Carolina Coastal Plain Ecological Forecasting

Utilizing NASA Earth Observations to Map Suitable Venus Flytrap Habitat in an Effort to Inform Conservation, Seed Banking, and Reintroduction in the Carolina Coastal Plain and Sandhills Regions

**Technical Report**

Final – November 18th, 2021

Monika Rock (Project Lead)

Katie Caruso

Jayne Lampley

Ashna Siddhi

Seamore Zhu

***Advisor:***

Marguerite Madden, University of Georgia (Science Advisor)

# 1. Abstract

Although the carnivorous plant Venus flytrap (*Dionaea muscipula*) is recognized globally, its native range is restricted to a small portion of the North and South Carolina Coastal Plain and Sandhills. Within this limited range, Venus flytrap populations are threatened by habitat loss, fire suppression, and poaching. NASA DEVELOP partnered with the North Carolina Botanical Garden (NCBG), the University of North Carolina Herbarium (NCU), and the North Carolina Natural Heritage Program (NCNHP) to support Venus flytrap conservation. The team developed models based on species presence data and environmental variables using the Software for Assisted Habitat Modeling (SAHM) to create a 2021 habitat suitability map for Venus flytrap. These models incorporated Earth observations collected by Landsat 8 Thermal Infrared Sensor (TIRS), Terra Moderate Resolution Imaging Spectroradiometer (MODIS), Advanced Land Observation Satellite (ALOS) Phased Array type L-band Synthetic Aperture Radar (PALSAR), and Sentinel-2 Multispectral Instrument (MSI). To predict areas at high risk of development, the team produced a 2050 land-use change map using TerrSet Land Change Modeler. The team found potential areas of conflict between predicted habitat and forecasted future development. Many areas of suitable habitat were concentrated along the coast where development was likely to occur, placing populations there at risk of extirpation. Overlaying suitable habitat with forecasted land change also identified suitable habitats with minimal risk of development, which may serve as lasting Venus flytrap habitat. These results can inform the NCBG and NCNHP’s conservation decision-making, including targeted seed banking, reintroduction, and prioritization of enduring habitats for protection and management.

**Key Terms**

Habitat suitability modeling, Random Forest, MARS, MaxEnt, Generalized Linear Model, Boosted Regression Tree, Google Earth Engine, vulnerable species

# 2. Introduction

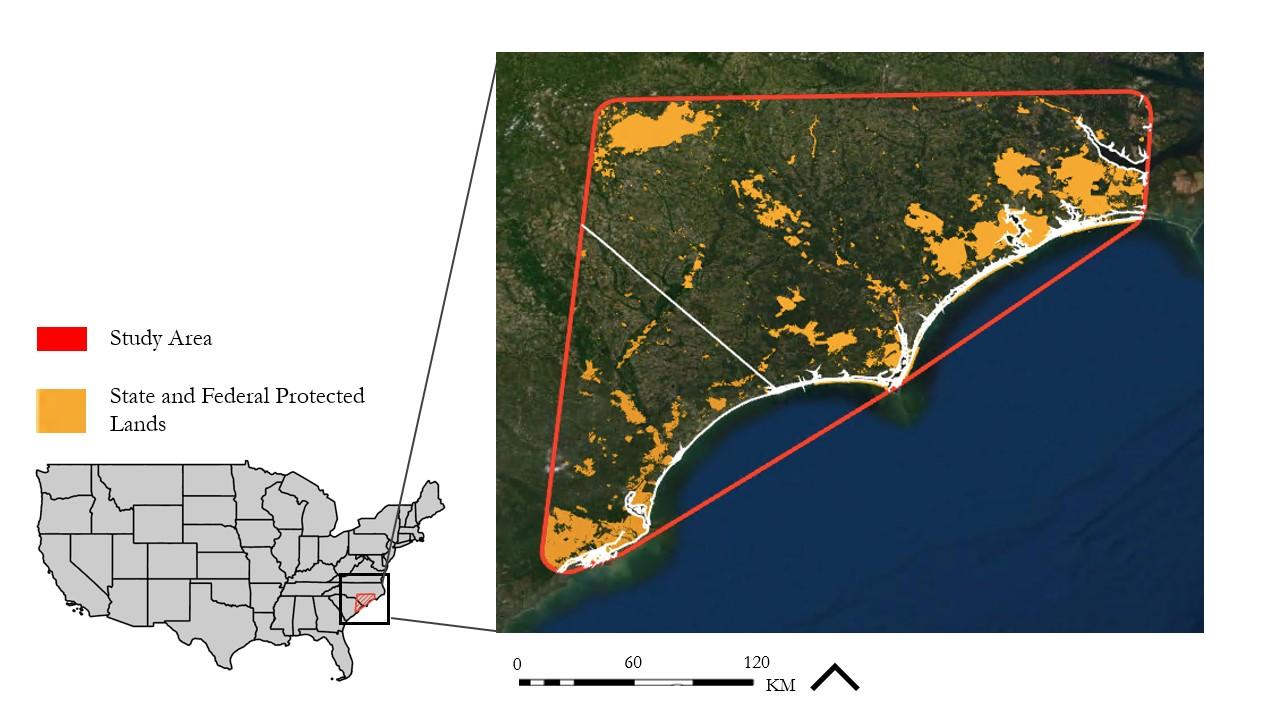
***2.1 Background Information***

Venus flytrap (*Dionaea muscipula* Ellis; VFT) is a carnivorous plant species endemic to the Coastal Plain and Sandhills regions of North and South Carolina (Luken, 2005; Roberts & Oosting, 1958). Throughout its range, VFT grows in wet pine savannas, bogs, pocosins, and in ecotones at the edges of Carolina bays (Luken, 2005; Roberts & Oosting, 1958). These habitats were historically maintained by high-frequency fire regimes and are characterized by sunny, open to semi-open conditions with acidic, sandy, nutrient-poor soils with little available nitrogen or phosphorus (Affolter et al., 2016; Gray et al., 2011). VFT supplements low nutrient availability by capturing and digesting insects using leaves modified into a unique active snap-trap. This mechanism evolved in a single evolutionary event and, among the approximately 810 known carnivorous plants, exists only in the monotypic sister genera *Dionaea* and *Aldrovandra* (Adamec et al., 2021; Cameron et al., 2002).

