudoabsence data used to train the range-wide ensemble SDM for *P. integrilabia*. Study area was arbitrarily selected to encompass the range of available presence locations, resulting in a 201,634km² area encompassing 7 states and 6 physiographic regions. Pseudoabsence points (N=500) were randomly generated throughout the study area in order to characterize the background environmental parameters for comparison to presence point locations.

Presence and Absence Points

The presence points used to inform the model were generated from EO data for *P. integrilabia* provided by state NHPs across the species' range (ALNHP 2022, GDNR 2022, MDWFP 2022, NCDENR 2022, OKNP 2022, SCDNR 2022, TDEC 2022) (Fig. 5). Though the focus of this study was to model suitable habitat within the Blue Ridge physiographic region, there are only 11 documented occurrences points within this portion of the species' range, three of which are based on historic collections with obscure location records. Hence, the modeled study area was expanded to the entire range to increase the sample size of presence locations used to inform a general model. The total number of NHP EO polygons representing all known *P. integrilabia* occurrences is 124 across 7 states. These EOs represent approximately 89 populations, only 50 of which are considered extant according to the latest species status assessment in 2021 (USFWS 2021).

Using information available from the USFWS enabling legislation designating *P. integrilabia* as a threatened species, subsequent USFWS species status assessments, and NHP metadata for each EO, I determined which polygons had enough geospatial accuracy (+/- 30m) to be used in the model (USFWS 2016). Considering environmental predictor variables have (or were resampled to) a 30m resolution, this was the preferred accuracy level of each presence location. Of the 124 presence records provided by NHPs, 85 were chosen to have enough spatial accuracy to be used in the model.

To maximize the sample size of presence locations, historic occurrences were included in the analysis as long as the location information was detailed enough to meet the 30m accuracy threshold. The 85 EOs chosen to be used as presence locations for the model comprised 61 extant and 5 extirpated occurrences, and another 17 occurrences of unknown status. There can be

a tradeoff in sensitivity (proportion of accurately predicted presence points) and specificity (proportion of correctly predicted absences) of the model when using historic locations, especially when the environmental variables include potentially impersistent conditions (temperature, precipitation, and land use) (Bracken et al. 2022). Three of the historic presence points used to train the model represented extirpated populations within the Blue Ridge physiographic region. Considering the main goal of this study was to discover undocumented populations within this region, I chose to include these and other historic populations to capture the entire range of preferred environmental conditions (achieving maximum sensitivity). However, this comes with the potential cost of specificity leading to over-prediction of suitable habitat as it relates to impersistent environmental variables such as land use changes in areas historically inhabited by the target species (Bracken et al. 2022).

Original EO data from NHPs are represented by polygons outlining an occurrence, but presence points must be in XY format for the model. To best represent the habitat of the entire occurrence, presence points were generated throughout each polygon of interest. To avoid overfitting, points had a minimum distance of 30m from any other point (considering the 30-m resolution of the environmental predictor variables). This resulted in $n=158$ presence points generated within 85 EO polygons chosen to represent habitat where *P. integrilabia* either occurs or has been known to occur within 30m.

Random pseudoabsence points were generated in ArcPro to train the model on background environmental conditions to be contrasted with presence points (Phillips et al. 2009). I created a 'background' polygon bounded by the study area and a 2km buffer around any presence points to be used in the model and randomly generated 500 points within the polygon. This number of background pseudoabsences were generated based on the methods of Ramirez-

Reyes et al. (2021) and Lawson et al (2022). Both presence and pseudoabsence point locations were exported as XY coordinates in a .csv file containing a dataset of n=654 presence and absence response variables to be used in training the model on specific environmental conditions at each point (Fig. 5).

Environmental Predictor Variables

Environmental predictor variables were chosen based on observed habitat conditions, variables used in previously developed SDMs for the species which proved effective for subsets of the species range, and the availability of geospatial data for the entire study area (see Table 3 in Results). Average annual temperature and precipitation from 1991-2020 were acquired through the PRISM Climate Group online repository (PRISM Climate Group 2020). The best available spatial resolution of 800m was resampled to a 30m resolution. The National Elevation Dataset (NED 30) was used to calculate slope, curvature, and elevation at a 30m resolution (USGS 2019). The National Hydrography Dataset (specifically NHD Flowline selected for streams and rivers) was downloaded by state across the study area and combined in ArcPro to calculate the Euclidian distance to streams at a 30m resolution (USGS 2018).

Environmental Protection Agency Stream Catchment (StreamCat) repository was used to download stream catchment-based metrics across the study area (Hill et al. 2016; <https://www.epa.gov/national-aquatic-resource-surveys/streamcat-dataset>). StreamCat variables summarize metrics for the nearby landscape that flows directly into a stream segment. Indexes are special metrics available through the StreamCat database which model or combine multiple metrics to describe catchment features such as baseflow, catchment integrity and wetness. I chose the following metrics to describe soil characteristics and substrate moisture within *P. integrilabia* habitat patches: baseflow index, catchment integrity, wetness index, soil

permeability, percent organic matter, and percent sand. Variables were downloaded for each of the 6 hydro-regions across the study area, then combined and reprojected to an equal area projection in ArcPro. Metrics were averaged across catchment basins ranging from 0.0009 to 79.1226 km² across the study area and were resampled to a 30m resolution.

Baseflow is the component of streamflow that can be attributed to ground-water discharge into streams (Wolock 2003) and can be particularly important in maintaining stream flow in the absence of precipitation events. The baseflow index is the ratio of baseflow to total flow, expressed as a percentage. Catchment integrity is a metric that evaluates the “capacity of a watershed to support and maintain the full range of ecological processes and functions essential to the sustainability of biodiversity and of the watershed resources and services provided to society” (Thornbrugh et al. 2018, page 1144). It is derived from an analysis of 22 StreamCat metrics representing anthropogenic features thought to negatively affect a watershed's ability to maintain six key functions: hydrologic connectivity, temperature regulation, habitat provision, hydrologic regulation, regulation of water chemistry, and sediment regulation. Baseflow index and catchment integrity were shown to have a high percent contribution in an Alabama-only SDM for *P. integrilabia* which led to the discovery of three new occurrences during targeted surveys (Lawson et al. 2022). The Wetness Index metric is a Compound Topographic Index which identifies areas with the potential to be wet enough to create wetlands. It uses the 30m NED to assess the upstream area and slope as it relates to overland flow and potential accumulation of water in a catchment basin (Bryce and Horvath 2019).

The mean permeability (cm/hour), mean percent organic matter, and mean percent sand soil metrics are characteristics of mapped soil associations averaged within a catchment basin, and are based on the Natural Resource Conservation Service's State Soil Geographic database

(STATSGO; <https://catalog.data.gov/dataset/u-s-general-soil-map-statsgo2>). The STATSGO database is a generalized version of the more detailed SSURGO soil survey maps and offer soil characteristics for use on a multistate or regional scale (Mednick et al. 2006). These soil metrics were chosen to capture common soil characteristics among mapped soil associations where *P. integrilabia* is known to occur. Soil permeability is known to affect important hydrologic functions of wetlands such as hydroperiod and ground water infiltration thus affecting vegetation composition (Correa-Araneda et al. 2012). Considering *P. integrilabia* is known to inhabit sandy acidic soils with little organic matter (USFWS 2021), the mean percent sand and mean percent organic matter metrics were intended to detect this correlation among presence locations.

Once the geospatial layers describing environmental predictor variables were downloaded, they were post-processed and prepared for modeling using ArcPro, and evaluated for correlation using the `cor()` function in the base R package (R Core Team 2021). StreamCat metrics were sourced as vector data. Each catchment was represented by a polygon with a mean value for each metric. Catchment polygons were rasterized into 30m pixels and aligned to match DEM derivative rasters (slope and curvature) using the ‘Mosaic to Raster’ function. All raster layers then were then clipped to the study area and aligned using the ‘Extract by Mask’ function in ArcPro, resulting in identical extents, projections, and pixel alignment (exact numbers of columns and rows) for all variables. Presence points, absence points, and predictor variables were projected to North America Albers Equal Area Conic (NAD 1983 Geographic Coordinate System), an equal-area projection, to maintain pixel size and shape considering the large multi-state extent of the study area (Ramirez-Reyes et al. 2021). Values of environmental predictor variables were extracted for each presence and pseudoabsence point, and were evaluated for

correlation using the `cor()` function in base R (R Core Team 2021). Elevation and mean annual temperature had a correlation coefficient of 0.87, so elevation was removed from the analysis.

Running the model and cross validation

The georeferenced presence points, pseudoabsence points, and environmental predictor variables were analyzed for empirical correlations which were then projected across the study area to highlight areas of suitable habitat using an ensemble SDM method described by Ramirez-Reyes et al. (2021). The R code published by Ramirez-Reyes (2021) was adapted to run, cross-validate, and combine the three models to produce the final ensemble output (Appendix A). The code employs the *mgcv* package for running the GAM (Wood 2018), *gbm* for running the GBM (Greenwell et al. 2018), and the *dismo* package for running the MaxEnt model (Hijmans et al. 2017). For generating performance metrics, the package *ROCR* was used for area under the receiver operator characteristic curve (Sing et al. 2005) and *MLmetrics* for calculating the area under the precision recall curve (Yachen 2016).

Each unique algorithm created a different predictive output based on differences in their fitted functions, which were then projected onto the study area, assigning each 30x30m pixel a relative Habitat Suitability Index (HSI) from 0 (low suitability) to 1 (high suitability). Results were then cross-validated using a leave-one-out approach to generate test values (Hastie et al 2009). The model was run 654 times (154 presence points + 500 pseudoabsence points), each time withholding one observation and using $n-1$ observations to train the model and generate a test value to compare to the withheld observation. This process generated a set of 654 ‘test values’ which were then compared to the actual ‘observed values’ to generate model performance metrics. Each model algorithm calculates own contribution index describing the relative importance of each predictor variable to habitat suitability predictions. Contribution

indices were calculated as p-values for GAM (generalized additive model), relative influence for GBM (generalized boosted model), and percent contribution for MaxEnt (maximum entropy model).

Performance metrics were calculated for each individual model, then used to create the final weighted ensemble output. The area under the receiver operator characteristic curve (AUC-ROC, or simply AUC) is the most common performance metric used to evaluate the accuracy of species distribution models (Hijmans and Elith 2019). The AUC-ROC is the area under the curve of the plot of sensitivity (correctly predicting presences) versus 1-specificity (correctly predicting absences) (Sofaer et al. 2019). AUC-ROC represents the ability of the model to discriminate between presence and absence, or probability that a randomly selected presence will have a higher suitability prediction than a randomly selected absence across all thresholds. This metric has been criticized as overstating model accuracy for species with low prevalence over a large geographic range, especially in models using pseudoabsences (Lobo et al. 2008). Therefore, the area under the precision recall curve (AUC-PR) was also generated (Ramirez-Reyes et al. 2021). This approach plots precision versus sensitivity, is independent of specificity or true absence predictions, and is not affected by study area extent or the use of pseudoabsences (Sofaer et al. 2019). Lastly, the Pearson Correlation Coefficient (COR) compared the 'test values' for each presence and pseudoabsence point to the actual 'observed values' at each point as an additional robustness check (Ramirez-Reyes et al. 2021).

The predicted HSI output from the three individual models were weighted and combined to produce the final ensemble output. Individual model HSI scores were multiplied by their relative model performance (weight=individual AUC-PR/sum of the three AUC-PR) then summed to create the final ensemble model HSI predictions (Ramirez-Reyes et al 2021). As in

the three individual model outputs, the result was a projection of predicted suitable habitat, with each 30x30m pixel within the study area assigned a relative HSI from 0 to 1 based on the combined results of the three individual models. Results were cross-validated using the same leave-one-out approach as individual models and AUC-PR, AUC-ROC and the correlation coefficient (COR) performance metrics were calculated for the ensemble model.

Model output analysis

The ensemble model output created a predictive surface assigning a habitat suitability index (0-1) to each 30m pixel within the study area. To accurately analyze proportions of the study area predicted to be suitable habitat, 3,804 km² designated as ‘open water’ in the NLCD layer were removed from the output dataset. The remaining predictive surface of the raster output was 197,830km², 26,395km² falling within western North Carolina. Next, I reclassified the continuous model output raster into HSI bins and created a vector file in order to calculate the area within each bin.

For further analysis of the model output, such as proportion of suitable habitat within a given area, it is common practice to transform the probability of suitable habitat to a binary output of suitable versus unsuitable habitat (Liu et al. 2005). This requires setting a threshold HSI value above which will be considered suitable habitat, and below which will be considered unsuitable habitat. A threshold value of 0.4 was chosen based on where sensitivity was equal to specificity (Freeman and Moisen 2008). The threshold value was determined using the R package *PresenceAbsence* to analyze predicted habitat suitability scores of presence points and pseudoabsence points (Freeman and Moisen 2008b). Areas assigned a habitat suitability score less than 0.4 were considered unsuitable, and areas assigned an HSI greater than 0.4 were considered to be suitable habitat.

After thresholding the model output, binary analyses by physiographic region and conservation status were performed using the Tabulate Area and Intersection tools in ArcPro. I determined the proportions of predicted suitable habitat in each physiographic region throughout the study area using the province classes designated by the Physiographic Regions of the Conterminous United States shapefile downloaded from USGS sciencebase.gov (Fenneman and Johnson 1946). For land ownership and conservation status of the predicted suitable habitat in western North Carolina, I used the NC NHP Managed Areas shapefile which identifies fee-simple properties and easements of conservation interest downloaded from <https://ncnhde.natureserve.org/content/data-download> (NCNHP 2023). Each of the properties within the Managed Areas database is assigned a GAP score from one to four which rates management intent to conserve biodiversity on conserved lands (USGS GAP 2022). GAP Status of 1 or 2 represent areas having permanent conservation of natural land cover with a mandated plan to manage for biodiversity. GAP Status 1 area management plans include allowing or mimicking natural disturbance events (for example wildfire), whereas GAP Status 2 areas do not. GAP 3 areas are also protected from natural land cover conversion, but are subject to extractive uses (for example logging or OHV use). GAP 4 represents land with no known conservation status or unknown management intent. For the purpose of this study, I assigned a GAP 4 status to any private property not included in the NC NHP Managed Areas database. Proportion of the predicted suitable habitat in developed versus undeveloped land was derived from the National Land Cover Database cover classes (NLCD 2019; Dewitz and USGS 2021). For this analysis, areas designated as developed (high, medium and low intensity), barren land, hay/pasture, and cultivated crops were designated as developed land, and all other cover classes designated as undeveloped land.

Targeted Surveys

Areas in western North Carolina highlighted by the model as having a high suitability index were prioritized for targeted surveys during the peak *P. integrilabia* bloom period, mid-August through early September 2023. The goals of the targeted surveys were to generate presence and absence points for future iterations of the model, ground truth the modeled prediction of suitable habitat, and to identify any areas suitable for reintroduction of the species.

