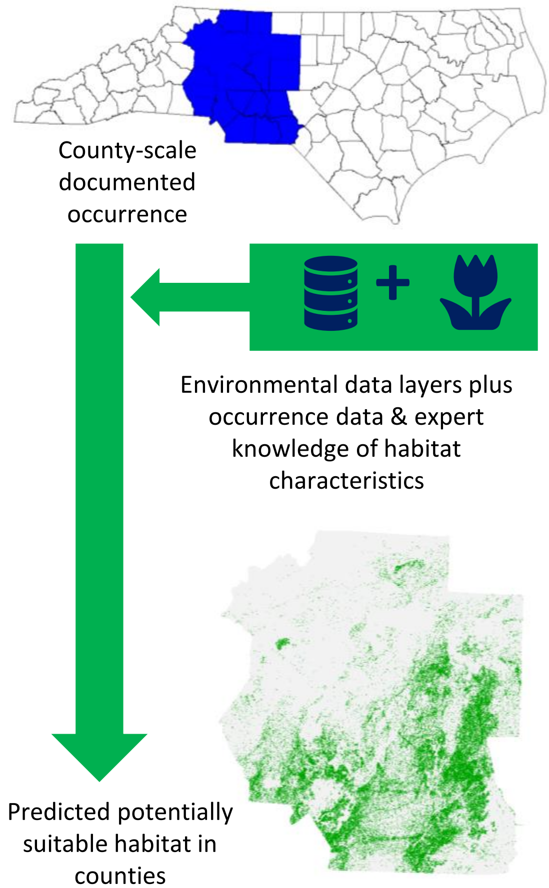
NCDOT ATLAS Project

Species' Models Support
Process Improvements
April 2018



The species' team identified models as useful for process improvement. Planners and biologists currently reference county occurrence lists to anticipate potential effects. **Models (1) facilitate finer-scale risk assessment and resource needs during project scoping/pre-screening, and (2) summarize multivariate, complex data to inform field and desktop site evaluations**. To identify modeling candidates, the team referenced occurrence data from the Natural Heritage Program and the NCDOT STIP layer to prioritize species with expected conflicts and reviewed current NCDOT/agency procedures to deprioritize species with established protocols or procedures (e.g., several marine species). For species for which models were deemed most valuable, the quantity and quality of available presence data determined if we pursued machine-learning (data-rich) or expert-based (data-poor) model methods.

Above: Diagram of model start and end point. Map at bottom illustrates a binary potential habitat map based on chosen threshold (e.g., require potential habitat to include 98% of known occurrences). Below: Diagram of our assessment of species' model needs and recommended approaches.

Federally listed species have been documented in the county

Models
if knowledge of higher resolution than county list would be valuable

No Models
if list is adequate given established protocols or agreements

- *Rusty patch bumble bee*
- *Red wolf*
- *Virginia big-eared bat*[1]
- *Gray bat*[1]
- *Indiana bat*[1]
- *American alligator*
- *Bog turtle*

[1]Process improvements for bats include models, but not state-wide habitat models

Expert-based Models
if data-poor, simple criteria, or existing model

- Small-whorled pogonia
- Spreading avens
- Blue Ridge goldenrod
- Carolina Northern flying squirrel
- Loggerhead turtle
- Piping plover
- St. Francis satyr butterfly
- …and more!

TOTAL: 13 plants, 4 aquatic animals, 17 terrestrial animals

Machine-learning Models
if adequate occurrence data, complex habitat criteria, or existing model

- Schweinitz's sunflower
- Dwarf-flower heartleaf
- Michaux's sumac
- Rough-leaf loosestrife
- Swamp pink
- Virginia spirea
- White irisette
- Smooth coneflower

TOTAL: 15 plants, 16 aquatics animals[2], and 2 terrestrial animals

[2]The USFWS has shared Maxent models for most listed fish and mussel

The species' team has reviewed data, protocols, and potential STIP effects for 76 federally listed species. All species have an assigned review or model strategy and lead team member. The Schweinitz's sunflower and Dwarf-flower heartleaf serve as pilot species to refine machine-learning model methods as we choose between Maxent and Random Forest algorithms. Results from both machine learning methods are in review via a customized Arc GIS Online tool created for this project.