Despite widespread recognition of this carnivorous plant, VFT is threatened by habitat loss, fire suppression, climate change, and overexploitation from poaching (Affolter et al., 2016). Habitat loss, due to development and climate change, is a primary threat because this species is highly adapted to particular conditions and is highly sensitive to any environmental changes (Affolter et al., 2016; Luken, 2005). Fire suppression also drives habitat loss for VFT, which depends on frequent fire disturbance to reduce overstory light competition, increase access to insect prey, stimulate flower production, and promote seedling establishment (Roberts and Oosting, 1958; Hamon et al., 2019).

For rare and threatened plant species with narrow habitat requirements and patchy distributions, habitat suitability models are an important tool for predicting species distributions (Gogol-Prokurat, 2011). Such suitability models support species’ conservation by identifying land to be prioritized for protection, management, and/or reintroduction from seed banks and living collections (Costa et al., 2010). To support VFT conservation, this study developed a habitat suitability model for VFT using all known population occurrences and a variety of environmental predictor variables. Although other studies created range-wide maps of VFT based on population surveys, no other published studies have mapped habitat suitability for VFT across its range.

The study area used in this project covers most of the historical native range of VFT (*Figure 1*). This region occurs entirely within the Coastal Plain physiographic province over a portion of southeastern North Carolina and eastern South Carolina. In addition, the study area stretches northwestward from the Atlantic Coast to cover known population sites in the North Carolina Sandhills region, an area which occurs in the margin between the Coastal Plain and the Piedmont physiographic province. The physical features of the region included in the study area are characterized by relatively low elevation topography, river drainages, wetlands, and increasing land development.



*Figure 1.* The map of the study area, which includes the known range of VFT in North and South Carolina with a 10km² buffer around the known range; protected land was collected from the North Carolina Natural Heritage Program and The Nature Conservancy.

***2.2 Project Partners & Objectives***

The DEVELOP team partnered with the North Carolina Botanical Garden (NCBG), the University of North Carolina Herbarium, and North Carolina Natural Heritage Program (NCNHP) to address VFT habitat loss. Our partner organizations are currently working to protect the species and its habitat. The NCBG conducts fieldwork, genetic research, land conservation, and outreach to protect native and culturally significant plants, including the carnivorous VFT. The NCNHP gathers information from respective studies and research to ensure public awareness and access to information necessary to assess the potential ecological impacts of conservation and development. In addition to this work, between 2018 and 2020 the US Fish and Wildlife Service (USFWS) contracted NCNHP to conduct a status survey of VFT to inform a status assessment of the species. The goal of this assessment was to guide decisions to federally list VFT as either “Threatened” or “Endangered” under the Endangered Species Act of 1973.

The partner organizations do not currently use Earth observations in their research or decision making. They are, however, interested in the utilization of maps derived from satellite imagery to advance the protection and conservation of native plants. Hence, the objectives of this project were to utilize satellite imagery to model current habitat suitability for VFT based on species presence data and environmental predictor variables for 2021 and create a land cover map for 2050. The Software for Assisted Habitat Modelling (SAHM) package within VisTrails v2.2.4 platform was used to model the current habitat suitability for VFT within and outside of its current known range (Morisette et al., 2013; Bavoil et al., 2005). Land Change Modeler (LCM) within TerrSet was then used to model the potential for land to transition to developed land cover by 2050. An ArcGIS Online StoryMap was also created to communicate the project’s results to the partner organizations along with other VFT stakeholders and the general public.

# 3. Methodology

***3.1 Data Acquisition***

A total of 400 polygons and species presence points were provided by project partners based on herbarium records and field survey data. These data were collected from 1920 to 2020 and include an extensive list of known VFT populations throughout the extent of its range in North and South Carolina. North Carolina data also included location data for VFT associate species: golden sedge (*Carex lutea*), Hooker’s milkwort (*Polygala hookeri*), Baldwin’s nutrush (*Scleria baldwinii*), rough-leaf loosestrife (*Lysimachia asperulifolia*), savannah nutrush (*Scleria verticillata*), and hooded pitcher-plant (*Sarracenia minor*). The term associate species refers to co-occurring species of plant assemblages which are often good indicators of a particular habitat or other associated species (Dufréne and Legendre, 1997). The study area encompasses all known locations of present-day and historical VFT populations provided by project partners. This list of population sites included locations from records dating back at least to 1920. The study area was created by drawing a polygon around known sites using the ArcGIS Pro application. Additionally, a 10km buffer zone was created around the polygon, expanding the considered area to approximately 41,400 km2.

Google Earth Engine (GEE) was the primary resource used to gather predictor variables for habitat suitability modeling (Table 1). Land surface temperature data, which influence habitat suitability based on VFT’s surface temperature tolerances, were derived from Landsat 8 Thermal Infrared Sensor (TIRS) and Terra Moderate Resolution Imaging Spectroradiometer (MODIS). Fire frequency of approximately every three years plays a key role in maintaining suitable habitat (Affolter et al., 2016). Therefore, several fire variables (see Table 1) from an ancillary dataset, LANDFIRE, were gathered. As VFT prefers flat terrain at approximately two to four meters above sea level, topographic variables were obtained from the Japan Aerospace Exploration Agency (JAXA) Advanced Land Observation Satellite (ALOS) Phased Array type L-band Synthetic Aperture Radar (PALSAR) digital surface model (DSM) and digital elevation models (DEM) (Waller et al., 2016). Precipitation data was acquired through the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS). Vegetation indices, which were obtained from the Sentinel-2 Multispectral Instrument (MSI) Level-2A dataset, are also predicted to correlate with the very distinctive and specific VFT habitats. Soil characteristics, chiefly soil type, pH, nutrient level and moisture, among others, are also an important determinant of VFT habitat (Affolter et al. 2016). Consequently, soil type data from the United States Department of Agriculture (USDA) Web Soil Survey as well as soil pH and texture from International Soil Reference and Information Centre (ISRIC: World Soil Information) were obtained. Along with soil data, all belowground water content variables were obtained from the USDA gridded Soil Survey Geographic (gSSURGO) geodatabase, which is derived from the USDA Natural Resources Conservation Service (NRCS) SSURGO Database.

ArcGIS Online was the primary resource used to gather four explanatory variables to map land cover changes using LCM, which forecasts potential land transitions, within TerrSet (Table 2). Each variable acquired explained the driving forces for development in the study area. These driving variables included distance to coast, distance to golf courses, distance to urban areas, and slope. The raw layers gathered from ArcGIS Online to create these explanatory variables were, USA Golf Courses and major urban areas in the United States designated in the U.S. Census Populated Place Areas. Coastal boundaries of the United States were identified in ArcGIS Pro using the polygon tool to calculate the distance to the coast. Lastly, landcover images required for LCM were gathered from the National Land Cover Database (NLCD) from the Multi-Resolution Land Characteristics Consortium (MRLC) website.