Surveys were timed to capitalize on *P. integrilabia*'s most conspicuous phenophase - flowering and early fruiting - in order to increase the chance of detection. Survey locations were determined based on the modeled habitat suitability index and were focused on accessible parcels containing undeveloped natural areas. Areas containing a high concentration of cells with an HSI of 0.80 and above were assessed for current land use (prioritizing undeveloped patches) and accessibility using the NC parcel map (NC OneMap 2019), aerial imagery, and the National Land Cover Database (NLCD; Dewitz and USGS 2021). After identifying parcels with relatively large areas of undeveloped land predicted to have highly suitable habitat, I used the NC Parcel map and county tax assessors' office information to locate landowner contact information and request permission to survey. Some survey locations with lower than a 0.8 HSI were surveyed based on proximity to historic occurrences, expert recommendations of areas of suitable habitat, or for ease of accessibility.

I shared the model output with regional experts and land trusts using ArcGIS online seeking advice on which areas to prioritize for surveys based on field experience (<https://www.arcgis.com/apps/mapviewer/index.html?webmap=e8bb3c43b9db4fa98f802be7b15d547e>). Dan Pittillo, Todd Crabtree, Wesley Knapp, Alan Smith, and David Lee all provided important information to help guide survey planning. Staff at Conserving Carolina, Foothills Conservancy,

and Mainspring Conservation Trust were very cooperative in granting access to all of their fee properties and recommending conserved sites to be surveyed.

Although no new presence points of *P. integrilabia* were obtained during targeted surveys (see Results below), a list of true absence points was generated for inclusion in future iterations of the model. It has been shown that incorporating true absence points from a small subregion of the study area into the background pseudoabsence dataset can significantly improve the performance of SDMs (Koshkina et al. 2017). Track logs were imported into ArcPro as polylines, buffered by 30m (approximate survey area beyond track), and using the Extract by Polygon tool, subsamples of the ensemble raster output representing areas surveyed were generated. Absence points were generated by using the Raster to Point tool, which generates one point per 30x30m cell that was surveyed. True absence points were removed from areas I considered to be highly suitable habitat based on observed characteristics of the site. Coordinates of true absence points are available at <https://doi.org/10.5281/zenodo.10280442> (Harding 2023).

In addition to generating absence data, survey areas were assessed for actual habitat suitability. Each survey location was subjectively ranked as high, medium, or low actual habitat suitability based on observed substrate moisture, canopy cover, topography, and presence of target indicator species known to be associated with *P. integrilabia* (see Table 1 in Results). Target attributes were based on my observations of known Blue Ridge populations of *P. integrilabia* the previous year and detailed population descriptions published in species status assessments (USFWS 2021, Shea 1992). Sites ranked as ‘high’ actual habitat suitability contained at least some flat, partially shaded, perennially wet areas with well-developed *Sphagnum* mats, a significant *Carex* component in the herbaceous layer, and hosted a high number of plants known to be associated with the target species, especially congeners known to

share a mycorrhizal symbiont with *P. integrilabia*. Sites ranked as medium actual habitat suitability contained some aspects of highly suitable habitat, but lacking others. For example, a site with partially shaded saturated soils and some associated species, but lacking a significant *Sphagnum* or graminoid component were ranked as having 'medium' actual habitat suitability. Sites ranked as 'low' actual habitat suitability lacked the correct substrate moisture and canopy cover, and only hosted a few if any target associated species. The observed habitat suitability ranks were compared to the mean predicted HSI for each survey area which was determined by averaging the predicted HSI score for each pixel within the survey polygon, and tested for significance of difference using a one-way ANOVA. Survey areas hosting highly suitable habitat were evaluated as receptor sites for translocation based on proximity to known occurrences and site conservation status. I assessed the furthest distance any known population occurs from any other known populations to consider the patch dynamics of natural populations. The recently discovered *P. integrilabia* population in Cumberland Gap, Kentucky (~500 individuals) is the most geographically isolated occurrence, at 68.42km away from any other documented populations. Therefore, suitable habitat patches within 68 km of known populations were considered for translocation.

CHAPTER 3: RESULTS

Habitat characterization of known occurrences

Bat Fork Bog, Starr Mountain, Sawmill Branch, Pine Log Mountain WMA, and Big Canoe habitat consisted of non-alluvial sphagnum wetlands with saturated soil located in flat areas near stream head waters. The Lee Mountain and Oil Camp Creek occurrence locations were more of a cataract bog, where small stream tributaries flow over exposed granite bedrock on small slopes causing permanently wet sphagnum areas. A checklist of easily identifiable target indicator species was generated from observed species present at the five sites (Table 1). The herbaceous layer in areas where *P. integrilabia* was observed generally had a significant graminoid component dominated by *Carex sp.* interspersed with ferns and decomposing woody debris. The midstory shrub layer was generally sparse, with less than ~30% cover. Overstory canopy cover was variable and ranged from about 25 to 90 percent.

Table 1. List of associated species observed growing with *P. integrilabia* during initial site visits in 2022. This list was used during 2023 targeted surveys to identify and rank habitat suitability.

Scientific Name (Common name)	Scientific Name (Common name)
<i>Acer rubrum</i> (red maple)	<i>Nyssa sylvatica</i> (blackgum)
<i>Alnus serrulata</i> (smooth alder)	<i>Osmunda spectabilis</i> (royal fern)
<i>Amauropelta novaboracensis</i> (New York fern)	<i>Osmundastrum cinnamomeum</i> (cinnamon fern)
<i>Carex intumescens</i> (bladder sedge)	<i>Oxypolis rigidor</i> (stiff cowbane)
<i>Chasmanthum laxum</i> (slender spikegrass)	<i>Parnassia asarifolia</i> (kidneyleaf grass-of-Parnassus)
<i>Eutrochium fistulosum</i> (hollow Joe-pye-weed)	<i>Platanthera ciliaris</i> (yellow fringed orchid)
<i>Liquidambar styraciflua</i> (sweetgum)	<i>Platanthera clavellata</i> (small green wood orchid)
<i>Liriodendron tulipifera</i> (tuliptree)	<i>Platanthera cristata</i> (crested fringed orchid)
<i>Lobelia cardinalis</i> (cardinal flower)	<i>Sphagnum sp.</i> (peat mosses)
<i>Lorinseria areolata</i> (netted chain fern)	<i>Viburnum nudum</i> (possumhaw viburnum)
<i>Lycopus virginicus</i> (Virginia bugleweed)	

The Laurel Branch location (population extirpated) no longer contained suitable habitat. The small stream located on private property was incised with four-foot banks disconnecting

water flow from the floodplain. A small patch of woods remains but is surrounded on all sides by agriculture fields. Although some of the vegetation suggested historical increased and prolonged soil saturation (*Acer rubrum*, *Nyssa sylvatica*, *Lorinseria areolata*, *Osmunda spectabilis*, *Iris virginica*) the soil was dry with no sphagnum and a negligible sedge component. *Ligustrum sp.*, *Microstegium vimineum*, and *Hedera helix* were encroaching the site, further decreasing habitat suitability.

Model output

Individual models

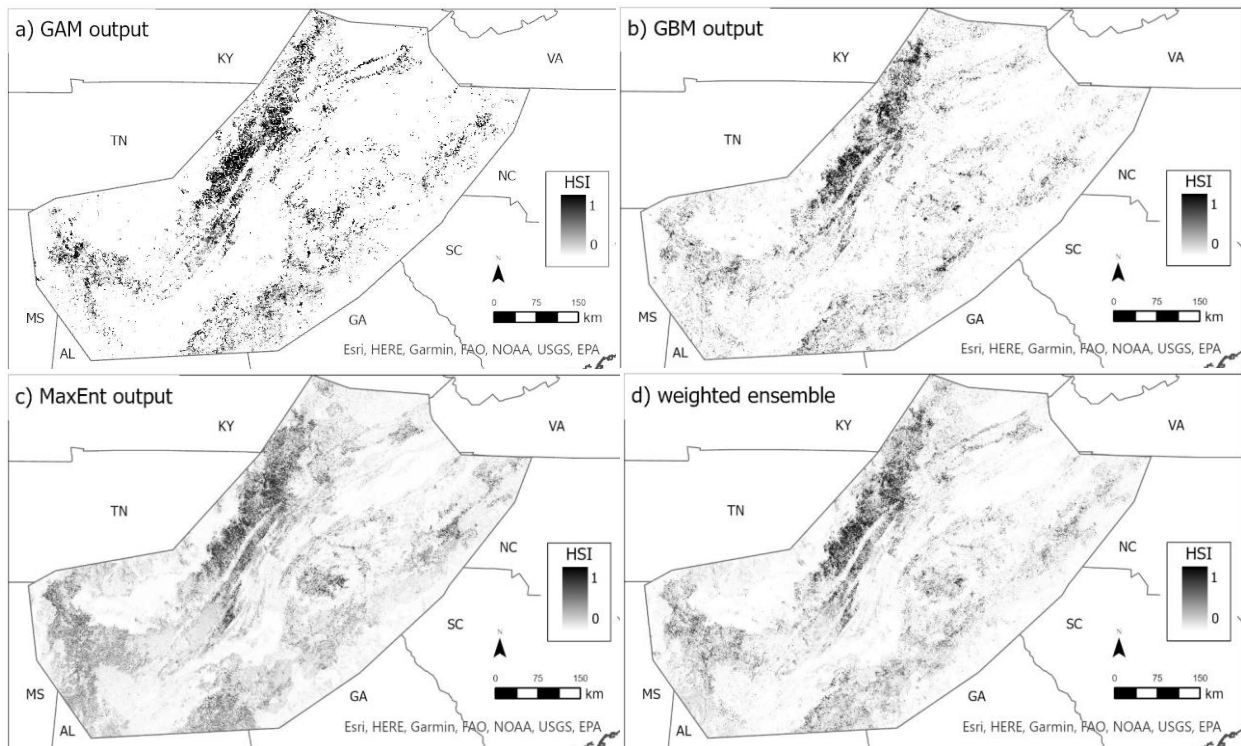


Figure 6. Habitat Suitability Index (HSI) predictions for *P. integrilabia* from three modeling algorithms, a) generalized additive model (GAM), b) generalized boosted model (GBM), and c) maximum entropy model (MaxEnt), and the final d) weighted ensemble model output. All four models highlighted eastern Tennessee and southeast Kentucky as highly suitable habitat. Based on cross-validation, the final weighted ensemble model output had the highest predictive accuracy compared to the three individual algorithms.

All three individual models highlighted eastern Tennessee and southeast Kentucky as having the highest concentration of suitable habitat throughout the range (Fig. 6). The GAM model appeared to produce the most discriminatory output, lacking widespread mid-range values (symbolized as gray areas) as seen in the MaxEnt output. The Maxent model had the highest AUC-PR and therefore the highest relative weight (0.3434), followed by the GBM (0.3430), and last the GAM (0.3136) (Table 2). The relative weights were then multiplied by the individual models' HSI values, which were then summed to assign a final ensemble HSI value to each pixel across the study area.

Table 2. Model performance metrics for the individual models and final weighted ensemble model. Area under the precision-recall curve (AUC-PR), area under the receiver operating curve (AUC), and correlation coefficient (COR) metrics were generated through cross validation using a leave-one-out approach (Hastie et al 2009, code adapted from Ramirez-Reyes et al 2021).

Model	AUC-PR	AUC	COR	Relative weight
Generalized Additive (GAM)	0.772	0.932	0.764	0.3136
Generalized Boosted (GBM)	0.844	0.944	0.758	0.3430
Maximum Entropy (MaxEnt)	0.845	0.939	0.752	0.3434
Weighted Ensemble	0.991	0.999	0.952	

All four models had AUC scores close to one, indicating significantly better-than-random predictive ability. An AUC score of 0.50 would indicate the model is as good as a random guess, but a score close to one indicates the model consistently predicts higher HSI scores at presence locations, and lower HSI scores at pseudoabsence locations (Hijmans and Elith 2019). AUC-PR does not incorporate correctly predicted absences (specificity) and is considered a more robust reflection of precision than an AUC score, especially when modelling a low-prevalence species using pseudoabsences across a large geographic area (Sofaer et al 2019b). The ensemble AUC-PR of 0.991 indicates very high discrimination ability of the model, or that 99% of the time a

randomly selected presence point will have a higher assigned HSI than a randomly selected absence point. The correlation coefficient of 0.952, indicates the test values and observed values of the presence and absence points were strongly correlated.

The relative importance of each environmental predictor variable varied across the three input algorithms (Table 3). Mean permeability of mapped soils within the catchment basin and curvature had high relative importance across all three algorithms. Base flow index had high importance in the GAM and GBM outputs but was one of the lowest contributing variables in the MaxEnt model. Alternatively, slope was not statistically significant in the GAM output, was the sixth relative influence in the GBM, and the third highest percent contribution in the MaxEnt model.

Table 3. Estimated relative importance of predictor variables in each of three habitat suitability models which were averaged to create the final ensemble output. Each model algorithm has its own contribution index: percent contribution for MaxEnt (maximum entropy model), relative influence for GBM (generalized boosted model), and p-values for GAM (generalized additive model). Values representing the most important variables for each model are highlighted in red.

Variables	GAM p-values	GBM relative influence	MaxEnt % contribution
Soil permeability ^a	<0.001	17.50	32.3
Base flow index ^a	<0.001	12.80	0.7
Curvature ^b	<0.001	12.73	17.3
Catchment integrity ^a	<0.001	10.07	8.8
Distance to stream ^c	0.002	9.37	2
Slope ^b	0.104	8.72	13
Mean annual temperature ^d	0.246	7.04	8.1
Wetness index ^a	0.003	6.33	5.8
Soil percent organic matter ^a	0.020	6.12	0.3
Mean annual precipitation ^d	0.013	5.17	7.5
Soil percent sand ^a	0.196	4.14	4.2

^a StreamCat database (Hill et al 2015)

^b Derived from 30m DEM (USGS 2019)

^c Derived from National Hydrography Dataset (USGS 2018)

^d PRISM Climate Group (2020)

