



# Creating Predictive Models to Streamline Inventory: Lessons from the Cumberland Plateau

Regan Lyons

April 22, 2009



**NatureServe**

A Network Connecting Science With Conservation

A vertical strip on the left side of the slide shows a lush forest floor with various green ferns and tree trunks in the background.

# Objective

---

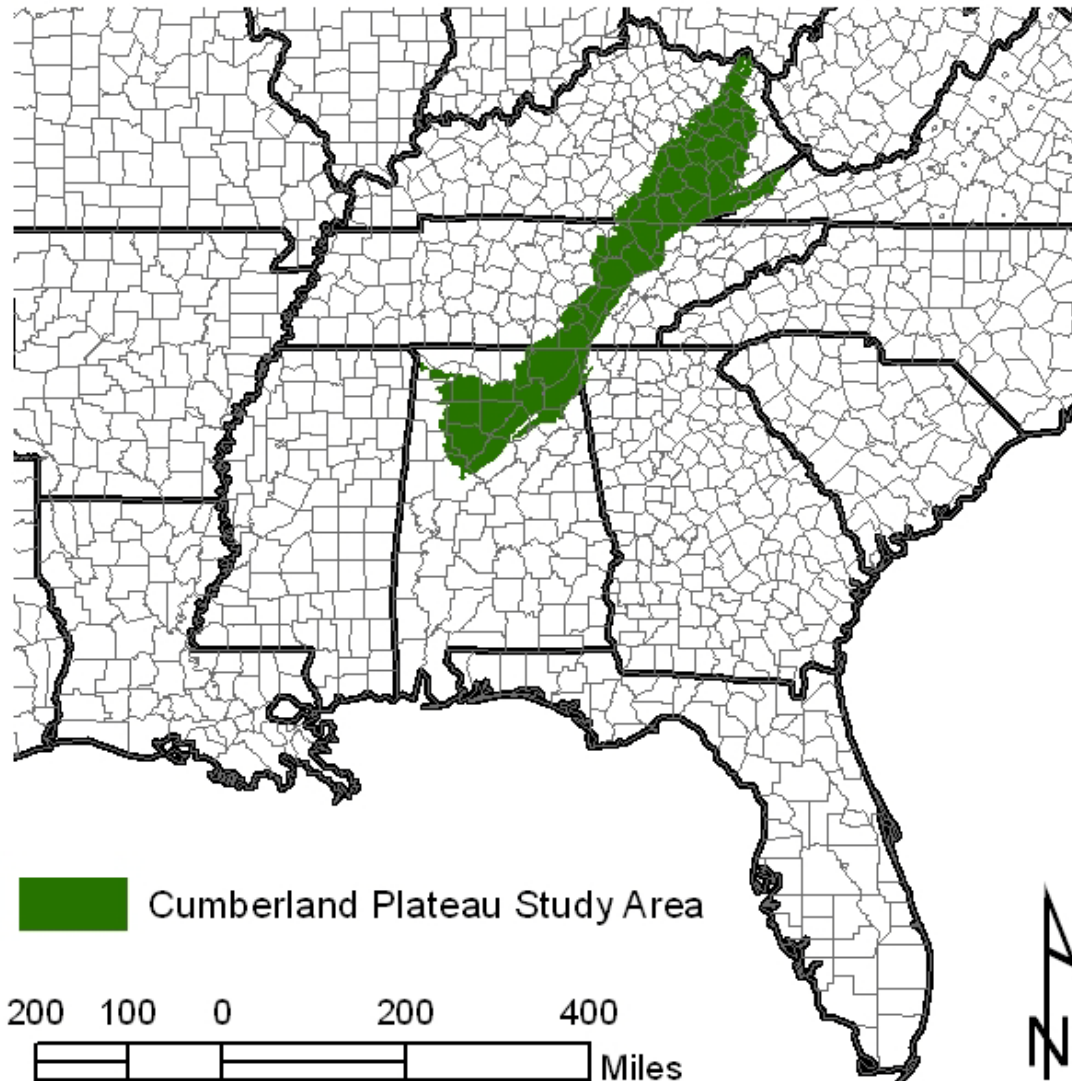
Use existing data to create predictive models to help locate new occurrences of rare communities on the Cumberland Plateau

Funding: Office Depot

Partners: KY, TN, & AL NHPs

# Study Area

---



A vertical strip on the left side of the slide shows a lush forest floor with various green ferns and tree trunks in the background.

# Process

---

Step One: What communities work best?