Table 1

*Predictor variables acquired for VFT habitat suitability modeling.*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **SN** | **Predictor Variable** | **Source** | **Native Resolution** | **Image Date** | **Access/** **Processing** | **Included in final model (Yes/No)** |
| 1 | Land Surface Temperature | Landsat 8 TIRS | 30m | 4/1/21 - 7/31/21 | GEE | Yes |
| 2 | Land Surface Temperature | Terra MODIS | 250m | 4/1/21 - 7/31/21 11/1/20 - 2/28/21 | GEE | Yes |
| 3 | Soil Type | gSSURGO | 10m | 01/26/21 | USDA/ArcGIS Pro | Yes |
| 4 | Available Soil Water Storage (at a depth from 0-25, 25-50cm) | gSSURGO | 10m | 01/26/21 | USDA/ArcGIS Pro | Yes |
| 5 | Drainage Class | gSSURGO | 10m | 01/26/21 | USDA/ArcGIS Pro | Yes |
| 6 | Ponding Frequency | gSSURGO | 10m | 01/26/21 | USDA/ArcGIS Pro | No |
| 7 | Water Table Depth | gSSURGO | 10m | 01/26/21 | USDA/ArcGIS Pro | Yes |
| 8 | Hydric Class | gSSURGO | 10m | 01/26/21 | USDA/ArcGIS Pro | Yes |
| 9 | Soil Moisture (at a depth of 0, 10, 30 cm) | gSSURGO | 10m | 01/26/21 | USDA/ArcGIS Pro | Yes |
| 35 | Flood Frequency Maximum | gSSURGO | 10m | 01/26/21 | GEE | Yes |
| 10 | Soil pH (at a depth of 0, 10, 30 cm) | ISRIC: World Soil Information | 250m | 01/01/18 | GEE | Yes |
| 11 | Soil Texture (at a depth of 0, 10, 30 cm) | ISRIC: World Soil Information | 250m | 01/01/18 | GEE | No |
| 12 | Soil Nitrogen (at a depth of 5, 15, 30 cm) | ISRIC: World Soil Information | 250m | N/A | GEE | Yes |
| 13 | Clay Content (at a depth of 0, 10, 30cm) | ISRIC: World Soil Information | 250m | 01/01/18 | GEE | No |
| 14 | Sand Content (at a depth of 0, 10, 30 cm) | ISRIC: World Soil Information | 250m | 01/01/18 | GEE | No |
| 15 | Elevation | ALOS DSM | 30m | 1/24/06- 5/12/11 | GEE | Yes |
| 16 | Slope | ALOS DSM | 30m | 1/24/06- 5/12/11 | GEE | Yes |
| 17 | Topographic Diversity (TD) | ALOS DEM | 30m | 1/24/06- 5/12/11 | GEE | Yes |
| 18 | Continuous Heat Insolation Load Index (CHILI) | ALOS DEM | 30m | 1/2/06- 5/12/11 | GEE | No |
| 19 | Multi-Scale Topographic Position Index (mTPI) | ALOS DEM | 30m | 1/24/06- 5/12/11 | GEE | Yes |
| 20 | Landform | ALOS DEM | 30m | 1/24/06- 5/12/11 | GEE | Yes |
| 21 | Normalized Difference Vegetation Index (NDVI) | Sentinel-2 MSI | 10m | 5/1/20 - 7/31/20 | GEE | No |
| 22 | Modified Soil Adjusted Vegetation Index 2 (MSAVI2) | Sentinel-2 MSI | 10m | 5/1/20 - 7/31/20 | GEE | No |
| 23 | Normalized Difference Water Index (NDWI) | Sentinel-2 MSI | 10m | 5/1/20 - 7/31/20 | GEE | Yes |
| 24 | Tasseled Cap Brightness | Sentinel-2 MSI | 10m | 5/1/20 - 7/31/20 | GEE | No |
| 25 | % Low Severity Fire | LANDFIRE | 30m | 12/31/10 | GEE | No |
| 26 | % Mixed Severity Fire | LANDFIRE | 30m | 12/31/10 | GEE | No |
| 27 | % Replacement Severity Fire | LANDFIRE | 30m | 12/31/10 | GEE | No |
| 28 | Historical Mean Fire Interval | LANDFIRE | 30m | 12/31/10 | GEE | No |
| 29 | Historical Fire Regime | LANDFIRE | 30m | 12/31/10 | GEE | No |
| 30 | Vegetation Departure from Historical Fire | LANDFIRE | 30m | 12/31/10 | GEE | No |
| 31 | Vegetation Succession Classes (Fire) | LANDFIRE | 30m | 12/31/10 | GEE | Yes |
| 32 | Vegetation Classes (Fire) | LANDFIRE | 30m | 12/31/10 | GEE | No |
| 33 | Land Cover | National Land Cover Database | 30m | 2016 | GEE | No |
| 34 | Precipitation | Climate Hazards Group CHIRPS | 30m | 4/1/20 - 7/31/20 | GEE | Yes |
| 36 | Tree Cover | Landsat 8 | 30m | 2000, 2005, 2010 | GEE | Yes |
| 37 | Tasseled Cap Wetness | Sentinel-2 MSI | 10m | 5/1/20 - 7/31/20 | GEE | Yes |
| 38 | Tasseled Cap Greenness | Sentinel-2 MSI | 10m | 5/1/20 - 7/31/20 | GEE | No |
| 39 | Existing Vegetation Cover | LANDFIRE | 30m | 12/31/10 | GEE | Yes |

Table 2

*Explanatory variables acquired for future potential land cover transition mapping.*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **SN** | **Predictor Variable** | **Source** | **Native Resolution** | **Image Date** | **Access/** **Processing** | **Included in final model (Yes/No)** |
| 1 | Distance to Coast | ArcGIS Online: National Oceanic and Atmospheric Administration (NOAA) | 0.0001m | N/A | ArcGIS Pro | Yes |
| 2 | Distance to Golf Course | ArcGIS Online:  USGS-GNIS, Esri | 0.000000001 Degree | N/A | ArcGIS Pro | Yes |
| 3 | Distance to Urban Areas | ArcGIS Online: ESRI | 0.000000001 Degree | N/A | ArcGIS Pro | Yes |
| 4 | Slope | ALOS DSM | 30m | 1/24/06- 5/12/1 | GEE  ArcGIS Pro | Yes |