Ensemble model

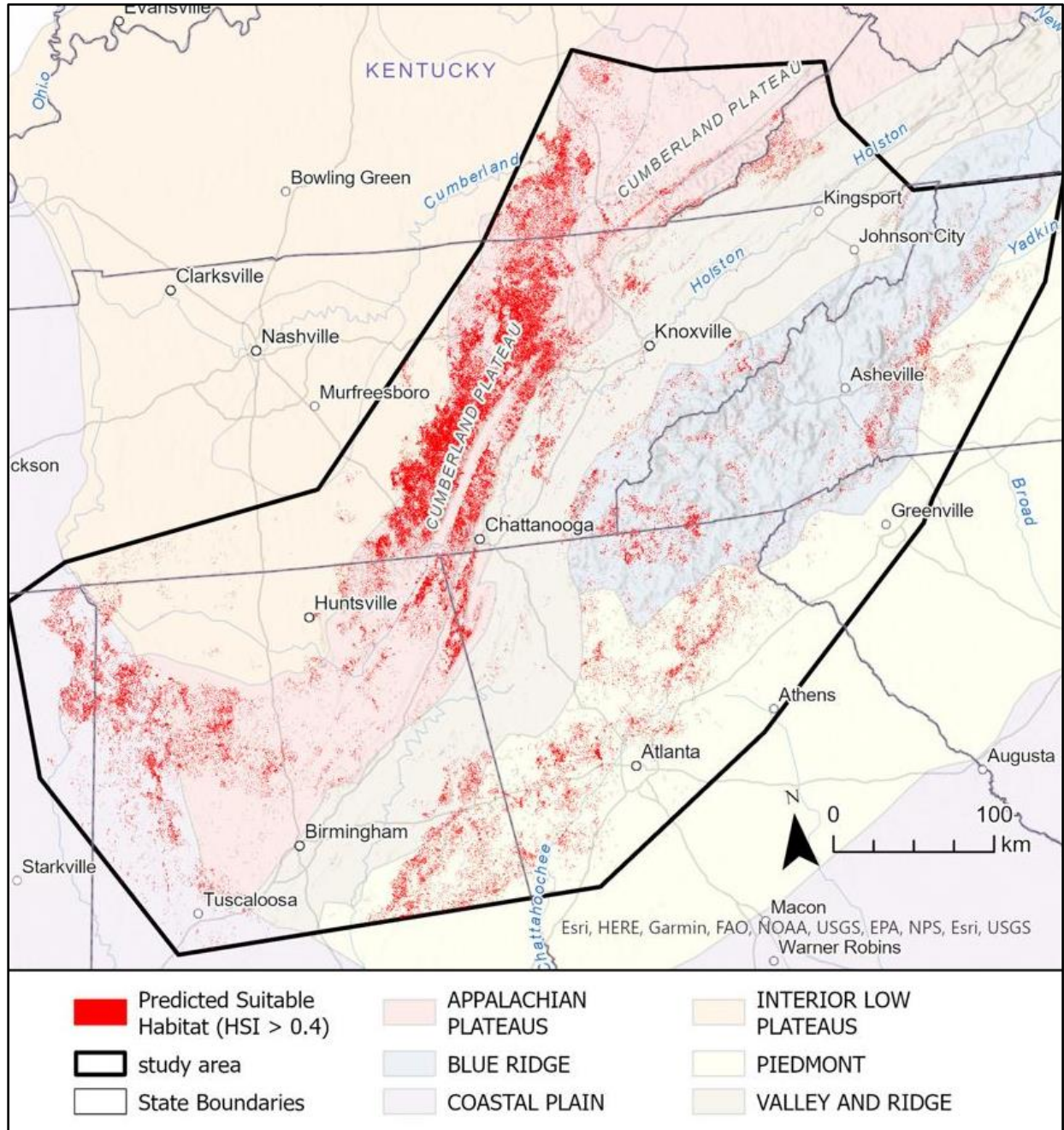


Figure 7. Probability surface output for range-wide ensemble species distribution model (weighted average of three individual algorithms). Red represents areas predicted to be suitable habitat for *P. integrilabia*. Suitable habitat is defined by areas assigned a habitat suitability index greater than 0.4. Approximately 6% of the study area was predicted to be suitable habitat. The Appalachian Plateaus physiographic region contains the highest concentration of suitable habitat, while the Valley and Ridge physiographic region contains the least concentration of predicted suitable habitat.

Using an HSI threshold value of 0.4, where sensitivity is equal to specificity, 11,223 km² of the entire study area, or approximately 6%, was predicted to be suitable habitat for *P. integrilabia* (Fig. 7). The mean predicted habitat suitability index for the presence points locations used to train the model was 0.83 (0.14). The AUC-PR of 0.991, and the clear bimodal distribution of predictions separating presence and absence indicate high model accuracy (Fig. 8). Only approximately 1% of the study area was predicted to have a habitat suitability greater than 0.8 (Table 4).

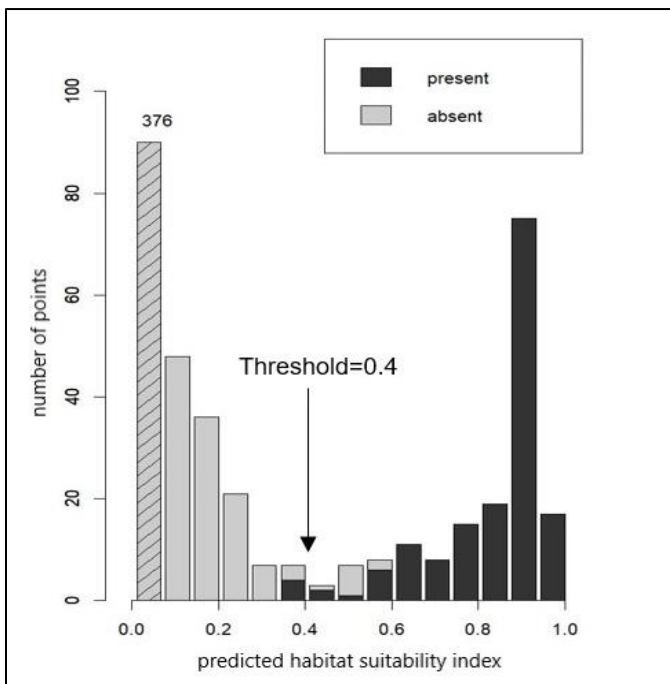


Figure 8. Predicted probabilities (habitat suitability index) for presence and absence points used to train the *Platanthera integrilabia* ensemble model output (n=654). Only 4 of the presence locations occurred within areas predicted to be less than 0.4 HSI. The 0 bar is truncated to enhance detail. The double humped histogram with minimal overlap between presence and absence predictions indicates high model accuracy (Freeman and Moisen 2008). For binary analyses, a threshold value of 0.4 was chosen to optimize sensitivity and specificity.

Table 4. Proportion of the entire study area (%) and the proportion of the study area falling within western North Carolina (WNC) predicted in each habitat suitability index (HSI) bin. Red line indicates the 0.4 threshold to distinguish areas of suitable and unsuitable habitat.

HSI	km ²	% study area	km ² in WNC	% WNC study area
0.0-0.2	172,654	85.63	24,203	90.92
0.2-0.4	12,664	6.28	1,404	5.27
0.4-0.6	6,000	2.98	531	1.99
0.6-0.8	3,753	1.86	199	0.75
0.8-1	2,470	1.22	58	0.22

The physiographic region containing the highest percentage of suitable habitat by area was the Appalachian Plateau in eastern Tennessee, which also contains the highest concentration of EOs used to train the model (Fig. 7, Table 5). The physiographic region containing the next highest proportion of suitable habitat is the Coastal Plain in northwest Alabama and northeast Mississippi, which was represented by 8 occurrences in the training dataset. Although only 5 populations were included in the model presence dataset from the Blue Ridge physiographic region, 3.52% of the study area falling within the region was predicted to be suitable habitat were. The Valley and Ridge physiographic region had the smallest proportion of suitable habitat, as well as the fewest occurrences included in the training dataset.

Table 5. Total area of each physiographic region included in the model study area and the proportion (%) of each physiographic region predicted to be suitable habitat (HSI >0.4). The number of element occurrences (EOs) included in the model training dataset per physiographic region indicates prevalence and is included for comparison.

Physiographic Region	suitable habitat (km²)	total area (km²)	% region containing suitable habitat	# EOs represented
Appalachian Plateaus	6,972	50,394	13.83%	53
Coastal Plain	1,169	16,540	7.07%	8
Piedmont	1,586	40,008	3.96%	13
Blue Ridge	1,193	33,924	3.52%	5
Interior Low Plateaus	630	20,288	3.10%	4
Valley And Ridge	673	40,652	1.66%	2

Predicted Habitat in WNC

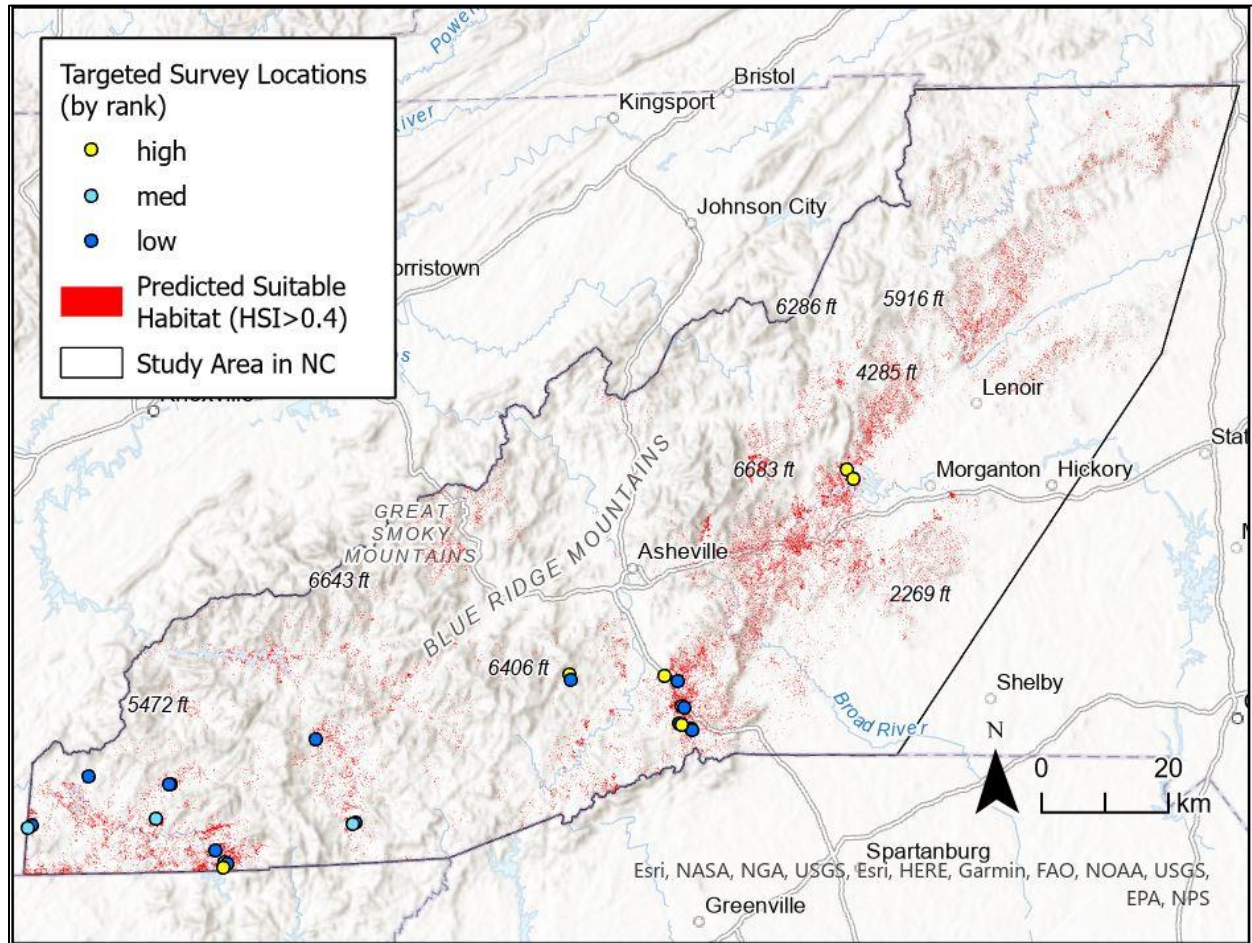


Figure 9. Predicted suitable habitat and targeted surveys for *P. integrilabia* within western North Carolina (habitat suitability index ≥ 0.4). Approximately 3% of the study area in North Carolina was predicted to be suitable habitat. Targeted surveys in Cherokee, Clay, Macon, Henderson, and Burke counties were completed in August and September 2023.

Of the 26,395 km² study area in western North Carolina, 788km² (approximately 3%) is predicted to contain suitable habitat for *P. integrilabia* (Fig. 9, Table 4). When overlaid by the National Land Cover Database, 178km² (approximately 22%) of predicted suitable habitat in WNC occurs on developed land (NLCD 2019). Of the remaining 610 km² of undeveloped predicted suitable habitat, approximately half occur within managed lands (347km²), the other half occur on private land (263km²) (Fig. 10). Approximately 131km² of undeveloped predicted

suitable habitat in western North Carolina (only 0.5% of the WNC study area) has a GAP designation of 1 or 2, where the land is managed for biodiversity (Fig. 10; NCNHP 2023).

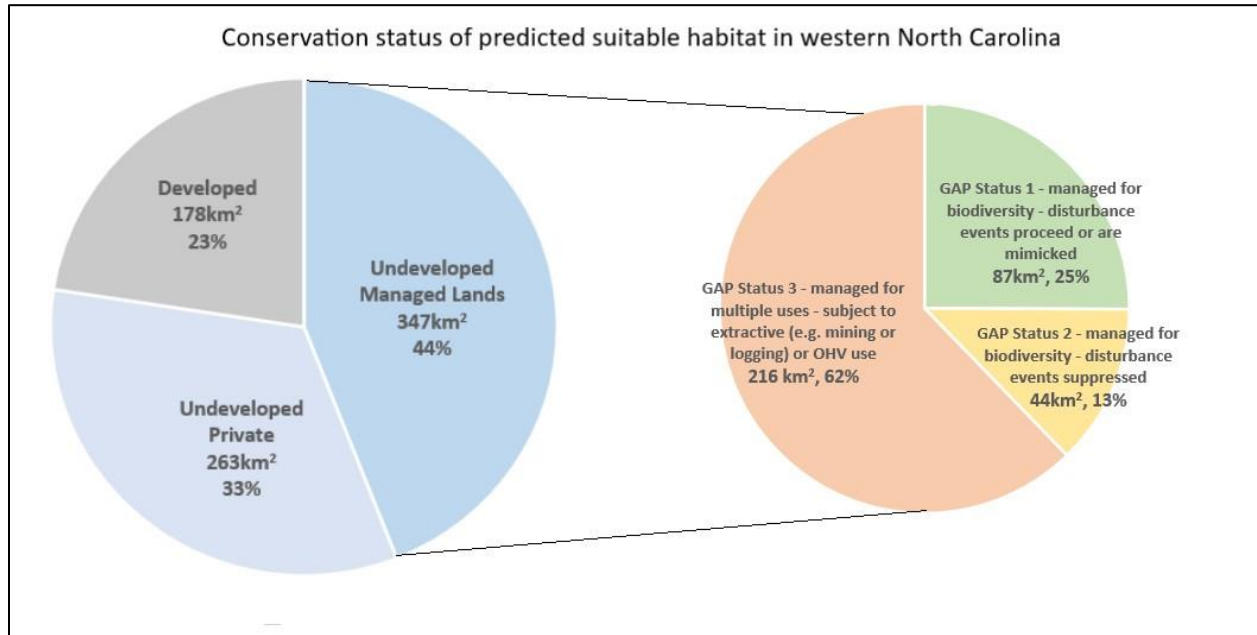


Figure 10. Conservation status of ensemble SDM predicted suitable habitat for *P. integrilabia* in western North Carolina (HSI>0.4). Left pie chart represents all 788km² of predicted suitable habitat in WNC, and right pie chart represents the suitable habitat within undeveloped managed lands. GAP status indicates conservation goals of the managed land (USGS 2022). 17% of the predicted suitable habitat in WNC falls within land being managed for biodiversity (GAP 1&2), which are of important conservation value for *P. integrilabia* habitat.