Tried: All rare (G1-G3)  
communities occurring w/in  
forests

Step Two: refined models for best  
candidates

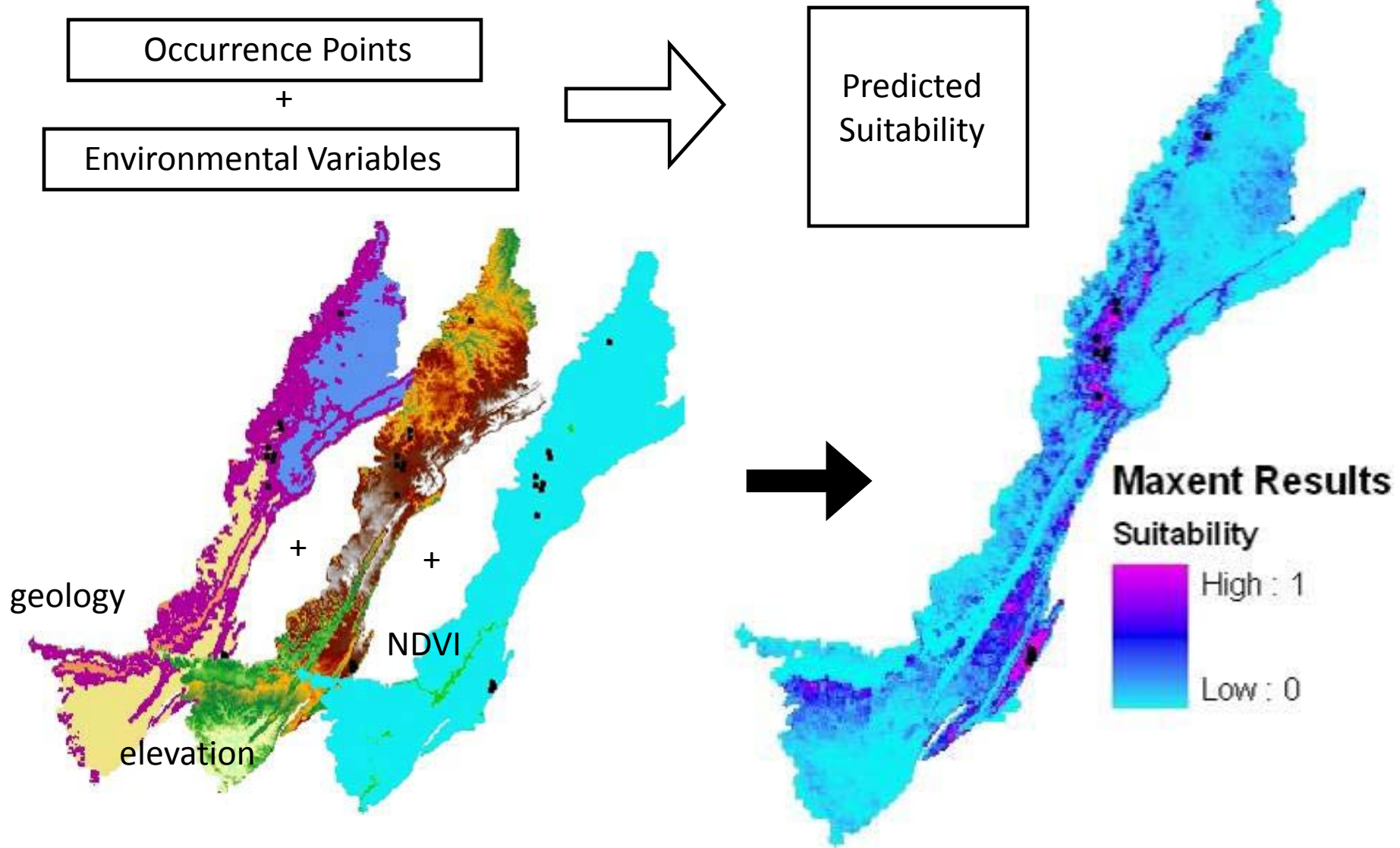


# Maximum Entropy Modeling

---

- Maximum Entropy – “MaxEnt” Modeling
  - (Phillips et al. 2006)
- Performs as well or better than other approaches (e.g. CART, GLM)
  - (Elith et al. 2007)
- Only requires presence data
- Seems to perform relatively well with small sample sizes
- Estimates an unknown distribution using environmental predictor variables as constraints

# Maxent



A vertical photograph on the left side of the slide shows a forest floor covered in green ferns and other vegetation. The background is slightly blurred, showing more of the forest.

# Occurrence Data

---

Project emphasized use of existing, easily obtained data:

State community EOs, supplemented by

- Classified plot data from National Parks
- Classified plot data from Bankhead NF
- Some species EO