***3.2 Data Processing***

***3.2.1 Presence Points***

The presence polygons of North Carolina VFT populations received from the NCNHP contained population quality rankings and location accuracy information. Population quality rankings were as follows: A (Excellent), B (Good), C (Fair), D (Poor), E (Extant), F (failed to find), H (historical), and X (extirpated). Accuracy rankings ranged from 1-5 (1 being considered very high accuracy by the NCNHP, and 5 being very low accuracy). Presence polygons ranked F, H, or X or with rankings of 4 or 5 were removed from the project due to likely absence of present populations in these locations and for low geographic accuracy. Similarly, the species presence data for South Carolina provided by the South Carolina Natural Heritage Program (SCNHP) contained population quality and accuracy rankings. Based on these rankings, polygons with higher ranking, which indicated lower accuracy were removed from the project. Furthermore, although historical data informed the creation of our study area, they were not used to train the SAHM model if they were listed by NCNHP or SCNHP as F, H, or X. This was done to ensure that the model of currently suitable habitat was trained by extant population data only. Lastly, presence data provided by NCNHP of VFT associate species were not used in the current study due to time constraints on the project.

***3.2.2 Predictor Variables***

Predictor variables were preprocessed in GEE before being imported into SAHM. The preprocessing performed in GEE included clipping each variable to the extent of the study area as well as filtering images and masking to reduce cloud cover interference, which supports the accuracy and completeness of our model. Landsat 8 and Sentinel-2 images were filtered for cloud coverage to a minimum cloud coverage that provided at least 50 images and image coverage over our entire study area. Landsat 8 images over our date range were filtered to only include images with under 80% cloud cover and masked using the cloud, cirrus, and cloud shadow bit masks. Sentinel-2 images over our date range were filtered to only include images with under 17% cloud cover and masked using the Sentinel-2 cloud probability dataset to remove pixels identified as >50% cloud probability. Some variables such as precipitation and temperature were filtered by date range to provide consistency. Calculations to gather different indices that were used as predictor variables were also performed within GEE (Table 3). To determine the health of vegetation in the study area, Normalized Difference Vegetation Index (NDVI) was calculated in GEE as the ratio of the difference between portions of the red and near-infrared spectrum to the sum of portions of the red and near-infrared spectrum (Table 3; U.S. Geological Survey [USGS], n.d.). To minimize the effect of soil background, the team used Modified Soil Adjusted Vegetation Index 2 (Table 3, MSAVI2, Jiang et al. 2007). Since VFT is a small plant that is often covered by taller overstory plants it can be hard to identify remotely. To compensate for this, differenced greenness and tasseled cap indices were used to identify vegetation that grew in similar habitats and would therefore indicate suitable VFT habitat (Table 3, Tasseled cap indices; Lastovicka et al., 2020). Changes in vegetation water content were calculated with the Normalized Difference Water Index (NDWI; Table 3; Gao, 1996). For soil data with different depths (0, 10, 30cm), each individual layer at a different depth was extracted for a total of three separate predictor variables. Other soil predictor variables, gathered from the United States Department of Agriculture (USDA) gSSURGO geodatabase, which is derived from the USDA SSURGO database, were preprocessed in ArcGIS Pro by clipping soil layers to the extent of the study area. Each variable was then extracted and saved as a tiff file. These predicted variables were then further preprocessed within SAHM. First, each variable was defined as categorical (1) or continuous (0), then resampled using nearest-neighbor approach and finally aggregated using mean method.

Predictor variables used for land change modeling were preprocessed in ArcGIS Pro before being imported into TerrSet. The preprocessing performed in ArcGIS Pro included clipping each layer to the extent of the study area. It also required calculating the Euclidean distance from coasts, cities, and golf courses using the distance tool, and then saving these files as tiff files.

Table 3

*Equations for deriving indices used as predictor variables for habitat suitability modeling. All indices were calculated in GEE using Sentinel-2 data at 10m resolution.*

|  |  |  |
| --- | --- | --- |
| **Predictor Variable** | **Equation** | **Included in final model (Yes/No)** |
| Normalized Difference Vegetation Index (NDVI) |  | No |
| Modified Soil Adjusted Vegetation Index 2 (MSAVI2) |  | No |
| Normalized Difference Wetness Index (NDWI) |  | Yes |
| Tasseled Cap Brightness (TCB) |  | No |
| Tasseled Cap Greenness (TCG) | (−0.2848 × BLUE) − (0.2435 × GREEN) − (0.5436 × RED) +  (0.7243 × NIR) + (0.0840 × SWIR1) − (0.1800 × SWIR2) | No |
| Tasseled Cap Wetness (TCW) | (0.1509 × BLUE) + (0.1973 × GREEN) + (0.3279 × RED) +  (0.3406 × NIR) – (0.7112 × SWIR1) – (0.4572 × SWIR2) | Yes |

***3.3 Data Analysis***

***3.3.1 SAHM***

Species occurrence data and the predictor variables list were gathered into the SAHM project workflow. The processed study area raster provided a template layer, which was used to define the study extent, 30m resolution, and World Geodetic System 1984 (WGS 84) coordinate system for analysis. All other raster layers were processed upon project execution to match the template layer. Because the point data used to train the model were presence-only points collected in field surveys, we generated 3000 background points within a binary mask using a kernel density of our presence points to compensate for the lack of absence points.

Next, we selected final predictor variables for our modeling. This selection was informed by comparing how well each predictor explained the locations of sampled occurrence and background points, the correlation between variables, and background knowledge of the habitat requirements of VFT. Predictor variables that explained sample data distribution, signified by higher percent deviance explained, were prioritized. For variables that had a covariate correlation above 0.7, we removed the variable with a lower percent deviance explained. This was done to avoid overfitting our model. By updating the covariate correlation plots regularly between removing variables, we ensured key predictors that capture unique aspects of suitable habitat were not omitted. We refined our predictor variables from 50 to 23. These variables, in decreasing order of percent deviance explained, were soil pH at 0 cm, mean nitrogen at 15cm, canopy cover, drainage class, flood frequency, existing vegetation cover, available water storage from a depth of 25-50cm, hydric class, soil taxonomy, landform, topographic diversity, soil moisture, tasseled-cap wetness, fire succession class, NDWI, elevation, early growing season land surface temperature, slope, dormant season land surface temperature from Terra MODIS, mTPI, growing season land surface temperature from Terra MODIS, water table, and precipitation.