Targeted Surveys

A total of 259.8 hectares or 0.3% of predicted suitable habitat in western North Carolina was surveyed for *P. integrilabia* (Fig. 9). No new occurrences were located during targeted surveys. A total of 2,761 absence points were generated to be used in future iterations of the model (coordinates available upon request). Of the 25 survey locations, seven were ranked as ‘high’ actual habitat suitability based on observed site characteristics during targeted surveys. All seven areas ranked as high actual suitability had average predicted HSI scores above the threshold value of 0.4 (Fig. 11). Five survey sites were ranked as ‘medium’ actual habitat suitability, and 13 of the 25 survey sites were determined to have ‘low’ actual suitability. Of the

sites ranked as low observed suitability, five had an average predicted HSI less than 0.4, and one site (Bartram Trail) had an average predicted HSI of 0.89 (Table 5, Fig. 11). Using a significance value of $p < 0.05$, there was a significant difference between the average mean HSI predicted for low, medium, and high-quality sites ($F(2,25) = 3.834, p = 0.037$).

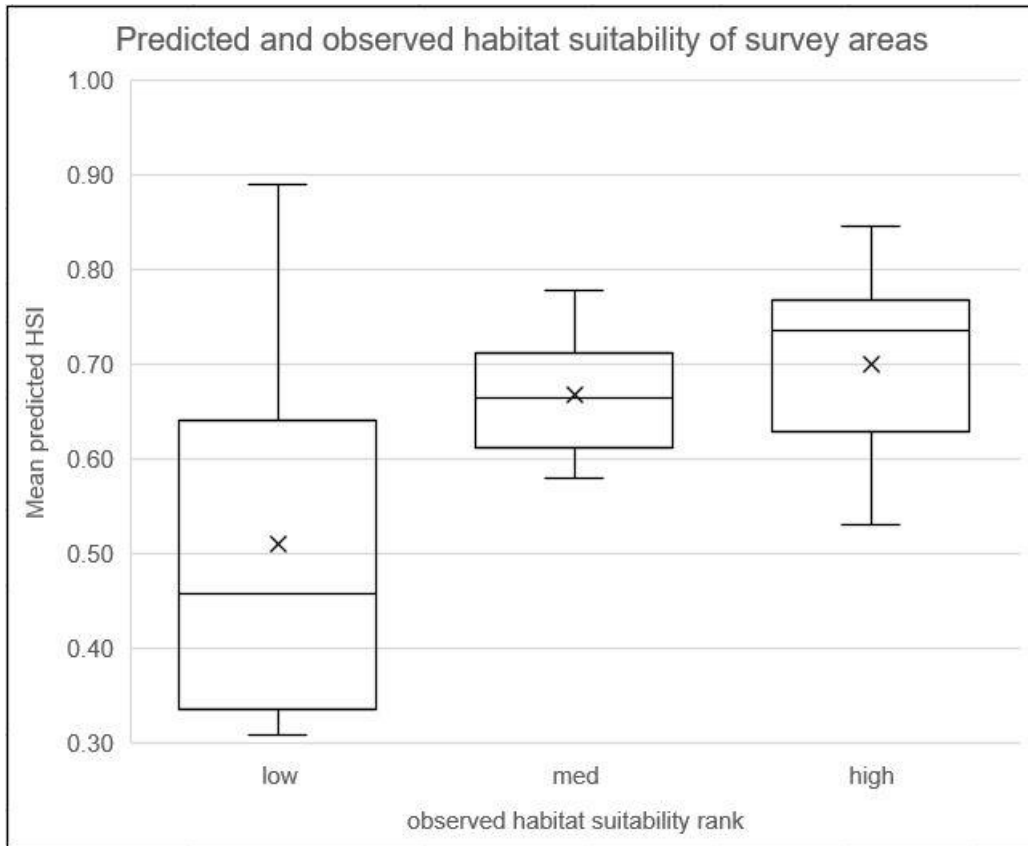


Figure 11. Average predicted habitat suitability (HSI) scores for 25 survey areas compared to the observed high, medium, or low quality habitat ranks based on observed characteristics during targeted surveys. Of the 25 survey areas, seven were ranked as high suitability, five medium, and 13 low actual suitability. Seven of the 13 low quality sites were predicted to be suitable habitat (HSI >0.4), but all high and medium quality sites were also predicted to be suitable habitat.

Targeted surveys confirmed invasive species encroachment and alteration of hydrologic function is a significant threat to the persistence of suitable *P. integrilabia* habitat in North Carolina. Many survey sites were heavily invaded by invasive species and were clearly affected by anthropogenic alterations to the local hydrology. Invasive species observed in areas predicted

to be suitable habitat include *Ligustrum sp.*, *Microstegium vimineum*, *Murdannia keisak*, *Rosa multiflorum*, and *Hedera helix*. A number of sites predicted to have high habitat suitability had some of the right conditions for good habitat, but were altered by monocultures of invasives, ditching of the floodplain, removal of native species, or separation of the floodplain due to incised water channels. The two survey sites with an average predicted HSI above 0.8 (Clear Creek in Henderson County and Eller Preserve in Clay County) were heavily invaded by *Ligustrum sp.* and *Microstegium vimineum*, causing a monoculture which is out competing native species. One survey site predicted to have highly suitable habitat in Cherokee County was actively being cleared of native vegetation and ditched to drain the flood plain (Fig. 12). Some other survey areas appeared to have had similar alterations in the recent past negatively impacting habitat quality (Clear Creek and Mud Creek in Henderson County for example).



Figure 12. Survey site in Cherokee County with a predicted habitat suitability score of 0.96 being actively destroyed during targeted surveys.

Of the seven targeted survey sites found to host high-quality habitat, six fall within conserved land (Appendix B). The two Clay County high suitability sites were the only sites to fall within a 68 km radius of a known natural population. However, one site is heavily invaded

by non-native species (Eller Preserve owned by the Nature Conservancy) and the other was located on private property with no known conservation protection. All the other sites of highly suitable habitat fell beyond the target 68km radius from a known population. If the reintroduction site at Bat Fork Bog in Henderson County becomes a viable population, two other conserved sites containing high-quality habitat fall within the 68km radius: King Creek Bog and Pink Beds North. The two Burke County high-quality sites are approximately 67 km from the Bat Fork Bog reintroduction site.

Surveys yielded the discovery of a new occurrence of the critically endangered *Sagittaria fasciculata* on private property in Hendersonville (Fig. 13). *Sagittaria fasciculata* is endemic to North and South Carolina and is listed as Federally Endangered and S1 (Critically Imperiled) in North Carolina. It has been documented to occur in 11 extant populations, 5 of which occur in Henderson County, NC. The area at large is a degraded swamp forest bog complex heavily invaded by *Rosa multiflora*, *Ligustrum sp.*, and *Microstegium vimineum*. One colony of many hundreds of stems were observed in a wet remnant ditch within a spring fed wetland. Associated herbaceous species are *Ludwigia paulustris*, *Sparganium americanus*, *Biden connata*, *Impatiens capensis*, *Carex sp.*, *Leersia virginica*, *Persicaria sp.*, *Dicanthelium sp.*, and *Osmundastrum cinnamomea*. Woody associates at the site include *Acer rubrum*, *Nyssa sylvatica*, *Liriodendron tulipifera*, *Pinus echinata*, *Rhododendron canescens*, *Viburnum nudum*, *Ilex montana*, and *Smilax rotundifolia*. The occurrence was confirmed by Nature Serve botanist, Wesley Knapp, and was reported to the NCNHP. The NCNHP plans to coordinate with the landowner to

encourage formal protection of the site.



Figure 13. Wesley Knapp confirms new occurrence of the critically endangered *Sagittaria fasciculata* discovered during targeted surveys. The population was found in an artificial drainage ditch on private property in Henderson County. The property was heavily invaded by *Rosa multiflora* and *Ligustrum sinense*.

Another interesting find was an occurrence of *Platanthera flava* var. *flava* on state land owned by the NC Department of Transportation (NCDOT). The species is listed as S3 in North Carolina (Vulnerable), and according to Weakley’s flora is considered rare in the mountains of western North Carolina (Weakley 2020). A voucher was collected and will be added to the WCU herbarium collection. The portion of NCDOT land where the *P. flava* population was found is designated as a mitigation site, and has a GAP Status of 2, implying the site is being managed for biodiversity. However, there was significant invasive species encroachment on the site including *Ligustrum* sp, *Rosa multiflora*, and *Microstegium vimineum*.

CHAPTER 5: DISCUSSION

Knowing where populations and areas of suitable habitat exist for rare species is paramount to their persistence and recovery. Over the last decade, species distribution models are becoming increasingly popular tools for informing rare species conservation (Qazi et al. 2022). Using SDMs to direct targeted surveys for undiscovered populations has proven to increase efficiency and efficacy compared to other sampling techniques (Sofaer et al. 2019, McCune 2020, Borokini et al. 2023). The results of this study support previous findings that species distribution modeling can be used to effectively locate suitable habitat for the rare orchid, *P. integrilabia* (Crabtree 2014, Lawson et al. 2022). I developed the first range-wide species distribution model for *P. integrilabia*, and performance metrics indicate strong predictive accuracy of the model (AUC-PR=0.991). Although no new occurrences of *P. integrilabia* were discovered during ground truthing surveys, some patches of high-quality habitat were found within areas predicted to have a high habitat suitability index by the model. The model predicts suitable habitat for *P. integrilabia* in 6% of the species' range, and only 3% of the range falling within western North Carolina (WNC). More than half of the WNC predicted suitable habitat is either developed or on non-conserved private property (Fig. 10). Therefore, according to the model, only about 1.3% of the historical range within NC contains conserved suitable habitat for *P. integrilabia*. This is likely an overestimate of habitat availability in WNC based on observed overprediction during targeted surveys (Fig. 11). Considering only 0.3% of the predicted suitable habitat in WNC was surveyed during this study, results do not rule out the possibility that *P. integrilabia* may still persist within the state. However, results do indicate that suitable habitat in the state still exists, but is rare and seriously threatened by environmental degradation.

Identification, conservation, and restoration of vestigial habitat patches must be prioritized for the successful reintroduction of *P. integrilabia* into its historical range in western North Carolina.

Of the seven targeted survey sites found to host high-quality habitat, I recommend one area for incorporation into the reintroduction efforts underway through a partnership between US Fish and Wildlife Service and Atlanta Botanic Gardens. The site is owned by Conserving Carolina, and occurs within the King Creek Bog, just 5km south of Bat Fork Bog where *P. integrilabia* was reintroduced in 2022. This site had an average predicted HSI score of 0.54, and surveys confirmed the presence of highly suitable habitat in an area being actively managed for biodiversity through invasive species removal and woody plant encroachment abatement. The King Creek Bog is categorized as a French Broad Valley Bog and is extensively described by Weakley and Schafale (1994). It hosts a high number of *P. integrilabia* associated species, including *Platanthera clavellata* and *P. ciliaris*, both of which share a common mycorrhizal symbiont required for germination with *P. integrilabia*. I suggest this site be considered as a suitable location for possible translocation of *P. integrilabia*.

If the reintroduction at Bat Fork Bog becomes a viable population, introducing the species to King Creek Bog could provide an important neighboring population boosting fitness and genetic diversity of both reintroduced populations (USFWS 2021). *P. integrilabia* is not considered an aggressive colonizer given the low percent flowering and subsequent seed production observed in small isolated populations potentially due to inbreeding depression (Zettler and Fairey 1990). Declining occurrences of *P. integrilabia* due to environmental degradation further isolates remnant populations causing lower viable seed production and decreased genetic diversity, which can lead to further decline of occurrences. Therefore, in order to execute a successful recovery of *P. integrilabia* in North Carolina, reducing isolation of

reintroduced populations must be considered. The 2022 reintroduction site at Bat Fork Bog is more than 138km from the closest documented population. Translocating *P. integrilabia* into King Creek Bog could reduce isolation of the reintroduced population and provide an opportunity for increased outcrossing which has been associated with increased fruit set in the species (Zettler et al. 1996). Increased fruit set should increase viable seed production, and therefore increase the potential for dispersal of the species throughout its historical range in western North Carolina.

Two other sites containing suitable *P. integrilabia* habitat were found near Bat Fork Bog: north Pink Beds in Pisgah National Forest and the Mud Creek wetland on an NC DOT mitigation property in Hendersonville. The NC DOT mitigation property could become an important receptor site if reintroductions at Bat Fork Bog (and potentially King Creek Bog) become viable populations. The degraded red maple swamp is heavily invaded by native and non-native species, but with some concerted restoration work could provide an important potential habitat patch for *P. integrilabia*. Pink Beds in Pisgah National Forest could also be a suitable site for eventual colonization, as it has very little presence of invasive species, and is being actively managed to promote biodiversity. Continued conservation and restoration of these habitat patches, as well continuing to search for additional suitable habitat sites near Bat Fork Bog, is crucial to the success of reintroduction efforts currently underway.

I also suggest prioritizing the restoration of another site ranked as highly suitable habitat, the Eller Preserve in Clay County. The site is owned and managed by the Nature Conservancy, but is heavily invaded by *Microstegium vimineum* and *Ligustrum*. If the Nature Conservancy begins active restoration of the portion of the preserve deemed suitable for *P. integrilabia* (south of the clearing being managed for pitcher plants), it too could become a suitable receptor site for

introduction considering it is only 44 km from the Tallulah Gorge population in north Georgia and occurs on a site with GAP Status of 1.

The results of this model should be considered preliminary, and as a basis for more hypotheses regarding the true distribution of *P. integrilabia* and suitable habitat patches available to the species. Ground truthing surveys suggest high sensitivity but low specificity of the model leading to overprediction in many areas (Fig.12). Iterative modeling of rare species distributions has been shown to increase predictive performance (Williams 2009, Sofaer et al. 2009, Borokini et al. 2023). Reducing spatial and environmental autocorrelation can improve transferability of habitat suitability patterns into underpopulated regions such as WNC (Phillips et al. 2009, Sofaer et al. 2019, Helmstetter et al. 2020). Here I provide a framework for future iterations of the model aimed at increasing specificity and reducing autocorrelation by refining presence, pseudoabsence, and environmental predictor inputs.

The presence point dataset used to train the range-wide model could be further refined to reduce autocorrelation, better reflect contemporary drivers of *P. integrilabia* distribution, and enhance the accuracy of performance metrics. I make the following recommendations: (1) reducing presence points to one observation per pixel of the *coarsest* predictor dataset to reduce autocorrelation. I reduced presence points to one point per 30m pixel, however some of the StreamCat metrics had a greater resolution than 30m causing potential oversampling in some areas where presence points were clustered. (2) Including only extant populations to train the model, which will allow for the use of important environmental predictor variables that change with time such as land use and canopy cover. To enhance the accuracy of performance metrics, I recommend (3) withholding a random subset of the presence locations throughout the range to test predictions. It is possible that the leave-one-out cross-validation approach used to generate

model performance metrics led to an inflated prediction of the models' discrimination ability potentially caused by autocorrelation where presence points tend to be clustered (Helmstetter et al. 2020). Although these recommended actions will reduce sample size, research shows a minimum sample size of 30 is sufficient for producing reliable results (Qazi et al. 2022).