-e.g. *Platanthera integrilabia*



# Environmental Variables

---

Models were based on topographic, geologic, and remotely sensed variables

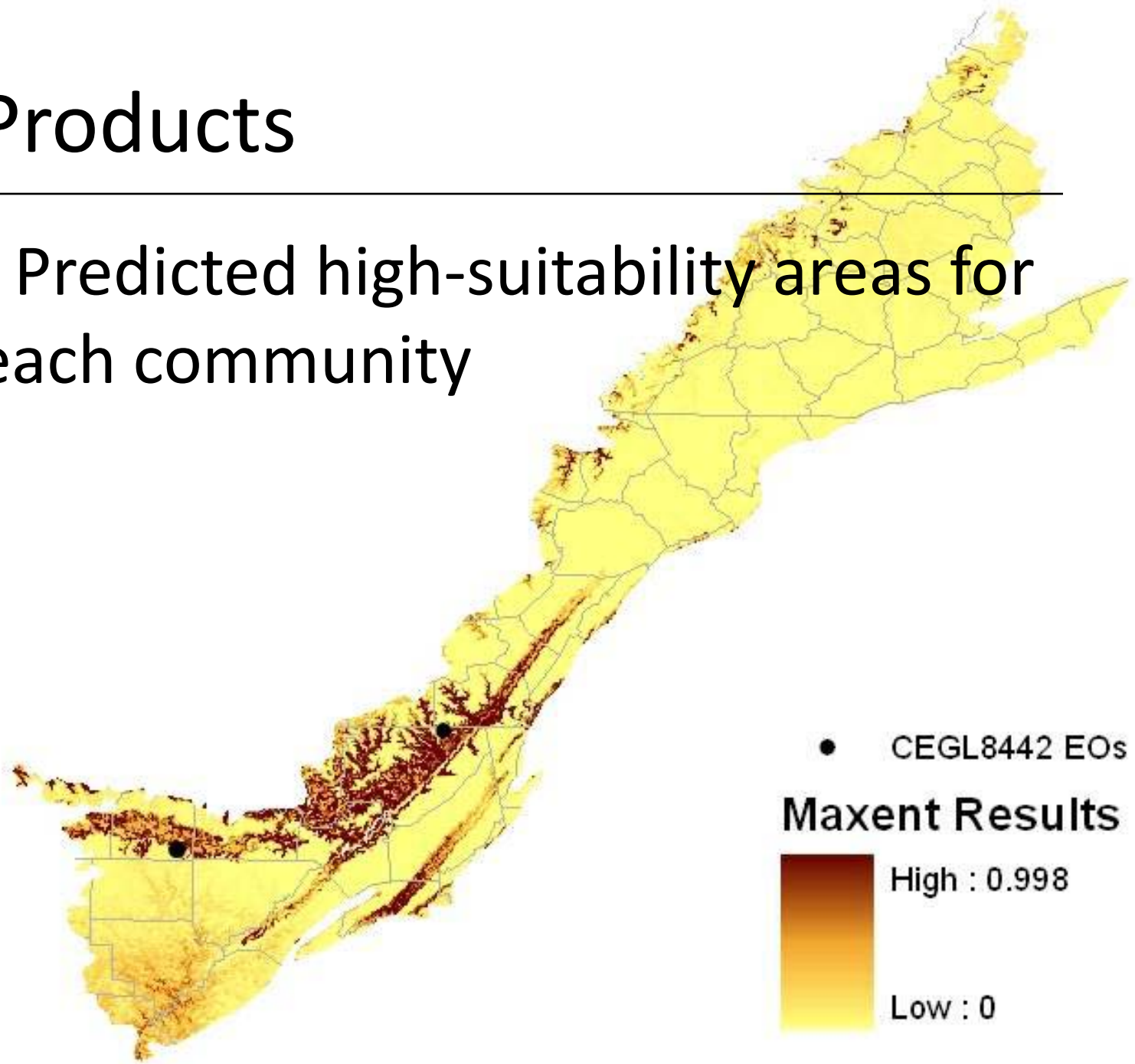
- Elevation
- Slope
- Aspect
- Curvature
- Geology – broad classes
- 2001 Landsat Imagery, multiple seasons
  - Normalized Difference Vegetation Index
  - Tassel-cap Transformed Landsat Imagery (wetness, greenness, brightness)