Five separate models within SAHM were evaluated: Boosted Regression Tree (BRT), Generalized Linear Model (GLM), MaxEnt, Multivariate Adaptive Regression Spline (MARS), and Random Forest (RF; see model explanations in Breiman, 2001; Phillips et al., 2006; Salas et al., 2017). These regression models are commonly used in habitat modeling. For MARS, all five categorical variables were excluded due to the difficulty of processing categorical data in that model. Similarly, soil taxonomy was excluded from the RF model because RF cannot process over 53 no-data categories within one dataset. From these models, we generated probability maps for habitat suitability as well as response curves, confusion matrices, and plots. To select the most suitable model for predicting suitable habitat for VFT among the five models run, we considered area under the curve (AUC), the calibration plot for the cross-validation split, response curves for each predictor variable used, and the spatial pattern of deviance residuals.

***3.3.2 LCM***

A land cover change map was generated using LCM within TerrSet to determine the potential of areas throughout the current and historical range of VFT to transition to a different land cover type by 2050. Land cover can transition to developed land cover from the following undeveloped categories in the NLCD: water, barren, forest, shrubland, herbaceous, planted/cultivated, and wetlands. Two land cover maps were inputted to forecast land transition potential: past land cover captured in 2001 and the most recent available land cover map, which was created from data gathered in 2019. Change was assessed between these two time periods to create a future land change map. The land change map created was further analyzed to identify the dominant transition of interest. Based on this analysis and suggestions from partner organizations, development was concluded to be a major cause of habitat loss for VFT. Consequently, transition to developed region was grouped and modeled to create a future land change map for 2050.

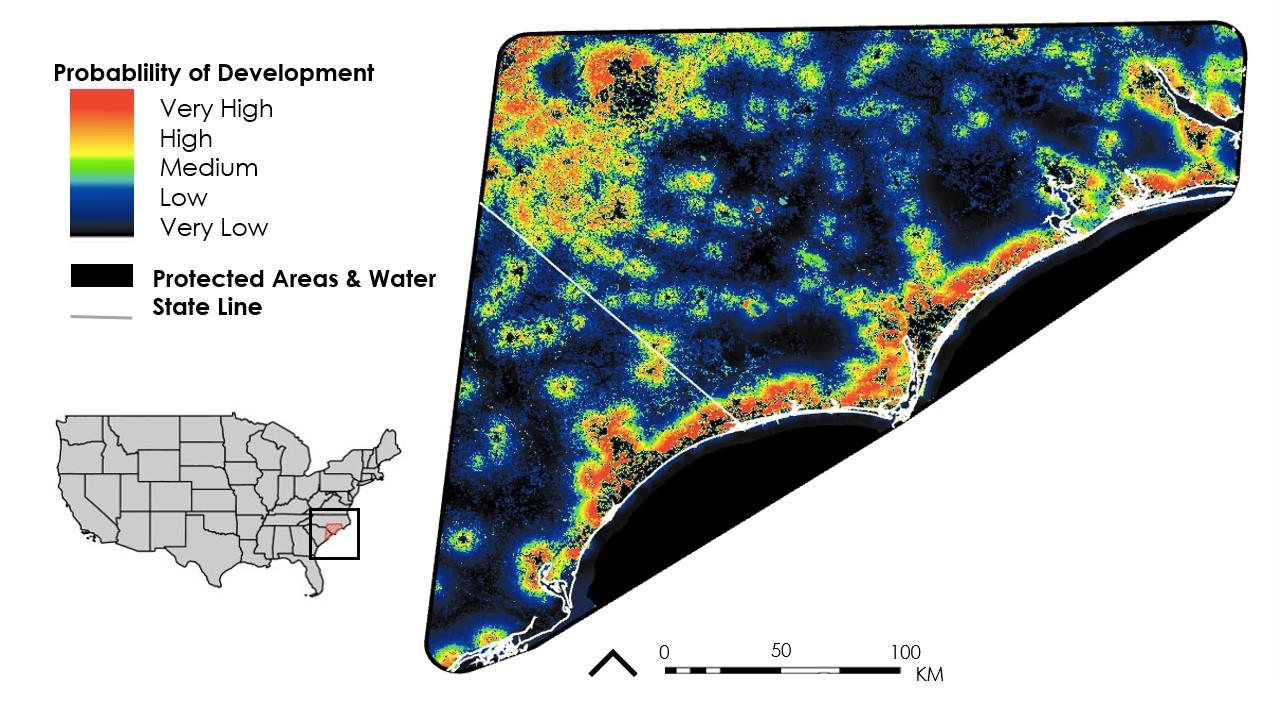
Each preprocessed explanatory variable was gathered in the LCM as an explanatory variable driving development in the range of VFT. With the help of these variables, the historical change from non-developed land cover to developed regions between 2001 and 2019 was generated. These historical changes to developed regions were further modeled to transition potential models using the Multi-layer Perceptron (MLP) neural network. This process was repeated if the accuracy assessment of the model was less than 80%.

Once calibrated, LCM used the historical rate of change and the transition potential models to predict a future scenario for the year of 2050. Since the final map would be used for habitat assessment, a soft prediction model was used. A soft prediction model yields a map of vulnerability to change to a developed region, whereas a hard prediction model is based on a competitive land allocation model. Once the final map of vulnerability to change was generated, the potential land transition map was overlayed with predicted suitable habitat of VFT to determine where the species may be threatened by development in the future.