Furthermore, increasing the number of background points (pseudoabsences) to comprehensively represent all environments in the study area should increase SDM performance. Research shows increasing background points up to 50,000 (where the burden of computational intensity outweighs the benefits of increased sampling) significantly enhances model predictive performance (Valavi et al. 2022). Given the extent of the range-wide model compared to prevalence of *P. integrilabia*, I recommend significantly increasing the number of pseudoabsence points used to train the model. Including the absence points generated in this study may enhance specificity of Blue Ridge predictions (Koshkina et al 2017). Phillips et al. (2009) suggest mitigating the sampling bias of presence points by similarly biasing pseudoabsence points. Including all NHP rare plant EOs that are not *P. integrilabia* as absence points will mimic the sample bias of the presence points. Increasing the number of randomly generated background points will better characterize environmental attributes of the entire study area. A combination of all three of these recommendations to increase the number of pseudoabsences to train the model will likely improve model accuracy.

Changes to the suite of environmental predictor variables based on the results of this study include using finer resolution data, adding lithology, land use, and canopy cover, and potentially removing soil characteristics from the analysis. Using the 30m DEM for slope analysis may have caused the model to over-predict in areas too steep for *P. integrilabia*. Using the 10m DEM would more accurately model slope. Slope based on the 10m DEM is only one of

three predictor variables used in the highly successful TN only model (Crabtree 2014). However, to realize the benefits of this modification would require resampling all data to a 10m resolution, significantly increasing computational intensity of the model. The coarseness of catchment-level data used in this model likely contributed to autocorrelation (Phillips et al. 2006), and may not have captured specific characteristics associated with small habitat patches occurring within catchments. I recommend using finer resolution soil metrics (such as POLARIS probabilistic 30m dataset, Chaney et al. 2016) or removing soil metrics from the analysis altogether based on inconsistencies between county soil maps and observed soils within *P. integrilabia* occurrences. The TN model did not incorporate soil characteristics because the actual soil type where *P. integrilabia* was known to inhabit in that region was too small and intricate to be accurately mapped (Crabtree 2007). In another study assessing hydrologic characteristics of three wetlands in Kentucky hosting *P. integrilabia* populations, soils at all three sites were incorrectly mapped compared to actual observations (Hoy 2012). I recommend adding lithology, as the species' distribution appears to follow patterns in bedrock geology effecting soil pH. When my model output is compared to known suitable habitat in Tennessee, the model tends to overpredict in areas of limestone. Including lithology in the analysis could reduce overprediction into non-acidic soils determined by bedrock characteristics. Including canopy cover and land use as predictor variables could reduce overprediction into rhododendron tunnels and developed areas where *P. integrilabia* does not occur.

The results of this study provide some insights on where future targeted surveys could be the most effective at discovering new occurrences of *P. integrilabia*. The Coastal Plain physiographic region had the second highest percent of predicted suitable habitat by area, but only eight populations are known to occur within that region. It is also the smallest

physiographic region within the study area. Targeted surveys in that region in 2020 yielded 4 new occurrences (Lawson et al. 2022). Based on these results, I recommend more targeted surveys for *P. integrilabia* be employed in the southwest portion of the species' range in northwest Alabama and northeast Mississippi. Despite the need for more botanical and rare plant surveys on private land in general, future targeted surveys for undiscovered populations of *P. integrilabia* may be most successful if they are directed towards conserved land. My discovery of the critically endangered *Sagittaria fasciculata* on private land further supports the importance of botanical inventory on private land. However, based on my targeted surveys, actual habitat suitability for *P. integrilabia* was more prevalent on conserved land compared to private land. Other factors that should be considered when prioritizing areas of predicted suitable habitat for targeted surveys is predicted habitat patch size and connectivity. Considering the negative anthropogenic affects to habitat quality appear be greater on smaller patches of suitable habitat, focusing on sites with the largest contiguous area of predicted suitable habitat may be more effective. As for connectivity, the most isolated known population of *P. integrilabia* is 68 km from any other population. Focusing surveys within that radius of known occurrences may be most effective for locating undiscovered populations.

The identification and conservation of rare species and their habitats is more important than ever as we are witnessing an unprecedented decline in biodiversity worldwide. Knowing where rare species occur is essential to conservation, and identifying where vestigial habitat patches remain within historic ranges is important for species recovery efforts. The results of this study show how a range-wide species distribution model for *P. integrilabia* can effectively predict areas of suitable habitat within its historical range in the Blue Ridge physiographic region of western North Carolina. The model predicted approximately 2.3% of western North Carolina

contains undeveloped suitable habitat for *P. integrilabia*. I surveyed 0.3% of the predicted suitable habitat in western North Carolina which did not yield the discovery of undocumented populations, but did locate some remaining patches of suitable habitat within the historical range of the species. Targeted surveys also provided powerful insight on model performance, informing a framework for future iterations of the model presented here. The results of this study do not rule out the possibility that *P. integrilabia* may still persist in North Carolina, but that patches of suitable habitat still remain in the state. Continued efforts to model suitable habitat and perform targeted surveys for *Platanthera integrilabia* will contribute to species conservation, and potential restoration of this threatened orchid into its historical range.

REFERENCES

- Alabama Natural Heritage Program (ALNHP). (2021). Element occurrence records for *Platanthera integrilabia*. Data available on request from the ALNHP, Auburn, AL. Accessed 29 March 2022.
- Borokini, I.T., Nussear, K., Petitpierre, B., Dilts, T. E., Weisberg, P.J. (2023). Iterative species distribution modeling results in the discovery of novel populations of a rare cold desert perennial. *Endangered Species Research*, 50: 47-62.
- Bracken, J.T, Davis, A.Y., O'Donnell, K.M., Barichivich, W.J., Walls, S.C., Jezkova, T. (2022). Maximizing species distribution model performance when using historical occurrences and variables of varying persistency. *Ecosphere*, 13(3). <https://doi.org/10.1002/ecs2.3951>
- Breiner, F.T., Guisan, A., Bergamini, A. and Nobis, M.P. (2015). Overcoming limitations of modelling rare species by using ensembles of small models. *Methods Ecol Evol*, 6: 1210-1218. <https://doi.org/10.1111/2041-210X.12403>
- Bryce, S., and Horvath, E. (2019). Potential Wetland Areas National Data Fact Sheet. *EnviroAtlas*. United States Environmental Protection Agency.
- Chaney, N.W., Wood, E.F., McBratney, A.B., Hempel, J.W., Nauman, T.W., Brungard, C.W. and Odgers, N.P. (2016). POLARIS: A 30-meter probabilistic soil series map of the contiguous United States. *Geoderma* 274:54–67.
- Connor, T., Viña, A., Winkler, J.A., Hull, V., Tang, Y., Shortridge, A., ... & Liu, J. (2019). Interactive spatial scale effects on species distribution modeling: The case of the giant panda. *Scientific Reports*, 9(1): 14563. <https://doi.org/10.1038/s41598-019-50953-z>
- Correa-Araneda, F.J., Urrutia, J., Soto-Mora, Y., Figueroa, R., Hauenstein, E. (2012). Effects of the hydroperiod on the vegetative and community structure of freshwater forested wetlands, Chile. *Journal of Freshwater Ecology*, 27(3), 459-470.
- Crabtree, T, Withers, D., Mathes, S., Condict, K. and McCoy, R. (2007). "An Ecological Assessment of Franklin State Forest Franklin and Marion Counties, Tennessee" Final Report Submitted to Tennessee Division of Forestry by Tennessee Division of Natural Areas in Partial Fulfillment of Contract ID-06-08092-00
- Crabtree, T. (2014). Predictive Habitat Modeling for Seven Plant Species of Federal Concern in Tennessee: An unpublished Report Prepared for U.S. Fish and Wildlife Service.
- Currah, R.S., Zettler, L.W., McInnis, T.M. (1997). *Epulorhiza inquilina* sp. nov. from *Platanthera* (Orchidaceae) and a key to *Epulorhiza* species. *Mycotaxon*, 61:335-342.

- Dewitz, J., and U.S. Geological Survey (2021). National Land Cover Database (NLCD) 2019 Products (ver. 2.0, June 2021): U.S. Geological Survey data release, <https://doi.org/10.5066/P9KZCM54>.
- Draper, D., Marques, I., Iriondo, J.M. (2019). Species distribution models with field validation, a key approach for successful selection of receptor sites in conservation translocations. *Global Ecology and Conservation*, 19(2019) e00653.
- Esri Inc. (2023). ArcGIS Pro (Version 3.1). Esri Inc. <https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview>
- Fenneman, N.M. and Johnson, D.W. (1946). Physiographic divisions of the conterminous U. S.: U.S. Geological Survey data release, <https://doi.org/10.5066/P9B1S3K8>
- Freeman, E.A. and Moisen, G.G. (2008). A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. *Ecological Modelling*, 217(2008):48-58.
- Freeman, E.A. and Moisen, G.G. (2008b). PresenceAbsence: An R Package for Presence Absence Analysis. *Journal of Statistical Software*, 23(11).
- Friedman J., Hastie, T., Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). *The Annals of Statistics*, 28:337–407.
- Georgia Department of Natural Resources (GDNR). (2022). Element occurrence data for white fringeless orchid. Unpublished GIS data from Georgia Department of Natural Resources, Wildlife Resources Division. Received 22 April 2022.
- Greenwell, B., Boehmke, B., Cunningham, J. (2018). Generalized boosted regression models “gbm”. R package version 2.1.8. Available: <https://cran.r-project.org/web/packages/gbm/>. Accessed June 2023.
- Hampe, A. and Petit, R.J. (2005). Conserving biodiversity under climate change: the rear edge matters. *Ecology letters*, 8:461-467.
- Harding, M. (2023). Searching for the White Fringeless Orchid, *Platanthera integrilabia*, in western North Carolina [Zenodo]. <https://doi.org/10.5281/zenodo.10280442>.
- Hastie, T., Tibshirani, R. (1986). Generalized additive models. *Statistical Science*, 3:297–318.
- Hastie, T., Tibshirani, R., Friedman, J.H. (2009). The elements of statistical learning: data mining, inference, and prediction. 2nd ed. New York, Springer. pp. 241-247
- Helmstetter, N.A., Conway, C.J., Stevens, B.S., Goldberg, A.R. (2021). Balancing transferability and complexity of species distribution models for rare species conservation. *Diversity and Distributions*, 27(1): 95-108.

- Hijmans, R.J., Phillips, S., Leathwick, J., Elith, J. (2017). Species distribution modeling “dismo”. R package version 1.1- 4. Available: <http://cran.r-project.org/package=dismo> (June 2023).
- Hill, R.A., Weber, M.H., Leibowitz, S.G., Olsen, A.R., Thornbrugh, D.J. (2016). The Stream-Catchment (StreamCat) dataset: A database of watershed metrics for the conterminous United States. *Journal of the American Water Resources Association (JAWRA)*, 52:120–128. <https://www.epa.gov/national-aquatic-resource-surveys/streamcat-dataset>
- Hooten, M.B., and Hobbs, N.T. (2015). A guide to Bayesian model selection for ecologists. *Ecological Monographs*, 85(1): 3–28. <https://doi.org/10.1890/14-0661.1>
- Hoy, C. (2012). A Hydrologic Characterization of Three Headwater Mountain Wetlands in Eastern Kentucky, USA. *Theses and Dissertations--Forestry and Natural Resources*, University of Kentucky. https://uknowledge.uky.edu/forestry_etds/5
- Jackson, C.R., Thompson, J.A., Kolka, R.K. (2014). Wetland soils, hydrology and geomorphology. In: Batzer, D.; Sharitz, R., eds. *Ecology of freshwater and estuarine wetlands*. Berkeley, CA: University of California Press: 23-60. Chapter 2., 23-60.
- Koshkina, V., Wang, Y., Gordon, A., Dorazio, R. M., White, M., Stone, L. (2017). Integrated species distribution models: combining presence-background data and site-occupancy data with imperfect detection. *Methods in Ecology and Evolution*, 8(4), 420-430.
- Lawson, K. M., Johnson, M.A.K., Schotz, A. (2022). Habitat Suitability Modeling and site verification for the White Fringeless Orchid (*Platanthera integrilabia*) in Alabama. *Southeastern Naturalist*, 21(1):28-41
- Levison, J., Larocque, M., Fournier, V., Gagne, S., Pellerin, S., and Ouellet, M.A. (2014). Dynamics of a headwater system and peatland under current conditions and with climate change. *Hydrological Processes*. 28: 4808-4822.
- Liu, C., Berry, P.M., Dawson, T.P. and Pearson, R.G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*. 28: 385-393. <https://doi.org/10.1111/j.0906-7590.2005.03957.x>
- Lomba, A., Pellissier, L., Randin, C., Vicente, J., Moreira, F., Honrado, J., Guisan, A. (2012). Overcoming the rare species modelling paradox: A novel hierarchical framework applied to an Iberian endemic plant. *Biological Conservation*. 143: 2647-2657.
- Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K., and Thuiller, W. (2009). Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*. 15: 59-69.
- McCune, J.L. (2016). Species distribution models predict rare species occurrences despite significant effects of landscape context. *Journal of Applied Ecology*. 53: 1871-1879.

- McCune, J.L., Rosner-Katz, H., Bennett, J.R., Schuster, R., Kharouba, H.M. (2020). Do traits of plant species predict the efficacy of species distribution models for finding new occurrences? *Ecology and Evolution*. 10: 5001-5014.
- Mednick, A.C. (2010). Does soil data resolution matter? State Soil Geographic database versus Soil Survey Geographic database in rainfall-runoff modeling across Wisconsin. *Journal of Soil and Water Conservation*, 65: 190 - 199.
- Mississippi Department of Wildlife, Fisheries, and Parks (MDWFP) (2022). Element occurrence data for white fringeless orchid. Unpublished GIS data from Mississippi Department of Wildlife, Fisheries, and Parks, Natural Heritage Program. Received 14 March 2022.
- Murdock, N.A. (1994). Rare and endangered plants and animals of the southern Appalachian wetlands. *Water, Air, and Soil Pollution*. 77:385-405.
- NC OneMap (2019). NC Parcel Map. North Carolina Department of Information Technology, Government Data Analytics Center, Center for Geographic Information and Analysis. Available at <https://www.nconemap.gov/pages/parcels> (Accessed: June 2023).
- North Carolina Department of Environment and Natural Resources (NCDENR) (2022). Element occurrence data for white fringeless orchid. Unpublished GIS data from NCDENR, Raleigh, NC. available by request at www.ncnhp.org. Received 05 April 2022.
- North Carolina Natural Heritage Program (NCNHP) (2023). Geographic Information System (GIS) data. NCDNCR, Raleigh, NC. Available at www.ncnhp.org. (Accessed: August 2023).
- Office of Kentucky Nature Preserves (OKNP) (2022). Element occurrence data for white fringeless orchid. Unpublished GIS data from Office of Kentucky Nature Preserves, Natural Heritage Program. Received 04 April 2022.
- Phillips, S.J., Anderson, R.P., Schapire, R.E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190:231–259.
- PRISM Climate Group (2020). PRISM climate data. Oregon State University, Corvallis, OR. Available online at <https://prism.oregonstate.edu>. Accessed 29 March 2023.
- R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Ramirez-Reyes, C., Nazeri, M., Street, G., Jones-Farrand, D.T., Vilella, F.J., Evans, K.O. (2021). Embracing ensemble species distribution models to inform at-risk species status assessments. *Journal of Fish and Wildlife Management* 12(1):98–111.
- Rehm, E.M., Olivas, P., Stroud, J., & Feeley, K.J. (2015). Losing your edge: Climate change and the conservation value of range-edge populations. *Ecology and Evolution*. 5(19): 4315-4326. <https://doi.org/10.1002/ece3.1645>