Data resolution = 30 meters

# Products

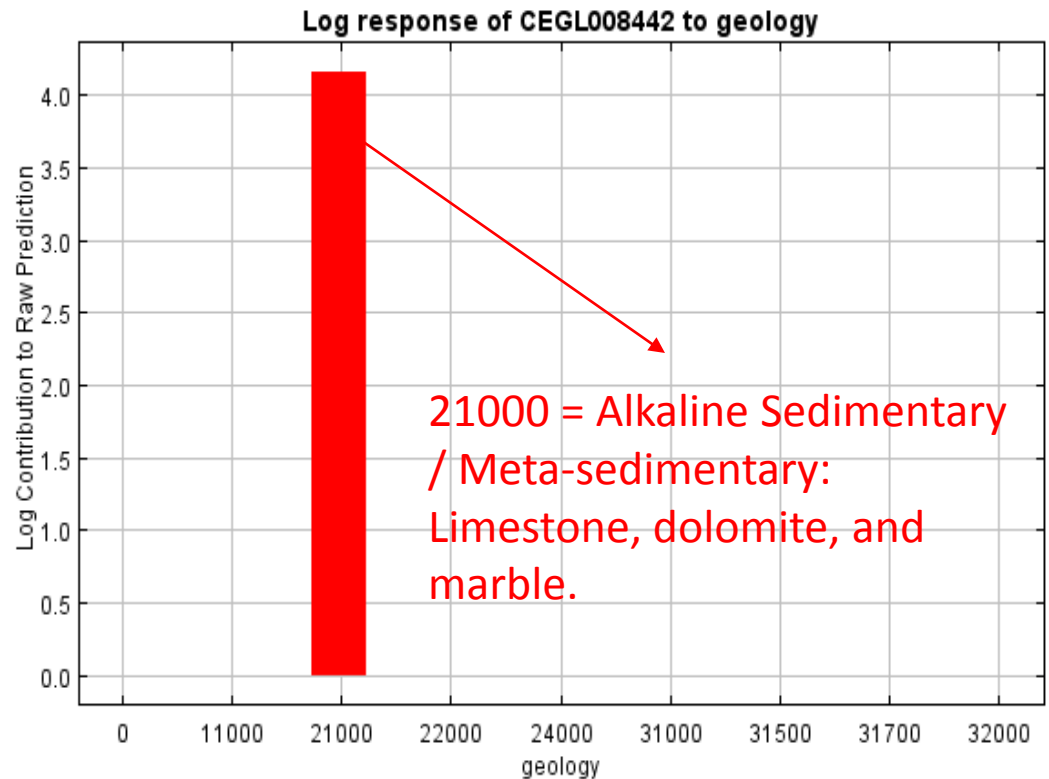
---

- Predicted high-suitability areas for each community



# Products

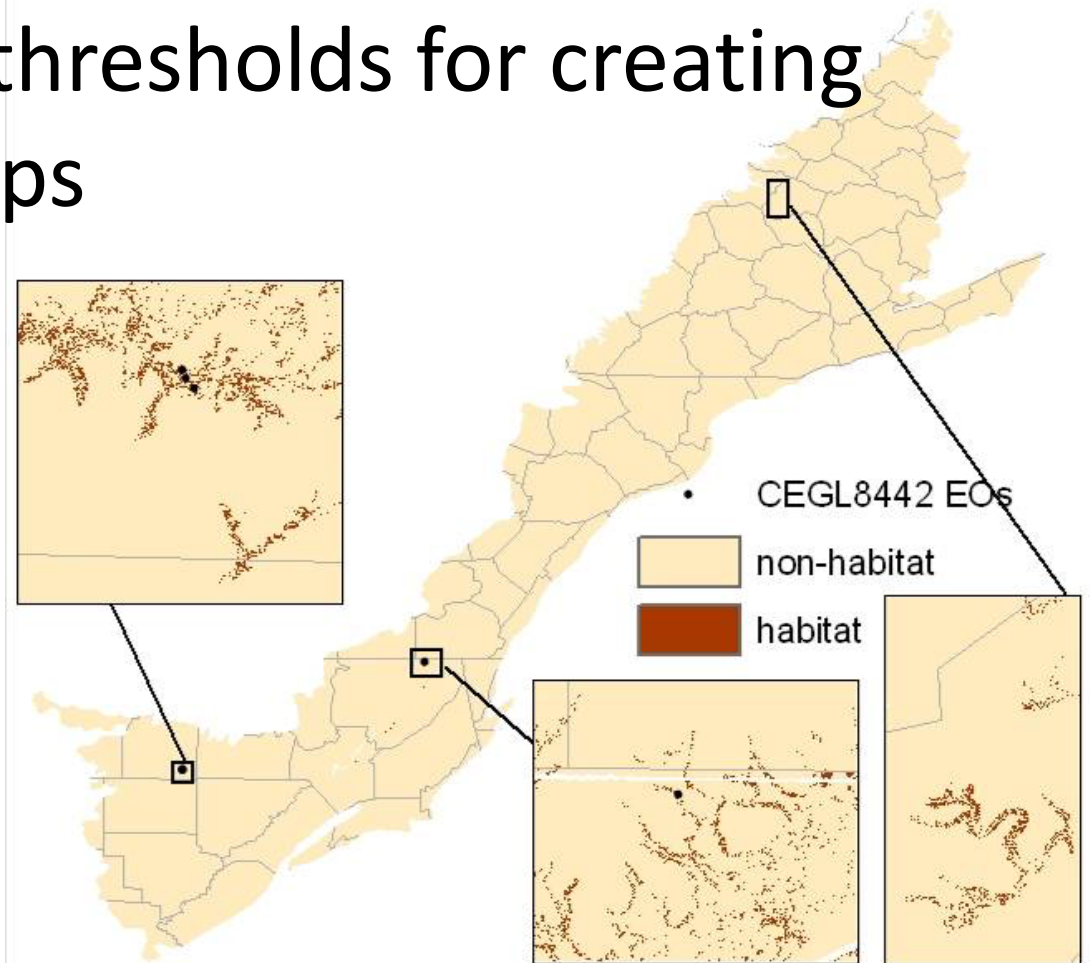
- Graphical displays of variable contributions



# Products

---

-Measures of model success,  
including thresholds for creating  
binary maps



A vertical strip on the left side of the slide shows a lush forest floor with various green ferns and tree trunks in the background.

# Results

---

Step One: What communities work best?

- Multiple occurrence records
- Limited selection bias
- Environmental affinities suitable for modeling at 30m-resolution

3 Target Communities for refined modeling

A vertical strip on the left side of the slide shows a lush forest floor with green ferns and tree trunks in the background.

# Things to Watch For

---

- Spatial accuracy of the occurrences
- Spatial accuracy of explanatory variables
- Scale of response to explanatory variables
- Categorical data: the value of lumping
- Temporal accuracy: old EOs
- Temporal accuracy: old explanatory data
- Biased EO locations