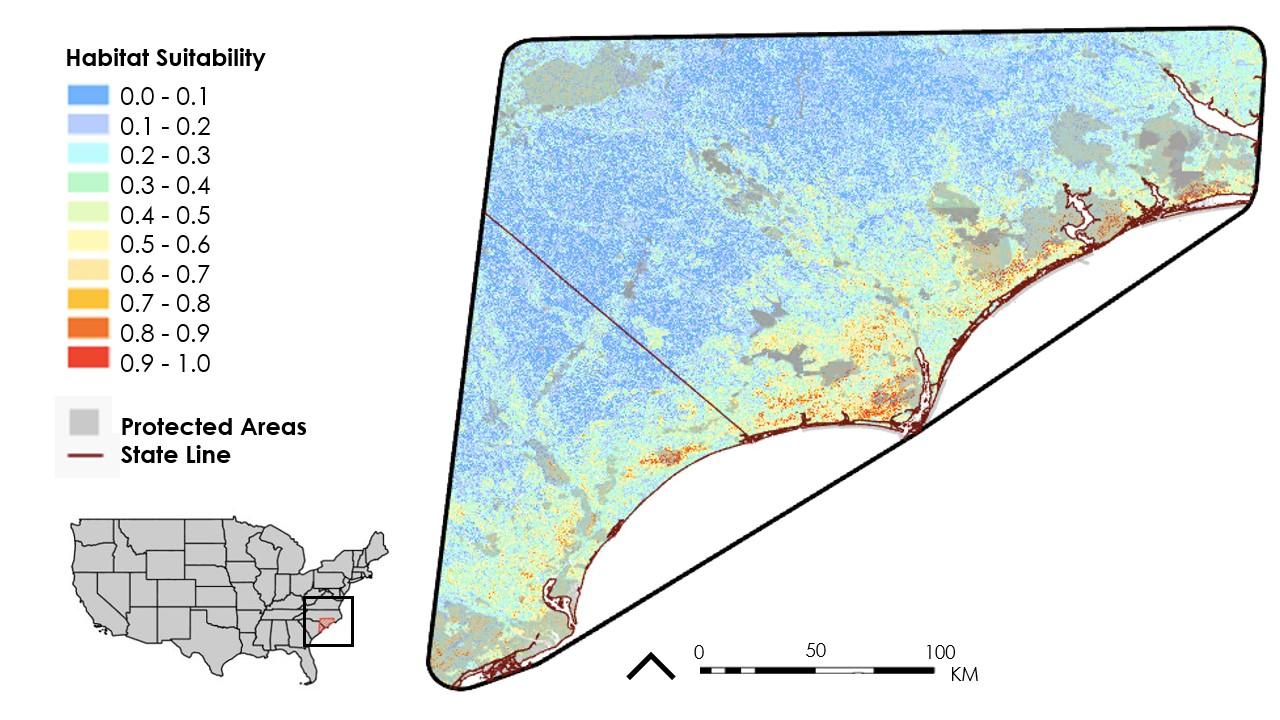
# 4. Results & Discussion

***4.1 Analysis of Results***

Of our five models, RF was most useful for identifying suitable habitat for VFT. Statistics that stood out in the RF model included an AUC of 0.92 for the training split and an AUC of 0.919 for testing data, the difference between the two being 0.001. The calibration plot, response curves, and plot of residuals also demonstrated equal or better performance than all other models. Variable importance in the RF model was calculated using the change in AUC when each predictor variable is permuted. The three most important variables for the RF model were available water storage from a depth of 25-50cm, followed by mean nitrogen at a depth of 15cm and existing vegetation cover.



*Figure 2.* Probability of land to transition to developed land cover by 2050.



*Figure 3*. Map of suitable habitat for VFT generated by RF model.

To analyze the results of our RF model and 2050 land cover change model, we used ArcGIS Pro. The RF probability map depicted predicted probability of habitat suitability over our study area at the 30m resolution defined by our template layer. The 2050 land cover change map was added as a layer in order to compare and identify areas of overlap (*Figure* 2). We also incorporated geometry data representing currently protected lands in North and South Carolina for analysis.

First, we classified probability of habitat suitability from 0% to 100% by increments of 10%. The map of current habitat suitability matched our prior knowledge about VFT habitat. The most important determinants of suitable habitat for VFT in the RF model were available water storage between 25-50cm and nitrogen content. These factors correspond with environmental factors identified to delineate VFT habitat in other studies. Soil moisture is an important determinant of VFT habitat as the species grows in narrow ecotones with suitable water regimes. VFT populations are sensitive to changes in moisture and must shift following changes in water table levels (Affolter et al., 2016). Suitable VFT habitat also requires particular soil nitrogen levels. In *in situ* experiments, Roberts and Oosting (1958) found that VFT growth was limited at both high and low nitrogen levels. The addition of nitrogen to suitable habitats threatens VFT populations by supporting the growth of plants that outcompete VFT (Affolter et al., 2016). Existing vegetation cover, topographic diversity, soil pH, canopy cover, NDWI, and landform were other important variables for determining suitable habitat for VFT.

High suitability was generally predicted in the Coastal Plain region of North and South Carolina. Overall, suitability sharply decreases further inland with a tail of moderately high suitability areas around and between the Cape Fear, Black, and South Rivers. There is another cluster of areas with moderately high probability to contain suitable habitat in the Sandhills region, which is found in the northwest portion of our study area (*Figure* 3). Our land cover change map also found that probability of development is highest along the coast and in the northwest portion of our study area. These areas of potential transition are visually similar to areas predicted to contain highly suitable VTF habitat (*Figure* 2).

We used a threshold of 80% probability of VFT presence to identify highly suitable habitat for the species in our analyses. This threshold was identified by comparing suitable habitat predicted by the RF model with locations of current and historical VFT populations. The 80% threshold identified an area that represents a conservative prediction of suitable habitat. We found this predicted area to be 355.1km2. Out of this total predicted suitable habitat, 17km2 overlapped with currently protected lands. This overlap represents predicted suitable habitat currently protected from development, which comprises 47.9% of the total predicted suitable area (Table 4).

Table 4

*Area of RF predicted habitat suitability overlapping with currently protected lands.*

|  |  |
| --- | --- |
| **Statistic** | **Area** |
| > 80% probable suitable habitat | 355.1km2 |
| > 80% probable suitable habitat on currently protected lands | 170.0km2 |
| Percentage of > 80% probable suitable habitat on currently protected lands | 47.9% |

We also determined the amount of suitable habitat at risk of degradation through development by 2050. We chose a probability threshold above 50% to identify areas where development is predicted to be more likely than not. However, our land cover change map did not take into account protected areas, so area of conflict may be an overestimate. This area of probable land change is 7532.1km2. The area of conflict between > 80% probable suitable habitat and predicted land change is 197.4km2. This area represents 55.6% of currently predicted suitable habitat, meaning over half of currently predicted suitable habitat could be threatened by land cover change by 2050 (Table 5).