- Shea, M. (1992). Status Survey Report on *Platanthera integrilabia*. Unpublished report to the U.S. Fish and Wildlife Service, Southeast Region.
- Sing, T., Sander, O., Beerenwinkel, N., Lengauer, T. (2005). ROCR: visualizing classifier performance in R. *Bioinformatics*. 21(20):3940-1.
- Sofaer, H.R., Hoeting, J.A., Jarnevich, C.S. (2018). The area under the precision-recall curve as a performance metric for rare binary events. *Methods in Ecology and Evolution* 10:565–577.
- Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. U.S. General Soil Map (STATSGO2). Available online at <https://sdmdataaccess.sc.egov.usda.gov>. Accessed 23 May 2023.
- South Carolina Department of Natural Resources (SCDNR). (2022). Element occurrence records for *Platanthera integrilabia*. Available upon request from the SCDNR Accessed 09 March 2022.
- Spira, T.P. (2011). Wildflowers and Plant Communities of the Southern Appalachian Mountains and Piedmont. The University of North Carolina Press.
- Tennessee Department of Environment and Conservation (TDEC). (2022). Element occurrence records for *Platanthera integrilabia*. Available upon request from the TDEC Division of Natural Areas, Nashville, TN. Accessed 07 April 2022.
- Thornbrugh, D.J., Leibowitz, S.G., Hill, R.A., Weber, M.H., Johnson, Z.C., Olsen, A.R., ... & Peck, D.V. (2018). Mapping watershed integrity for the conterminous United States. *Ecological indicators*, 85: 1133-1148.
- U.S. Fish and Wildlife Service (USFWS) (2021). Species status assessment report for *Platanthera integrilabia* (white fringeless orchid), Version 1.0. March 2021. Atlanta, Georgia.
- U.S. Fish and Wildlife Service (USFWS) (2023). Recovery Plan for White Fringeless Orchid (*Platanthera integrilabia*). Atlanta, Georgia.
- U.S. Geological Survey (USGS) (2014). National Elevation Dataset – 30-m resolution. Available online at <https://datagateway.nrcs.usda.gov/>. Accessed 23 May 2023.
- U.S. Geological Survey (USGS) (2018). National Hydrography Datasets – NHDPlusV2. Available online at <https://www.epa.gov/waterdata/get-nhdplus-national-hydrography-dataset-plus-data>. Accessed 3 May 2023.
- U.S. Geological Survey (USGS) Gap Analysis Project (GAP) (2022). Protected Areas Database of the United States (PAD-US) 3.0: U.S. Geological Survey data release, <https://doi.org/10.5066/P9Q9LQ4B>.

- US Fish and Wildlife Service (USFWS) (2016). Threatened Species Status for *Platanthera integrilabia* (White Fringeless Orchid) 2016, 81 FR 62826 Federal Register 81(177): 62826–62833.
- Valavi, R., Guillera-Arroita, G., Lahoz-Monfort, J. J., & Elith, J. (2022). Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. *Ecological Monographs*. 92(1), e01486.
- Weakley, A.S. (2020). Flora of the southeastern United States. University of North Carolina Herbarium, North Carolina Botanical Garden, Chapel Hill, NC. Pdf downloadable at <https://ncbg.unc.edu/research/unc-herbarium/floras/>
- Weakley, A.S., Schafale, M.P. (1994). Non-alluvial wetlands of the Southern Blue Ridge — Diversity in a threatened ecosystem. *Water Air Soil Pollut* 77:359–383.
<https://doi.org/10.1007/BF00478428>
- Williams, J.N., Seo, C., Thorne, J., Nelson, J.K., Erwin, S, O'Brien, J.M., Schwartz, M.W. (2009). Using species distribution models to predict new occurrences for rare plants. *Diversity and Distributions*, 15:565–576.
- Wolock, D.M. (2003). Base-flow index grid for the conterminous United States. U.S. Geological Survey Open-File Report 03-263. U.S. Geological Survey. Reston, Virginia.
<http://water.usgs.gov/lookup/getspatial?bfi48grd>.
- Wood, S. (2018). Mixed GAM computation vehicle with automatic smoothness estimation “gam”. R package version 1.8-31. Available: <https://cran.r-project.org/web/packages/mgcv/index.html> (June 2023).
- Wooten, S., Call, G., Datillo, A., Cruse-Sanders, J., Nagel Boyd, J. (2020). Impacts of Forest Thinning and White-Tailed Deer Herbivory on Translocation of the Rare Terrestrial Orchid *Platanthera integrilabia*. *Diversity* 12 (412).
- Yachen, Y. (2016). MLMetrics: machine learning evaluation metrics. R package version 1.1.1. Available: <http://github.com/yanyachen/MLmetrics>
- Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Case, A., Costello, C., Dewitz, J., Fry, J., Funk, M., Grannemann, B., Rigge, M., Xian, G. (2018). A new generation of the United States National Land Cover Database (NLCD): Requirements, research priorities, design, and implementation strategies. *ISPRS Journal of Photogrammetry and Remote Sensing* 146:108–123.
- Zettler, L.W. and McInnis Jr, T.M. (1992). Propagation of *Platanthera integrilabia* (Correll) Luer, an endangered terrestrial orchid through symbiotic seed germination. *Lindleyana* 7:154-161.
- Zettler, L.W., McInnis Jr., T. M., Ahuja, J.S. (1996). Insect pollination of the endangered monkey-face orchid (*Platanthera integrilabia*) in McMinn County, Tennessee: one last glimpse of a once common spectacle. *Castanea* 61:14-24.

APPENDIX A: R CODE

```
####Code Adapted from Carlos Ramirez-Reyes, "Embracing ensemble
species distribution models to inform at-risk species assessments"
2021, some annotations by C. Ramirez-Reyes##

#This code generates species distribution models in R using three
#algorithms:
#General additive models (GAM), generalized boosted regression (GBM)
#and Maximum entropy (MAXENT)
#Multiple elements are based on "Spatial Distribution Models" tutorial
#developped by Robert J Hijmans and Jane Elith 2019-
#https://rspatial.org/sdm/SDM.pdf
#author: Carlos Ramirez-Reyes cr20666@msstate.edu
#-----#
-----#

#####load required packages#####
# data manipulation
library(raster) #to work with raster files
library(sp) #to work with coordinates
library(rgdal) #to work with shapefiles
library(rgeos) #to perform shapefile gdifference
library(tidyverse) #to make data subsets
library(dismo) #Allows running Maxent in R and calculate variable
correlations
library(knitr)
library(data.table)

#load model packages
library(gbm) # To run GBM
library(mgcv) #To run GAM
library (rJava)# To run Maxent

# cross validation
library(ROCR) #To calculate AUC
library(philentropy) #To calculate jaccard index
library(MLmetrics) #To calculate AUC-PR

#### System setup####

#Change working directory if needed to your project directory
setwd("D:/WCU/Thesis ArcPro/PLIN_model")
```

```

#Now set a temporary files folder- Ideally a location with a larger
#capacity >500GB since we will use lots of large rasters
#This will probably require you to install R libraries again!

rasterOptions(tmpdir = "D:/WCU/Thesis ArcPro/PLIN_model/TempFiles",
progress="text", timer=TRUE, overwrite = T)

####Load data####

#####Load presence points for the species####
#We will work with a table containing occurrence records (coordinates)
#with these columns: "species_name", "x" and "y"

p <- read.csv("D:/WCU/Thesis ArcPro/PLIN_model/presence_pts2.csv",
header = T)
head(p)

#subset the table to leave only "x:" and "y" point coordinates
p_c <- p[ , c("x","y")]
head(p_c)
#Convert coordinates to points:
p_xy <-p_c
coordinates(p_xy)=~x+y
#Assign a projection to presence points (make sure that points' native
#projection the specified below, otherwise transform it to your
#coordinate system)I used Albers Equal Area conic proj (aea)
proj4string(p_xy)=CRS("+proj=aea +lat_0=40 +lon_0=-96 +lat_1=20
+lat_2=60 +x_0=0 +y_0=0 +datum=NAD83 +units=m +no_defs") # set
projection
plot(p_xy)

# Note,if you're working with a large area and large pixel sizes, then
#an equal-area projection system should be used. If the raster cells
#are not equal-area, then that will bias the pseudoabsences towards
#larger raster cells and their associated environmental data.

####Load predictor variables####

#Instead of loading predictor rasters one by one we will load all
#*.tif files in our folder as a raster stack
#This will require that all your environmental predictor raster files
#have the same extent (# of columns and rows)
#Also, it is recommended to use an equal-area projection system. If
#the raster cells are not equal-area, then that will bias the
#pseudoabsences towards larger raster cells and their associated
#environmental data.

```

```

dir <- "D:/WCU/Thesis ArcPro/PLIN_model/predictor_tifs"
#The location where the raster files are

WetnessIndex <- raster("D:/WCU/Thesis
ArcPro/PLIN_model/predictor_tifs/WetIndex.tif")
BFI <- raster("D:/WCU/Thesis
ArcPro/PLIN_model/predictor_tifs/BFI.tif")
CatchInteg<-raster("D:/WCU/Thesis
ArcPro/PLIN_model/predictor_tifs/CatchInteg.tif")
Curvature<-raster("D:/WCU/Thesis
ArcPro/PLIN_model/predictor_tifs/curve.tif")
Elevation<-raster("D:/WCU/Thesis
ArcPro/PLIN_model/predictor_tifs/elev.tif")
PctOM<-raster("D:/WCU/Thesis
ArcPro/PLIN_model/predictor_tifs/PctOM.tif")
PctPerm<-raster("D:/WCU/Thesis
ArcPro/PLIN_model/predictor_tifs/PctPerm.tif")
PctSand<- raster("D:/WCU/Thesis
ArcPro/PLIN_model/predictor_tifs/PctSand.tif")
AnnPrecip<-raster("D:/WCU/Thesis
ArcPro/PLIN_model/predictor_tifs/PRISMppt.tif")
MeanTemp<-raster("D:/WCU/Thesis
ArcPro/PLIN_model/predictor_tifs/PRISMtmean.tif")
Slope<-raster("D:/WCU/Thesis
ArcPro/PLIN_model/predictor_tifs/slope.tif")
DistToStream<-raster("D:/WCU/Thesis
ArcPro/PLIN_model/predictor_tifs/StreamDist.tif")
files=c(WetnessIndex,BFI,CatchInteg,Curvature,Elevation,PctOM,PctPerm,
PctSand,AnnPrecip,
        MeanTemp,Slope,DistToStream)
files #Check the names of the files
library(raster)
raster_stack=stack(files)

####Load study area####

#Use a polygon to delimit the study area
sa <- readOGR("D:/WCU/Thesis ArcPro/PLIN_model/Study_Area_proj.shp")
#do this only if you have a shapefile
plot (sa)#Visualize the study area

#Import state boundaries for reference if you have such file
#states <- readOGR("D:/WCU/Shapefile Downloads/States_shapefile.shp")
#do this only if you have a shapefile
#plot (states, add = T)
plot (sa, add = TRUE)
plot(p_xy, add =T)

```

```

#Make sure the projections are the same, if they are not, then
#reproject
proj4string(raster_stack)
proj4string(sa)
proj4string(p_xy)
proj4string(states)

####Load pseudoabsences points ####

# I created a 'background' polygon (bounded by the study area and a 2km
#buffer around any presence points to be used in the model and
#randomly generated 500 points within the polygon, then exported the
#xy coordinates into a .csv file, "pseudoabsence_pts2.csv"

a_xy <- read.csv("D:/WCU/Thesis
ArcPro/PLIN_model/pseudoabsence_pts2.csv", header = T)
head(a_xy)
proj4string(a_xy)=CRS("+proj=aea +lat_0=40 +lon_0=-96 +lat_1=20
+lat_2=60 +x_0=0 +y_0=0 +datum=NAD83 +units=m +no_defs") # set
projection
plot(a_xy)

#Visualize the presence and pseudoabsence points
plot(raster_stack, 2, add=TRUE) #Choose first raster in the predictors
stack, make add=TRUE
points(a_xy, col='red',)
points(p_xy, col='blue')
plot(sa, border= "blue", add = TRUE)

#### Extract environmental predictor values from rasters and make a
data table ####

#1. First extract the predictor values for presence points
p_vals <- raster::extract(raster_stack, p_xy)
head(p_vals) #Check table

#Now extract predictor values for the background points
#(pseudoabsences)
a_vals <- raster::extract(raster_stack, a_xy)
head(a_vals) #Check table

#2 Append the coordinates for the presence points
p_xy_coords <- data.frame(p_xy) #Converts spatial points to data
p_vals <- cbind(p_xy_coords, p_vals)
head(p_vals)
#Append the coordinates for the background points

```

```

#a_xy_coords <- as.data.frame(a_xy[,1:2]) #Converts spatial points to
#coordinate table
a_xy_coords <- as.data.frame(a_xy)
a_vals <- cbind(a_xy_coords, a_vals)
head(a_vals)

#3 Create row with "1"s to each presence record and a "0" to each
#background
pb <- c(rep(1, nrow(p_vals)), rep(0, nrow(a_vals)))
head(pb)
colnames(a_vals)[1] <- "x"
# Append the pb ("0"s and "1"s) to the extracted values of predictors
sdmdata <- data.frame(cbind(pb, rbind(p_vals, a_vals)))
head(sdmdata)

#Now we have the final data with presences and absences points and the
#associate predictor values to be used in models!
# Let's save them to the computer
saveRDS(sdmdata, "D:/WCU/Thesis
ArcPro/PLIN_model/r_outputs/sdmdata_06132023.Rds") #An
write.csv(sdmdata, "D:/WCU/Thesis
ArcPro/PLIN_model/r_outputs/sdmdata_06132023.csv")

#-----#
##### Generalized additive model      G   A   M      #####
#-----#
#Read the sdmdata to continue working if you closed R, you might need
#to open the predictors too
sdmdata <- readRDS("D:/WCU/Thesis
ArcPro/PLIN_model/r_outputs/sdmdata_06132023.Rds")

#To run GAM we need to select the variables to use as having all
#possible will exceed the capabilities of the model
#specially with limited presence data, so lets check for correlations

####Check for variable correlations####
#Lets start with an exploratory analysis to eliminate correlated
#variables
not_used_vars <- names(sdmdata) %in% c("pb", "x", "y") #to exclude
#coordinates and variables not used in correlations
bc1 <- bioclim(sdmdata[!not_used_vars]) #bioclim finds automatically
#the predictor columns
pairs(bc1) # plots the correlations, so we can start filtering
#out correlated variables