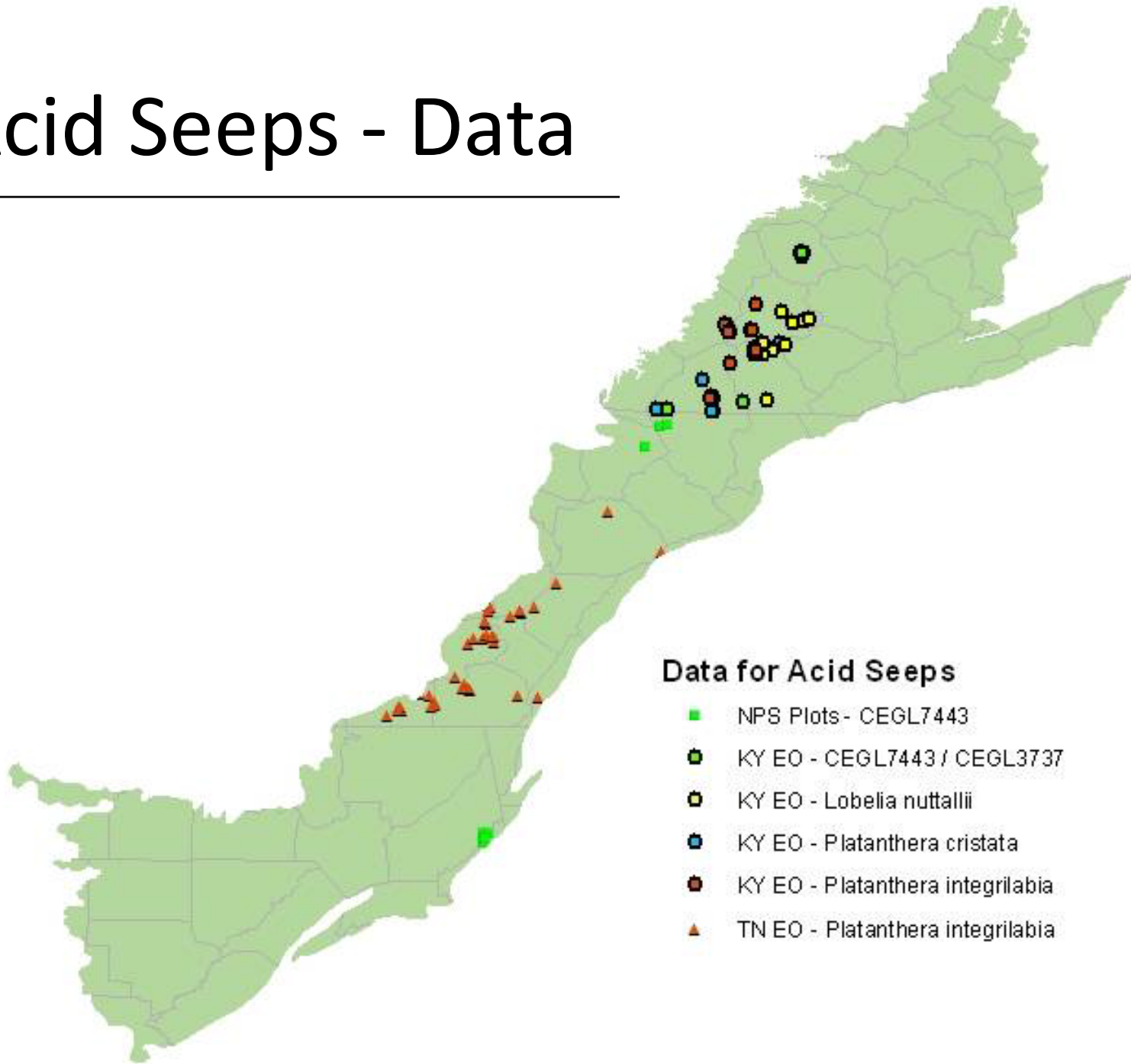
Step 2: Refined Models

THE SEARCH

FOR ACID SEEPS

# Acid Seeps - Data

---



A vertical photograph on the left side of the slide shows a forest floor covered in green ferns and other vegetation. The background is slightly blurred, showing more of the forest.

# Acid Seeps - Results

---

## Most Important Predictive Variables:

- Elevation (300 – 600 meters)
- Slope (<5 degrees)
- Brightness (winter)
- Greenness (summer)
- Curvature
- Wetness (winter)

# Acid Seeps - Results

---

Most Important Predictive Variables:

Elevation

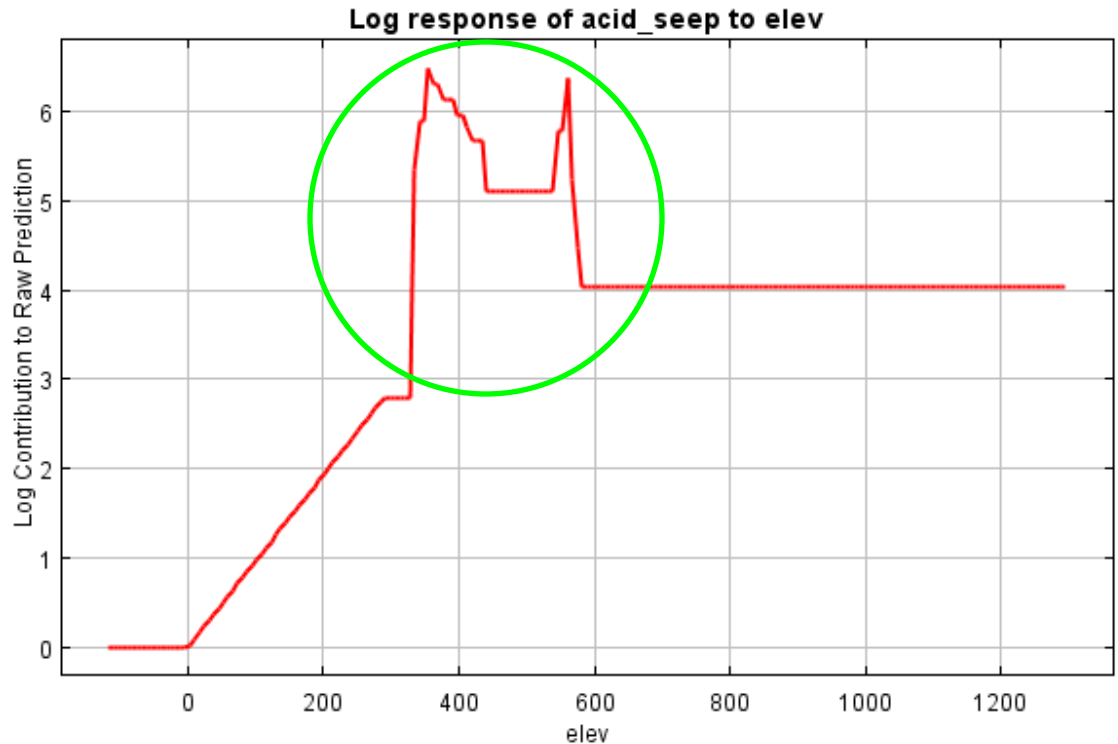
Slope

Brightness

Greenness

Curvature

Wetness



# Acid Seeps - Results

---

Most Important Predictive Variables:

Elevation

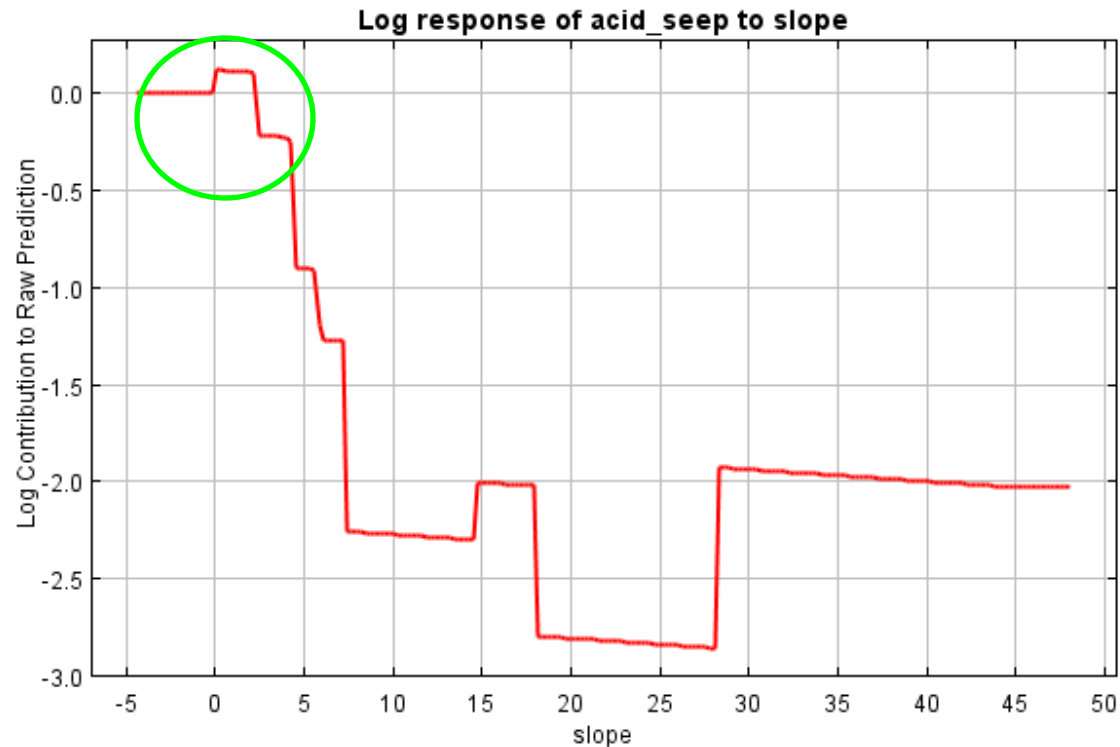
**Slope**

Brightness

Greenness

Curvature

Wetness



# Acid Seeps - Results

---

Most Important Predictive Variables:

Elevation

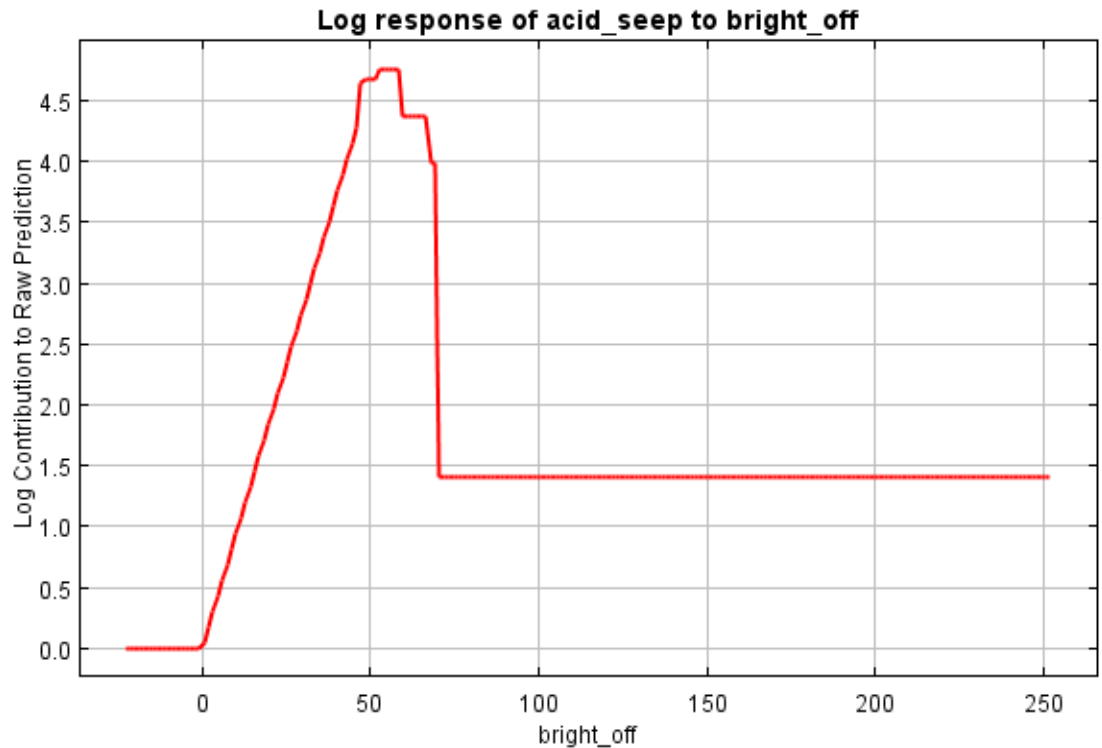
Slope

**Brightness**

Greenness

Curvature

Wetness



# Acid Seeps - Results

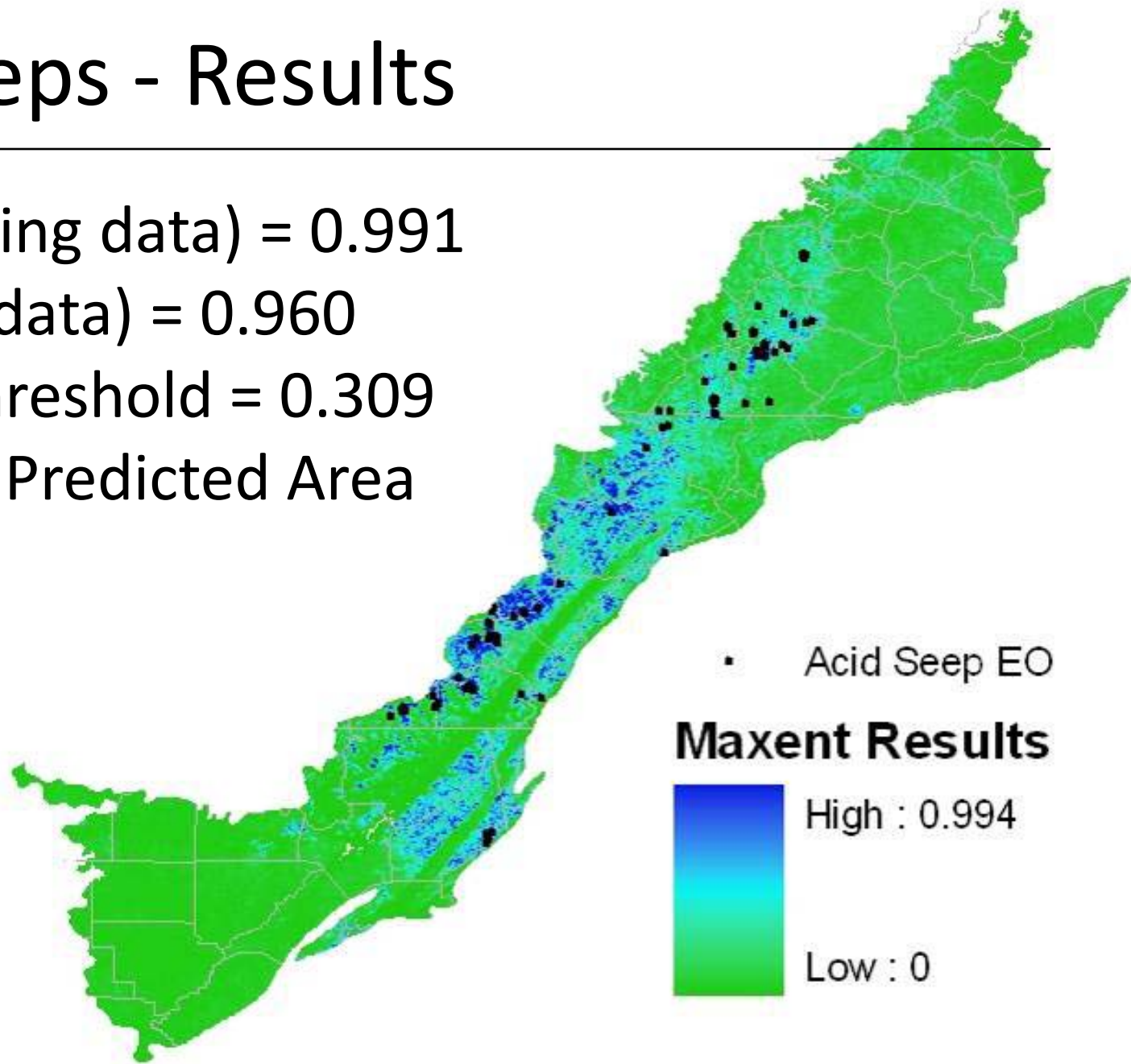
---

AUC (training data) = 0.991

AUC (test data) = 0.960

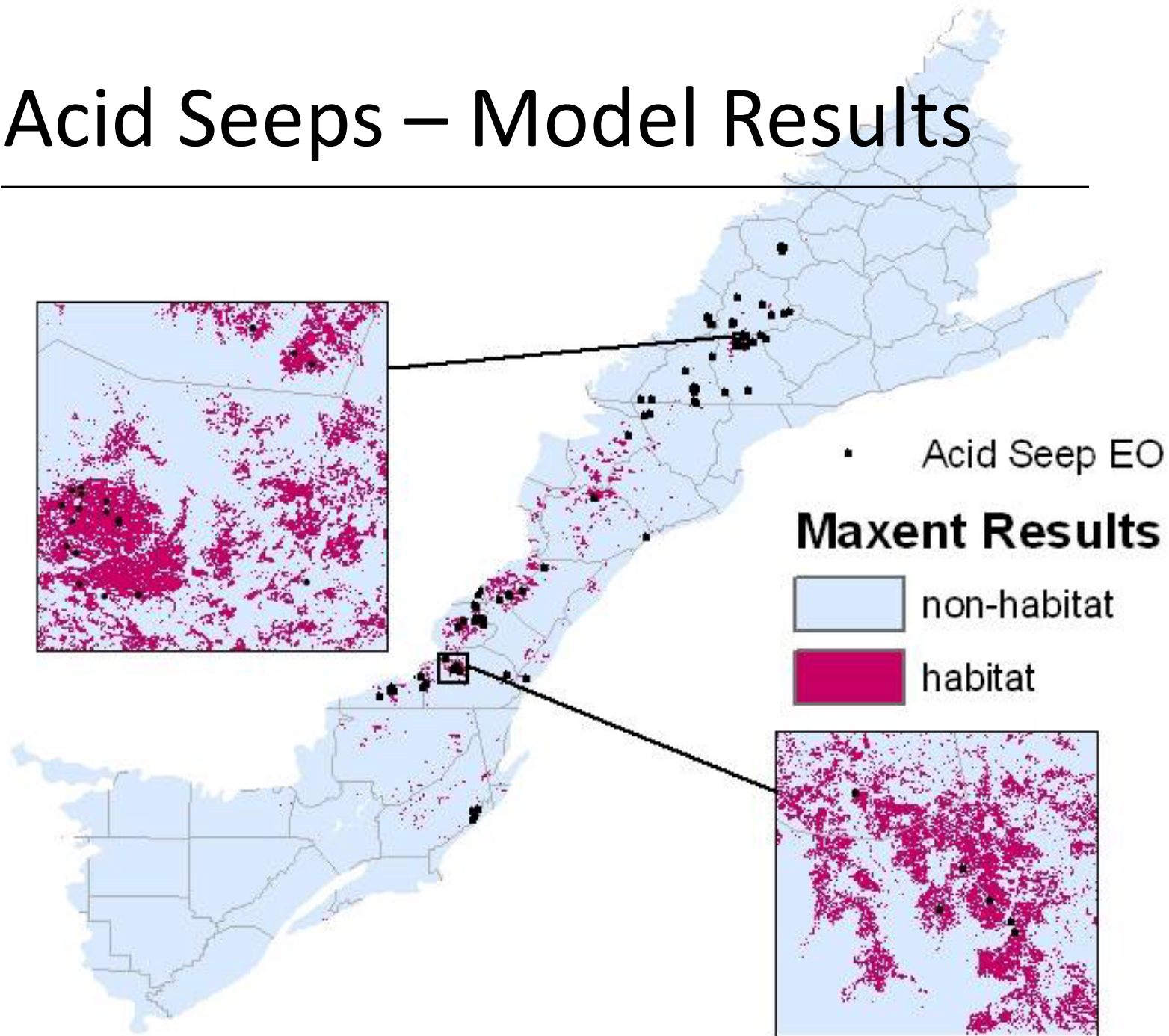
Logistic Threshold = 0.309


Fractional Predicted Area  
= 5.3%



# Acid Seeps – Model Results

---





# Acid Seeps - Conclusions

---

Seep model over-predicts, could be improved by:

- Remodel: Smaller areas, finer resolution
- Post Modeling:
  - Smaller areas, identify stream heads
  - Add data on canopy cover and/or powerline corridors

Despite errors of commission, effective for directing field work

# Acid Seeps – Field Inventory

---





# General Conclusions

---

Models are only as good as the data that goes in... but even imperfect data can help guide inventory

- For rare communities, errors of commission common

- Models most useful when paired with finer-scale analysis

- Particularly useful in unfamiliar areas