Table 5

*Area of RF predicted habitat suitability compared with area of probable development from the 2050 land cover change map.*

|  |  |
| --- | --- |
| **Statistic** | **Area** |
| > 80% probable suitable habitat | 355.1km2 |
| > 50% probable land change | 7532.1km2 |
| Area of conflict between predicted suitable habitat and predicted land change | 197.4km2 |
| Percentage of > 80% probable suitable habitat threatened by > 50% probable land change | 55.6% |

***4.2 Limitations***

While this project successfully generated a well-fit habitat suitability model and highlighted conflict between future development and currently suitable habitat for VFT, it was not without its limitations. First, our study area was limited to observed occurrences only. The extent of the study area may have inadvertently excluded areas of suitable habitat where populations grew historically but were extirpated before being recorded. Second, although SAHM generates absence points when absence data is not available, collected VFT absence points would create a more accurate model of suitable habitat; habitat suitability models trained using presence and absence data are considered the gold standard for habitat modeling. Third, SAHM reduces all populations of VFT to one point regardless of size. Therefore, populations of 2,000 and 10 are weighted equally and, thus, the modeling process does not distinguish habitats that support large numbers of VFT from habitat that support small populations. Fourth, there were fewer presence points collected in the northwest portion of the study area in the Sandhills region of North Carolina, where habitat differs from that of the Coastal Plain. This may have caused the model to extrapolate this area, leading to less accurate habitat modeling in the Sandhills region than in the Coastal Plain. Lastly, TerrSet modeling focused on anthropogenic development and did not forecast transition in other land cover types. By excluding transitions among undeveloped land cover types, the LCM may have excluded transitions from suitable to unsuitable habitat in natural areas such as wetland to forest.

***4.3 Future Work***

Although this study focused on modeling suitable habitat in the known range of VFT in North and South Carolina, future studies could forecast suitable habitat for the species beyond this region based on changes in climate and sea level. Newly identified populations could be integrated into the study area to further refine the models. Future work on the conservation of VFT will support the conservation of associated ecosystems because, as a habitat specialist, VFT is highly adapted to particular environments in the Carolina Coastal Plain and Sandhills regions. Associate species of the VFT could be included in habitat modeling to support the ecosystem as a whole. Lastly, our partners are interested in studying the sharp contrast between the VFT habitat in the Sandhills and the VFT habitat in the remainder of the Coastal Plain as the Sandhill region is a sharply distinct physiographic region. Future work can examine these two regions separately to compare the similarities and differences between suitable habitats in each region.

# 5. Conclusions

Throughout its limited range in North and South Carolina, VFT grows only in the Coastal Plain and Sandhills regions. The majority of suitable habitat for VFT identified in this habitat suitability model was concentrated near the Atlantic coastline (*Figure* 3). VFT likely once inhabited a range spanning nearly 500km² from central North Carolina to southern South Carolina, and the RF model of habitat suitability identified 355.1km² in this range with highly suitable habitat (greater than 80% suitability) for the species (*Figure* 3). These areas of highly suitable habitat can be prioritized for protection and reintroduction. Areas of highly suitable habitat may also contain previously unknown VFT populations. Survey efforts for new populations may focus on these areas, which have the proper environmental factors to support VFT growth.

Habitat loss through land use change is a primary threat to VFT, and this study identified a number of areas within VFT’s range with high potential to transition to developed land cover by 2050. Forecasted development, like suitable habitat, is concentrated along the coast of North and South Carolina (*Figure* 2). Based on comparisons between the habitat suitability model and LCM, 55.6% of the areas predicted to have highly suitable habitat for VFT also have high (greater than 50%) potential to become developed by 2050. Locations with known VFT populations and high potential to transition to developed land cover by 2050 represent areas where VFT’s long-term survival may be threatened. These areas of conflict can be prioritized for protection or seed banking to conserve the genetic diversity of populations at high risk of development.

VFT is a charismatic species that has dramatically declined due to fire suppression, habitat loss, and poaching. The purpose of this project was to create a habitat suitability model for VFT to predict the potential distribution of the species and to forecast habitat suitability for 2050. These models seek to inform conservation of VFT and the current assessment by the USFWS to determine if VFT should be listed under the Endangered Species Act. Additionally, since SAHM workflows are easily replicated, both the partners and future researchers can use these methods to refine or expand upon the habitat suitability model for VFT created in this study.

# 6. Acknowledgments

The authors wish to thank everyone who assisted in this study, but especially our project partners: Dr. Johnny Randall, Dr. Damon Waitt, Misty Buchanan, and Dr. Alan Weakley. We also thank Meredith Wojcik, Dale Suiter, Sarah McRae, Stephanie Horton, Joseph Lemeris, Keith Bradley, Dr. Clyde Sorenson, Dr. Natalie Kerr, Julie Moore, Brenda Wichmann, Emily Oglesby, Laura Hamon and the VFT Working Group for data and insights on the VFT and its habitat. We also thank Dr. Marguerite Madden for advising this study and Dr. Kenton Ross, Peder Engelstad, Alex Posen, Sean McCartney, and Sarah Payne for technical assistance.

This material contains modified Copernicus Sentinel data (2020 – 2021), processed by ESA.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Aeronautics and Space Administration.

This material is based upon work supported by NASA through contract NNL16AA05C.

# 

# 7. Glossary

**AIC** –The Akaike Information Criterion; AIC is a criterion or standard which is used to help select the most appropriate model from a set of models by measuring how closely each model fits the data

**AUC** – Area Under the [ROC] Curve; A value depicting the probability a model will rank a randomly chosen presence observation higher than a randomly chosen absence observation

**BRT** –Boosted Regression Tree; utilizes recursive binary splits to evaluate response predictors (regression trees) while also combining many smaller, simpler models to improve prediction (boosting)

**dAUC** – Differenced Area Under the [ROC] Curve; the difference between the training curve and cross validated curves, a fit model has a difference of ≤ 0.05

**Earth observations** – Measurements and remotely sensed imagery of the Earth’s surface collected using Earth-orbiting satellites with sensors which provide information about the physical, chemical, and biological systems over space and time

**ESA** –Endangered Species Act of 1973; a comprehensive law of the United States which protects imperiled species in order to prevent extinction and to recover species

**GLM** –Generalized Linear Model; a basic linear regression that begins with a null model and calculates the AIC score for each variable that may be included in the model. The AIC scores covariates based on how well they fit the data, and GLM uses a stepwise procedure to first choose single covariates with the highest scores, then two-covariate models, then three, and so on. The result is a combination of covariates that provides the highest overall AIC score

**Google Earth Engine** –An open-source platform to visualize and process geospatial data

**Remote Sensing** –The process of detecting and monitoring the elements of a physical area by measuring its reflective radiation from a distance, usually by satellite or aircraft

**MODIS** – Moderate Resolution Imaging Spectroradiometer; NASA imaging sensors on the Terra and Aqua satellites which capture earth observations in 36 spectral bands at varying spatial resolutions

**NCBG** –North Carolina Botanical Garden; botanical garden located in Chapel Hill, North Carolina and operated by the University of North Carolina at Chapel Hill

**NCNHP** –North Carolina Natural Heritage Program;a program of the Division of Land and Water Stewardship within the North Carolina Department of Natural and Cultural Resources