```

```

cor(sdmdata[,
c('WetIndex', 'BFI', 'CatchInteg', 'curve', 'elev', 'PctOM', 'PctPerm', 'PctSand',
  'PRISMppt',
  'PRISMtmean', 'slope', 'StreamDist')])
M<-cor(sdmdata)
library(corrplot)
corrplot(M, method="number")
#strong corr found between elev and PRISMtmean -0.82, so I removed
#elev from model

#Run a full model with all non-correlated variables and all points

gam_full <- gam(pb ~ s(WetIndex)+ s(BFI)+ s(CatchInteg) +
  s(curve) +s(PctOM) +
  s(PctPerm) + s(PctSand) + s(PRISMppt) +
  s(PRISMtmean) + s(slope) +
  s(StreamDist),
  data = sdmdata, family= binomial("logit"))

summary(gam_full) #Check the model
gam_full
as_table(summary(gam_full))
as.table(summary(gam_full))
as.table(gam_full)
testdf1<-data.table(gam_full$var.summary)

####Evaluate the GAM model ####

#First, we create a set of predicted values to contrast with our
#"actual" presence and absences(pseudoabsences)
#We use the "leave one out-llo" approach in which we remove one of the
#points in the data and create a model
#with the remaining set. We then use that model to predict the
#suitability of the point left out. And repeat the process
#for each row of the dataset.

llo_gam <- rep(1,nrow(sdmdata)) #create an object of the size of nrows
for (j in 1:nrow(sdmdata)) #Create the function
{
  fj <- gam(pb ~s(WetIndex)+ s(BFI)+ s(CatchInteg) +
    s(curve) +s(PctOM) +
    s(PctPerm) + s(PctSand) + s(PRISMppt) +
    s(PRISMtmean) + s(slope) +
    s(StreamDist),#select the same model variables used in
#the model above
    data = sdmdata[-j,], family= binomial("logit"))#Train the
#model with everything except one point

```



```

    llo_gam[j] <- predict(fj, sdmdata[j,], type="response") #Now make
#the prediction of the 1 point
}

#Now calculate the metrics
AUC_GAM <- AUC(y_pred = llo_gam, y_true = sdmdata$pb) #Calculates the
#AUC
PRAUC_GAM <- PRAUC(y_pred = llo_gam, y_true = sdmdata$pb) #Calculates
#the PR-AUC
COR_GAM <- cor(llo_gam, sdmdata$pb, "complete.obs") #Calculates the
#correlation
JAC_GAM <- rbind(llo_gam, sdmdata$pb)
JAC_GAM <- JAC_GAM[,!apply(is.na(JAC_GAM), 2, any)]
JAC_GAM <- 1-distance(JAC_GAM, method = "jaccard") #Calculate Jaccard
#index

#Finally make a prediction of the species potential suitable habitat
#over the full study area extent using the GAM model.

pred_gam_full <- predict(raster_stack,gam_full,progress='text',
type="response")
plot(pred_gam_full)
writeRaster(pred_gam_full, filename="D:/WCU/Thesis
ArcPro/PLIN_model/r_outputs/outputrasters/GAM_raster06132023",
format="GTiff", overwrite=TRUE) #Save it

#-----#
###----- M A X E N T -----###
#-----#
#To run Maxent in R we need to first install the package "dismo",
#which contains Maxent as a function
#We also NEED to have java installed in the pc and the maxent.jar file
#copied in the #R dismo folder, that in our case I had to put it as
D:/TempR/R/win-library/3.5/dismo/java/maxent.jar
#Go here if you need trouble shooting java
#https://stackoverflow.com/questions/37735108/r-error-onload-failed-
#in-loadnamespace-for-rjava
#Go here to look for more parameters for Maxent
#https://stackoverflow.com/questions/35082922/running-maxent-in-
#r/35464295
#Go here for a guide to select parameters
#https://onlinelibrary.wiley.com/doi/full/10.1111/j.1600-
#0587.2013.07872.x

```

```

#When working with MaxEnt, consider using the R package 'ENMeval' to
#optimize the MaxEnt parameters: https://cran.r-
#project.org/web/packages/ENMeval/vignettes/ENMeval-vignette.html

#Read the sdmdata to continue working if you closed R, you might need
#to open the predictors too
sdmdata <-readRDS("D:/WCU/Thesis
ArcPro/PLIN_model/r_outputs/sdmdata_06132023.Rds")

#let create a model using the same non-correlated predictor variables
#as in the GAM model

max_full <-
maxent(sdmdata[c("WetIndex", "BFI", "CatchInteg", "curve", "PctOM", "PctPer
m", "PctSand", "PRISMppt",
                "PRISMtmean", "slope", "StreamDist")],
      p=sdmdata[, "pb"],
      args=c('quadratic=true',
            'product=true',
            'threshold=true',
            'hinge=true',
            'threads=2',
            'responsecurves=true',
            'jackknife=false',
            'askoverwrite=false',
            'writeplotdata=true'))

max_full #Check the model output

####Maxent model evaluation####

#We will create a model evaluation in a similar way to GAM, using the
#leave-one-out approach, see above for description

llo_max <- rep(1,nrow(sdmdata)) #create an object of the size of nrows
for (j in 1:nrow(sdmdata))
{
  fj <- maxent(sdmdata[-j,
c("WetIndex", "BFI", "CatchInteg", "curve", "PctOM", "PctPerm", "PctSand", "P
RISMppt",
                "PRISMtmean", "slope", "StreamDist")],
#select the same model variables as the full model
      p=sdmdata[-j, "pb"],
      args=c('quadratic=true',
            'product=true',

```

```

        'threshold=true',
        'hinge=true',
        'threads=2',
        'responsecurves=true',
        'jackknife=false',
        'askoverwrite=false',
        'writeplotdata=true'))
  llo_max[j] <- predict(fj, sdmdata[j,], type="response")
}

#### Maxent AUC ####
llo_max <- rep(1,nrow(sdmdata)) #create an object of the size of nrow
for (j in 1:nrow(sdmdata))
{
  fj <- maxent(sdmdata[-j,c(4:13)], p=sdmdata[-j,"pb"])
  llo_max[j] <- predict(fj, sdmdata[j,], type="response")
}

#Now calculate the metrics
AUC_MAX <- AUC(y_pred = llo_max, y_true = sdmdata$pb) #Calculates the
AUC
PRAUC_MAX <- PRAUC(y_pred = llo_max, y_true = sdmdata$pb) #Calculates
the PR-AUC
COR_MAX <- cor(llo_max, sdmdata$pb, "complete.obs") #Calculates the
#correlation
JAC_MAX <- rbind(llo_max, sdmdata$pb)
JAC_MAX <-JAC_MAX[,!apply(is.na(JAC_MAX), 2, any)]
JAC_MAX <- 1-distance(JAC_MAX, method = "jaccard") #Calculate Jaccard
#index

# Finally make a prediction of the species potential suitable habitat
#over the full raster extent using the GAM model

pred_max_full = predict(max_full, raster_stack, progress='text',
timer=TRUE)
plot(pred_max_full)
writeRaster(pred_max_full, filename="D:/WCU/Thesis
ArcPro/PLIN_model/r_outputs/outputrasters/MAX_raster06152023",
format="GTiff", overwrite=TRUE)

#-----
#### Boosted regression trees /GBM Generalized boosted model####
#-----
#-----
#We will first run an exploratory model to identify the optimal number
#for trees

```

```

gbm_ex2<- gbm.step(
  data=sdmdata,
  gbm.x
=c("WetIndex","BFI","CatchInteg","curve","PctOM","PctPerm","PctSand","
PRISMppt",
      "PRISMtmean","slope","StreamDist"), #select the variables
#used from sdmdata, same set as GAM
  gbm.y = "pb",
  family = "bernoulli",
  tree.complexity = 5,
  learning.rate = 0.005, bag.fraction = 0.5)

#We got around optimal number of trees in the run above. Now lets fit
#the final #model with those trees.

gbm_full <- gbm(formula = pb ~ WetIndex + BFI + CatchInteg + curve +
PctOM + PctPerm + PctSand + PRISMppt + PRISMtmean + slope
+ StreamDist,
  data = sdmdata,
  distribution = "bernoulli",
  n.trees = 3050,
  bag.fraction = 0.5)

summary(gbm_full) #show the relative influence of each variable as
table

#### Evaluate GBM####

#We also use the leave-one-out approach as in the GAM and GBM, see
#explanation above
#start calculating the prediction made

l1o_gbm <- rep(1,nrow(sdmdata)) #create an object of the size of nrows
for (j in 1:nrow(sdmdata))
{
  fj <- gbm(pb ~ WetIndex + BFI + CatchInteg + curve + PctOM + PctPerm
+ PctSand + PRISMppt + PRISMtmean + slope
  + StreamDist,
  data=sdmdata[-j,],
  distribution = "bernoulli",
  n.trees = 3050,
  bag.fraction = 0.5)
  l1o_gbm[j] <- predict(fj, sdmdata[j,], n.trees=3050,
type="response")
}

```

```

#Now calculate the metrics
AUC_GBM <- AUC(y_pred = llo_gbm, y_true = sdmdata$pb)
#Calculates the AUC
PRAUC_GBM <- PRAUC(y_pred = llo_gbm, y_true = sdmdata$pb)
#Calculates the PR-AUC
COR_GBM <- cor(llo_gbm, sdmdata$pb, "complete.obs")
#Calculates the correlation
JAC_GBM <- rbind(llo_gbm, sdmdata$pb)
JAC_GBM <- JAC_GBM[,!apply(is.na(JAC_GBM), 2, any)]
JAC_GBM <- 1-distance(JAC_GBM, method = "jaccard")
#Calculate Jaccard index

#Finally make a prediction for the model with all variables to the
#study area
pred_gbm_full <- predict(raster_stack, gbm_full, progress='text',
n.trees=3050, type='response')
plot(pred_gbm_full)
#Save the output prediction as a tiff
writeRaster(pred_gbm_full, filename="D:/WCU/Thesis
ArcPro/PLIN_model/r_outputs/outputrasters/gbm_raster06152023",
format="GTiff", overwrite=TRUE)

#### Ensemble model ####

#####
#ensemble model based on weighted AUC-PR/
#It is a similar approach to using Somers' D as a weighting
#factor based on 'Overcoming limitations of modelling rare
#species by using ensembles of small models' (Breiner et al
#2015) Methods in Ecology and Evolution 2015, 6, 1210-1218
#
#####

AUCPRs_sum <- PRAUC_GAM + PRAUC_GBM + PRAUC_MAX #sum the individual
#PRAUC scores for each model approach

weight_gam <- (PRAUC_GAM/AUCPRs_sum) #Calculate the weights for each
#model
weight_gamv <- rep(weight_gam,nrow(sdmdata)) #Generate a vector of
#size sdmdata to be used in the ensemble algebra
weight_gbm <- (PRAUC_GBM/AUCPRs_sum)
weight_gbmv <- rep(weight_gbm,nrow(sdmdata))
weight_max <- (PRAUC_MAX/AUCPRs_sum)
weight_maxv <- rep(weight_max,nrow(sdmdata))
# The ensemble model prediction based on different models and weights

```

```

esm <- ((pred_gam_full * weight_gamv) + (pred_gbm_full * weight_gbmv)
+ (pred_max_full * weight_maxv))

plot(esm)
writeRaster(esm, filename="D:/WCU/Thesis
ArcPro/PLIN_model/r_outputs/outputrasters/PLIN_ensemble_06162023",
format="GTiff", overwrite=TRUE)

####Calculate AUC of the ensemble model ####
#Extract the raster values for the presences and absences
sdmdata_pts <- sdmdata[,c("x","y")] #get the coordinates
esm_pred_vals <- raster::extract(esm, sdmdata_pts) #Extract ensemble
#values for each coordinate pair

#Now calculate the metrics
AUC_ESM <- AUC(y_pred = esm_pred_vals, y_true = sdmdata$pb)
#Calculates the AUC
PRAUC_ESM <- PRAUC(y_pred = esm_pred_vals, y_true = sdmdata$pb)
#Calculates the PR-AUC
COR_ESM <- cor(esm_pred_vals, sdmdata$pb, "complete.obs") #Calculates
#the correlation
JAC_ESM <- rbind(esm_pred_vals, sdmdata$pb)
JAC_ESM <- JAC_ESM[,!apply(is.na(JAC_ESM), 2, any)]
JAC_ESM <- 1-distance(JAC_ESM, method = "jaccard") #Calculate Jaccard
#index

#### EOF (end of file)####

```

APPENDIX B: TARGETED SURVEY LOCATIONS

Table B1. Results of targeted surveys completed in August and September 2023.

HECTARES SURVEY DATE	LOCATION	HSI MEAN ±SD	GAP Status	DESCRIPTION	<i>P. INTEGRILABIA</i> ASSOCIATED SPECIES PRESENT	HABITAT RANK
3.42 ha 8/17/23	King Creek Bog Henderson Co. Conserving Carolina	0.54 ±0.05	1	French Broad Valley Bog; saturated substrate with <i>Sphagnum</i> mats and significant <i>Carex</i> component with variable canopy cover (some areas being managed for open canopy), many PLIN associated species, suggested location for expansion of the Bat Fork Bog reintroduction.	<i>Acer rubrum</i> , <i>Nyssa sylvatica</i> , <i>Viburnum nudum</i> , <i>Alnus serrulata</i> , <i>Osmundastrum cinnamomea</i> , <i>Osmunda spectabilis</i> , <i>Amauropelta noveboracensis</i> , <i>Lorinseria areolata</i> , <i>Lycopus virginicus</i> , <i>Lobelia cardinalis</i> , <i>Oxypolis rigidor</i> , <i>Eutrochium fistulosum</i> , <i>Platanthera clavellata</i> , <i>Platanthera ciliaris</i> , <i>Sphagnum</i> sp.	high
10.89 ha 8/18/23	Mud Creek, Henderson County NC Dept Transportation	0.75 ±0.12	2	Disturbed alluvial swamp with large pockets of saturated soils among braided runnels with <i>Sphagnum</i> mats and significant <i>Carex</i> component and variable canopy cover west of powerline cut, invaded by <i>Ligustrum</i> sp. and <i>Rosa multiflora</i>	<i>Acer rubrum</i> , <i>Liriodendron tulipifera</i> , <i>Nyssa sylvatica</i> , <i>Alnus serrulata</i> , <i>Viburnum nudum</i> , <i>Osmundastrum cinnamomea</i> , <i>Osmunda spectabilis</i> , <i>Lorinseria areolata</i> , <i>Amauropelta novaboracensis</i> , <i>Lycopus virginicus</i> , <i>Oxypolis rigidor</i> , <i>Lobelia cardinalis</i> , <i>Eutrochium fistulosum</i> , <i>Carex intumescens</i> , <i>Sphagnum</i> sp., <i>Platanthera clavellata</i>	high
23.94 ha 8/22/23	Pink Beds North, Transylvania County US Forest Service	0.53 ±0.11	3	Large swamp forest- bog complex with saturated soils, many pockets of well-developed <i>Sphagnum</i> with significant <i>Carex</i> component, variable canopy cover, and many PLIN associated species	<i>Acer rubrum</i> , <i>Liriodendron tulipifera</i> , <i>Nyssa sylvatica</i> , <i>Viburnum nudum</i> , <i>Dicanthelium</i> sp., <i>Chasmanthum laxum</i> , <i>Osmundastrum cinnamomea</i> , <i>Osmunda spectabilis</i> , <i>Parathylepteris noveboracensis</i> , <i>Lorinseria areolata</i> , <i>Lycopus virginicus</i> , <i>Lobelia cardinalis</i> , <i>Eutrochium fistulosum</i> , <i>Carex intumescens</i> , <i>Platanthera clavellata</i> , <i>Sphagnum</i> sp.	high