Expert-based Procedures

Build in ArcGIS Model Builder • Start with expert's knowledge, literature review, and Element Occurrence data from NHP • Review aerial imagery, quad maps, and environmental data in ArcGIS **• Identify key environmental attributes of species and use environmental data to define thresholds and ranges for suitable vs unsuitable habitat** • Use GIS tools to apply these rules over species' range to identify areas as suitable or not suitable • Output binary (Suitable, Unsuitable) or categorical (High, Medium, Low) habitat classification

Shared Procedures

Design and document to meet relevant standards and facilitate updating if new data become available • Evaluate predictions with independent data (DOT site survey data) • Evaluate predictions with expert-knowledge (AGOL)

Shared Data

NHP Element Occurrences and a collection of 60+ environmental layers are the primary data resources for the models. Machine learning begins with all environmental layers and retain those most valuable for discerning patterns. Experts *a priori* select layers identified as best proxies for conditions observed in the field.

Machine-learning Procedures

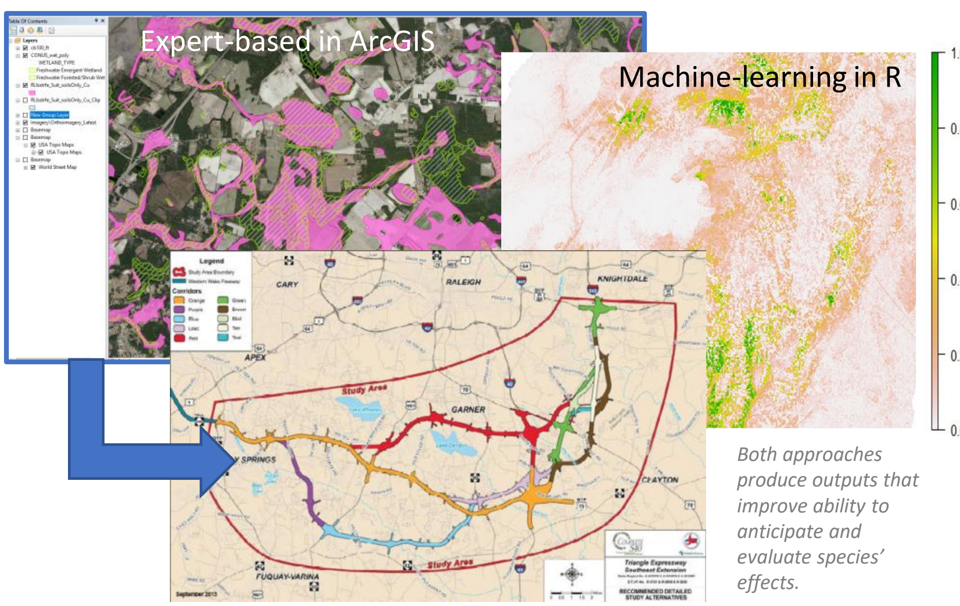
Build in R programming language • Start with Element Occurrence data from NHP (presences) • Draw random points from counties with presences (pseudoabsences) • Attribute presence and absence points with environmental data **• Construct many regression tree models using different splits of the data (Random Forest) to find best solution to accurately classify known presence** • Evaluate model statistically (cross-validation) • Output continuous probability of suitability habitat or convert to binary or categorical based on thresholds

Species are listed to occur in county, but…

What are relative expected effects?

Where are effects most likely?

What species are likely affected?



Both approaches produce outputs that improve ability to anticipate and evaluate species' effects.

Both model approaches draw from the same pool of GIS data. The species' team had collaborated with the GIS team to acquire, clean, document, and standardize over 60 environmental data layers to date. These statewide data provide measures of reflectance, landform, vegetation cover and structure, soil characteristics, disturbance indicators, and climate. Expert-based models primarily select variables with clear ecological relevance and close association to site characteristics measured in the field. Machine-learning models perform better with continuous data and can leverage complex environmental relationships that may not be evident from field surveys or represented via 'buffer and clip' style GIS commands.