**RF** – Random Forest; a general-purpose classification and regression model that combines multiple randomized decision trees and aggregates their predictions by averaging

**ROC** – Receiver Operating Characteristic Curve; used to display the performance of a model

**SAHM** –Software for Assisted Habitat Modelling; software which allows the user to incorporate multiple environmental predictor layers of a study area with species field sampling measurements into various habitat suitability models

**TerrSet Land Change Modeler** –Software which analyzes and models land cover change over time and provides information which may be used to assist land and conservation planning

**TSS** – True Skill Statistic; A summary statistic that considers specificity and sensitivity to measure the accuracy of a model’s predictions

**VisTrails** –An open-source scientific workflow data management system that supports data visualization and exploration

**Vulnerable Species** –A species identified by the IUCN Red List (International Union of Conservation of Nature) as being threatened with global extinction

# 8. References

Adamec, L., Matušíková, I., & Pavlovič, A. (2021). Recent ecophysiological, biochemical and evolutionary insights into plant carnivory. *Annals of Botany, 128*(3), 241–259. https://doi.org/10.1093/aob/mcab071

Bavoil, L., Callahan, S. P., Crossno, P. J., Freire, J., Scheidegger, C. E., Silva C. T., Vo, H. T. (2005). VisTrails: Enabling interactive multiple-view visualizations. *VIS 05. IEEE Visualization, 2005*, 135–142. https://doi.org/10.1109/VISUAL.2005.1532788

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32. https://doi.org/10.1023/A:1010933404324

Cameron, K. M., Wurdack, K. J., & Jobson, R. W. (2002). Molecular evidence for the common origin of snap-traps among carnivorous plants. *American Journal of Botany, 89*(9), 1503–1509. https://doi.org/10.3732/ajb.89.9.1503

Costa, G. C., Nogueira, C., Machado, R. B., & Colli, G. R. (2010). Sampling bias and the use of ecological niche modeling in conservation planning: A field evaluation in a biodiversity hotspot. *Biodiversity and Conservation*, *19*, 883–889. https://doi-org/10.1007/s10531-009-9746-8

Dufréne, M., & Legendre, P. (1997). Species assemblages and indicator species: The need for a flexible asymmetrical approach. *Ecological Monographs*, *67*(3), 345–366.

Earth Resources Observation and Science (EROS) Center. (2013). Collection 2 Landsat 8-9 OLI (Operational Land Imager) and TIRS (Thermal Infrared Sensor) Level-2 Science Product. USGS EROS Archive, accessed 19 October 2021. https://doi.org/10.5066/P9OGBGM6

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, *77*(4), 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Gao, B. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, *58*(3), 257–266. https://doi.org/10.1016/S0034-4257(96)00067-3

Gogol-Prokurat, M. (2011). Predicting habitat suitability for rare plants at local spatial scales using a species distribution model. *Ecological Applications, 21*(1), 33–47. https://doi.org/10.1890/09-1190.1

Gray, J. B., Wentworth, T. R., & Brownie, C. (2011). Extinction, colonization, and persistence of rare vascular flora in the longleaf pine-wiregrass ecosystem: Responses to fire frequency and population size. *Natural Areas Journal*, *23*(3), 210–219.

Hamon, L. E., Youngsteadt, E., Irwin, R. E., & Sorenson, C. E. (2019). Pollination ecology and morphology of VFT in sites of varying time since last fire. *Annals of the Entomological Society of America*, *112*(3), 141–149. https://doi.org/10.1093/aesa/say032

Jiang, Z., Huete, A. R., Li, J., & Qi, J. (2007). Interpretation of the modified soil-adjusted vegetation index isolines in red-NIR reflectance space. *Journal of Applied Remote Sensing*, *1*(1), Article 013503. https://doi.org/10.1117/1.2709702

Lastovicka, J., Svec, P., Paluba, D., Kobliuk, N., Svoboda, J., Hladky, R., & Stych, P. (2020). Sentinel-2 data in an evaluation of the impact of the disturbances on forest vegetation. *Remote Sensing*, *12*(12), Article 1914. https://doi.org/10.3390/rs12121914

Luken, J. O. (2005). Habitats of *Dionaea muscipula* (Venus’ fly trap), Droseraceae, associated with Carolina Bays. *Southeastern Naturalist*, *4*(4), 573–584. https://doi.org/10.1656/1528-7092(2005)004[0573:HODMVF]2.0.CO;2

Morisette, J. T., Jarnevich, C. S., Holcombe, T. R., Talbert, C. B., Ignizio, D., Talbert, M. K., Silva, C., Koop, D., Swanson, A., & Young, N. E. (2013). VisTrails SAHM: Visualization and workflow management for species habitat modeling. *Ecography*, *36*, 129–135. https://doi.org/10.1111/j.1600-0587.2012.07815.x

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*(3–4), 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Roberts, P. R., & Oosting, H. J. (1958). Responses of Venus fly trap *(Dionaea muscipula*) to factors involved in its endemism. *Ecological Monographs*, *28*(2), 193–218. https://doi.org/10.2307/1942208

Salas, E. A. L., Valdez, R., & Michel, S. (2017). Summer and winter habitat suitability of Marco Polo argali in southeastern Tajikistan: A modeling approach. *Science Direct*, *3*(11), Article e00445. https://doi.org/10.1016/j.heliyon.2017.e00445

Schulze, W., Schulze, E. D., Schulze, I., & Oren, R. (2001). Quantification of insect nitrogen utilization by the venus fly trap *Dionaea muscipula* catching prey with highly variable isotope signatures. *Journal of Experimental Botany*, *52*(358), 1041–1049. https://doi.org/10.1093/jexbot/52.358.1041

Talbert, C., & Talbert, M. (2012). User Documentation for the Software for Assisted Habitat Modeling (SAHM) package in VisTrails. USGS.

U.S. Geological Survey. (n.d.) *Landsat Surface Reflectance-derived Spectral Indices.* USGS. https://www.usgs.gov/landsat-missions/landsat-surface-reflectance-derived-spectral-indices

Waller, D. M., Johnson, Y. B., Evans, R., Gibson, T., & Morris, W. (2016). Petition to list the Venus flytrap (*Dionaea muscipula* Ellis) as Endangered under the 1973 Endangered Species Act. U.S. Fish & Wildlife Service. https://doi.org/10.13140/RG.2.2.24190.59204

Wan, Z., Hook, S., & Hulley, G. (2015). MOD11A1 MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC, accessed 30 September 2021. https://doi.org/10.5067/MODIS/MOD11A1.006