HECTARES SURVEY DATE	LOCATION	HSI MEAN ±SD	GAP Status	DESCRIPTION	<i>P. INTEGRILABIA</i> ASSOCIATED SPECIES PRESENT	HABITAT RANK
5.76 ha 8/24/23	Eller Preserve, Clay County The Nature Conservancy	0.85 ±0.17	1	Low Mountain Seepage Bog; southern-most drainage south of the meadow represents best PLIN habitat. Area of suitable habitat heavily invaded by <i>Ligustrum sp.</i> and <i>Microstegium vimineum</i>	<i>Acer rubrum, Alnus serrulata, Liriodendron tulipifera, Nyssa sylvatica, Viburnum nudum, Osmundastrum cinnamomea, Osmunda spectabilis, Amauropelta noveboracensis, Lorinseria areolata, Lycopus sp., Eutrochium fistulosum, Lobelia cardinalis, Oxypolis rigidor, Platanthera clavellata, Sphagnum sp.</i>	high
2.34 ha 8/24/23	Lake Chatuge east, Clay County Private, Cajam Properties	0.74 ±0.17	4	Wetland area along small tributary stream. Small patches of sphagnum mats among braided runnels. Many PLIN associated species. Many <i>Platanthera</i> leaves observed, some flowering <i>P. clavellata</i> . Heavily invaded by <i>Ligustrum sp.</i> and <i>Microstegium vimineum</i> . Leaf sample collected.	<i>Acer rubrum, Alnus serrulata, Liriodendron tulipifera, Nyssa sylvatica, Viburnum nudum, Osmundastrum cinnamomea, Osmunda spectabilis, Amauropelta noveboracensis, Lorinseria areolata, Lycopus virginicus, Lobelia cardinalis, Oxypolis rigidor, Parnassia asarifolia, Eutrochium fistulosum, Sphagnum sp, Platanthera clavellata</i>	high
12.78 ha 9/8/23	Paddys Creek South Burke County Foothills Conservancy	0.79 ±0.19	2	Large non-alluvial wetland in northwest part of property; approximately 10 acres of headwater seep habitat, extended beyond survey into neighboring private property. Heavily invaded by <i>Murdannia keisak</i> and <i>Microstegium</i> . 1000s of <i>Platanthera</i> leaves observed, samples collected at 7 locations throughout seep wetland	<i>Acer rubrum, Liriodendron tulipifera, Nyssa sylvatica, Liquidambar styraciflua, Alnus serrulata, Viburnum nudum, Chasmanthum laxum, Osmundastrum cinnamomea, Osmunda spectabilis, Amauropelta novaboracensis, Lorinseria areolata, Lycopus virginicus, Lobelia cardinalis, Rhexia mariana, Eutrochium fistulosum, Sphagnum sp, Platanthera clavellata</i>	high

HECTARES SURVEY DATE	LOCATION	HSI MEAN ±SD	GAP Status	DESCRIPTION	<i>P. INTEGRILABIA</i> ASSOCIATED SPECIES PRESENT	HABITAT RANK
16.65 ha 9/8/23	Paddys Creek North Burke County NC DNCR, NC Land and Water Fund	0.72 ±0.25	3	Well-developed sphagnum bog at headwaters of a tributary to Paddy's Creek in northwest section of survey area. Herb layer 80-100% cover mostly <i>Dicanthelium sp</i> , <i>Carex sp</i> , and <i>Scirpus sp</i> . Dominant tree <i>Pinus strobus</i> . Many <i>Platanthera</i> leaves, samples collected at 2 locations.	<i>Acer rubrum</i> , <i>Liriodendron tulipifera</i> , <i>Nyssa sylvatica</i> , <i>Liquidambar styraciflua</i> , <i>Viburnum nudum</i> , <i>Chasmanthum laxum</i> , <i>Osmundastrum cinnamomea</i> , <i>Osmunda spectabilis</i> , <i>Amauropelta novaboracensis</i> , <i>Lorinseria areolata</i> , <i>Lycopus virginicus</i> , <i>Carex intumescens</i> , <i>Sphagnum sp.</i> , <i>Platanthera clavellata</i>	high
3.15 ha 8/23/23	Little TN River tributary Macon County Mainspring Conservation Trust	0.67 ±0.09	2	Wetland area along tributary stream with <i>Platanus occidentalis</i> . Significant invasive encroachment of <i>Ligustrum japonicum</i> and <i>Microstegium vimineum</i> . Canopy cover variable and substrate saturated in areas still connected to floodplain, but lacks significant <i>Sphagnum</i> and <i>Carex</i> component. Stream incised degrading habitat.	<i>Acer rubrum</i> , <i>Liriodendron tulipifera</i> , <i>Nyssa sylvatica</i> , <i>Amauropelta noveboracensis</i> , <i>Lycopus virginianus</i> , <i>Osmundastrum cinnamomea</i> , <i>Platanthera clavelata</i>	medium
15.39 ha 8/25/23	Bushy Head Mt, Parker Branch, Cherokee County Private	0.73 ±0.29	4	Headwater stream with some pockets of saturated soil in flat areas where stream is not incised. Variable canopy cover, but mostly shaded. Land previously grazed and selectively logged. <i>Sphagnum</i> component missing or very minor.	<i>Acer rubrum</i> , <i>Liriodendron tulipifera</i> , <i>Nyssa sylvatica</i> , <i>Alnus serrulata</i> , <i>Amauropelta novaboracensis</i> , <i>Lobelia cardinalis</i> , <i>Viburnum nudum</i> , <i>Platanthera clavellata</i> , few infrequent <i>Osmundastrum cinnamomea</i> , <i>Osmunda spectabilis</i> , <i>Lorinseria areolata</i>	medium
17.37 ha 8/30/23	Peachtree, Cherokee County Private	0.59 ±0.37	4	Habitat variable throughout survey area. Some headwater streams with pockets of saturated soils small sphagnum mats in flat inundated areas. Some areas of predicted habitat overly shaded with large <i>Rhododendrons</i> .	<i>Acer rubrum</i> , <i>Liriodendron tulipifera</i> , <i>Nyssa sylvatica</i> , <i>Viburnum nudum</i> , <i>Alnus serrulata</i> , <i>Osmundastrum cinnamomea</i> , <i>Amauropelta noveboracensis</i> , <i>Lorinseria areolata</i> , <i>Lycopus virginicus</i> , <i>Lobelia cardinalis</i> , <i>Platanthera clavellata</i>	medium

HECTARES SURVEY DATE	LOCATION	HSI MEAN ±SD	GAP Status	DESCRIPTION	<i>P. INTEGRILABIA</i> ASSOCIATED SPECIES PRESENT	HABITAT RANK
8.55 ha 8/31/23	Wehutty west, Cherokee County Private	0.78 ±0.29	4	Small relatively flat headwater stream west of the dirt road. Small patches of saturated floodplain with <i>Sphagnum</i> appear to be good habitat. Mostly shaded, but some small open canopy patches. Surprised to find no <i>P. clavellata</i> .	<i>Acer rubrum</i> , <i>Liquidambar styraciflua</i> , <i>Viburnum nudum</i> , <i>Alnus serrulata</i> , <i>Osmundastrum cinnamomea</i> , <i>Osmunda spectabilis</i> , <i>Amauropelta noveboracensis</i> , <i>Lobelia cardinalis</i> , <i>Sphagnum sp.</i>	medium
6.84 ha 9/15/23	King Creek, Henderson County Private, Ingels	0.66 ±0.16	4	Degraded non-alluvial wetland. Saturated soils with significant <i>Sphagnum/Carex</i> component but dominating young willows creating dense mid-story canopy, heavily invaded by <i>Murdannia keisak</i> .	<i>Acer rubrum</i> , <i>Liriodendron tulipifera</i> , <i>Alnus serrulata</i> , <i>Nyssa sylvatica</i> , <i>Carex intumescens</i> , <i>infrequent</i> , <i>Osmundastrum cinnamomea</i> , <i>Osmunda spectabilis</i> , <i>Lorinseria areolata</i> , <i>Lobelia cardinalis</i> , <i>Platanthera clavellata</i> , <i>Sphagnum sp.</i>	medium
11.97 ha 8/16/23	Green River Game Lands south Henderson Co NC State	0.45 ±0.35	2	Stream corridor with some small tributaries, too steep, floodplain disconnected from stream.	<i>Acer rubrum</i> , <i>Alnus serrulata</i> , <i>Liriodendron tulipifera</i> , <i>Amauropelta noveboracensis</i>	low
3.51 ha 8/16/23	Green River Game Lands north Henderson County NC State	0.32 ±0.32	4	Stream corridor with some small tributaries, too steep, floodplain disconnected from stream.	<i>Acer rubrum</i> , <i>Alnus serrulata</i> , <i>Liriodendron tulipifera</i> , <i>Amauropelta noveboracensis</i>	low
3.62 ha 8/17/23	King Creek west Henderson County Conserving Carolina	0.75 ±0.24	2	Stream-side upland with some nice alluvial wetlands where still connected with floodplain dominated by <i>Scirpus sp.</i> , variable canopy cover, lacking sphagnum component.	<i>Acer rubrum</i> , <i>Liriodendron tulipifera</i> , <i>Amauropelta noveboracensis</i> , <i>Lobelia cardinalis</i> , <i>Oxypolis rigidor</i> , <i>Eutrochium fistulosum</i>	low

HECTARES SURVEY DATE	LOCATION	HSI MEAN \pm SD	GAP Status	DESCRIPTION	<i>P. INTEGRILABIA</i> ASSOCIATED SPECIES PRESENT	HABITAT RANK
10.17 ha 8/18/23	Clear Creek Henderson County Private	0.89 \pm 0.19	4	Most areas north of Clear Creek highly unsuitable habitat, cleared and ditched agricultural land and horse pasture. South of Clear Creek heavily invaded by <i>Ligustrum sinense</i> , <i>Microstegium vimineum</i> , <i>Rosa multiflorum</i> , and <i>Elaeagnus umbellata</i> . Some very small patches of <i>Sphagnum</i> and <i>Carex</i> , but area mostly ditched and disconnected from flood plain.	<i>Acer rubrum</i> , <i>Liriodendron tulipifera</i> , <i>Lobelia cardinalis</i> , <i>Amauropelta novaboracensis</i>	low
14.04 ha 8/23/23	Kelly Cove Rd Macon County Mainspring Conservation Trust	0.58 \pm 0.31	2	Most drainages on property heavily shaded by dense Rhododendron. Hydrology impacted by ditching and roads. Small seep at base of slope on west edge of a flat area west of Burningtown Creek w/ <i>Sphagnum</i> and many <i>P. clavellata</i> heavily invaded by dense <i>Microstegium vimineum</i> .	<i>Acer rubrum</i> , <i>Liriodendron tulipifera</i> , <i>Nyssa sylvatica</i> , <i>Viburnum nudum</i> , <i>Amauropelta noveboracensis</i> , <i>Eutrochium fistulosum</i> few <i>Osmundastrum cinnamomea</i> and <i>Osmuda spectabilis</i> w/ <i>Platanthera clavellata</i> in a small seepy area with <i>Sphagnum</i> sp.	low
13.77 ha 8/23/23	Bartram Trail Macon County US Forest Service	0.34 \pm 0.30	3	Too steep and dry, not good habitat.	presence/absence survey only	low
5.22 ha 8/24/23	Lake Chatuge west Clay County Private, Cajam Properties	0.76 \pm 0.23	4	Disturbed upland area with incised stream adjacent to hay field.	presence/absence survey only	low
7.11 ha 8/24/23	Lake Chatuge east Stillhouse Creek Clay County Private, Cajam Properties	0.55 \pm 0.26	4	Too steep and dry, some stream side habitat but lacking perennial wetlands, no significant <i>Sphagnum</i> or <i>Carex</i> component.	presence/absence survey only	low

HECTARES SURVEY DATE	LOCATION	HSI MEAN ±SD	GAP Status	DESCRIPTION	<i>P. INTEGRILABIA</i> ASSOCIATED SPECIES PRESENT	HABITAT RANK
9.90 ha 8/25/23	Bushy Head Mt, Laurel Branch Cherokee County Private	0.36 ±0.27	4	Headwater stream with lots of <i>Rhododendron</i> shading out stream corridor.	<i>Acer rubrum</i> , <i>Liriodendron tulipifera</i> , <i>Nyssa sylvatica</i> , <i>Amauropelta novaboracensus</i> , <i>Lobelia cardinalis</i>	low
9.90 ha 8/31/23	East of Unaka, roadside survey, Cherokee County Private	0.31 ±0.316	4	Patches predicted to be suitable habitat converted to pasture, no private land access granted.	presence/absence survey only	low
17.82 ha 8/31/23	Wehutty east, Cherokee County Private	0.61 ±0.33	4	Heavily disturbed flat area with remnants of a non-alluvial wetland. Was actively being cleared, complete removal of woody vegetation and draining via ditching.	in margins around recently cleared area: <i>Acer rubrum</i> , <i>Liquidambar styraciflua</i> , <i>Liriodendron tulipifera</i> , <i>Parathelypteris noveboracensis</i> , <i>Lorinseria areolata</i> , <i>Viburnum nudum</i> , <i>Eutrochium fistulosa</i> , <i>Alnus serrulata</i>	low
15.03 ha 9/2/23	Pink Beds South, Transylvania County US Forest Service	0.33 ±0.26	3	Swamp forest - bog complex where not overgrown with <i>Rhododendron</i> . Few small pockets of saturated areas with well-developed sphagnum with significant <i>Carex</i> component, many PLIN associated species. One small patch of highly suitable habitat at north tip of survey area, added to Pink Beds North.	<i>Acer rubrum</i> , <i>Liriodendron tulipifera</i> , <i>Nyssa sylvatica</i> , <i>viburnum nudum</i> , <i>Dicanthelium sp.</i> , <i>Chasmanthum laxum</i> , <i>Osmundastrum cinnamomea</i> , <i>Osmunda spectabilis</i> , <i>Parathelypteris noveboracensis</i> , <i>Lorinseria areolata</i> , <i>Lycopus virginicus</i> , <i>Lobelia cardinalis</i> , <i>Eutrochium fistulosum</i> , <i>Carex intumescens</i> , <i>Platanthera clavellata</i> , <i>Sphagnum sp.</i>	low
8.91 ha 9/6/23	Blue Ridge Community College Henderson County Private	0.47 ±0.22	4	Degraded Red Maple swamp forest-bog complex with some wetland areas along old drainage ditch. New occurrence of <i>Sagittaria fasciculata</i> discovered. Lacks <i>Sphagnum</i> and other <i>Platanthera sp.</i> Property heavily invaded with <i>Rosa multiflora</i> , <i>Ligustrum</i> , and <i>Microstegium vimineum</i> . <i>Murdannia keisak</i> in open wetland east of Bat Fork.	<i>Acer rubrum</i> , <i>Liriodendron tulipifera</i> , <i>Nyssa sylvatica</i> , <i>Viburnum nudum</i> , <i>Alnus serrulata</i> , <i>Osmundastrum cinnamomea</i> , <i>Osmunda spectabilis</i> , <i>Amauropelta noveboracensis</i> , <i>Lorinseria areolata</i> , <i>Carex intumescens</i>	low

