Classifying Marshland Plant Species by Processing Light Reflectance in Satellite Images

Aidan Draper[1, ‡], Luke Onken[2], Megan Vahsen[2], Dr. Jason McLachlan[2]
[1]Elon University, [2]University of Notre Dame, [‡]Correspondence: adraper2@elon.edu

Background

The Random Forest classification algorithm has become very popular in the remote sensing community for its ability to accurately identify such things as forests, bodies of water, urbanized areas, and agricultural plots in satellite images because these surfaces have greatly differing light reflectance bands. However, the remote sensing community has not used the algorithm as commonly for classifying surfaces that slightly differ in their light reflectance bands, like different forms of vegetation.

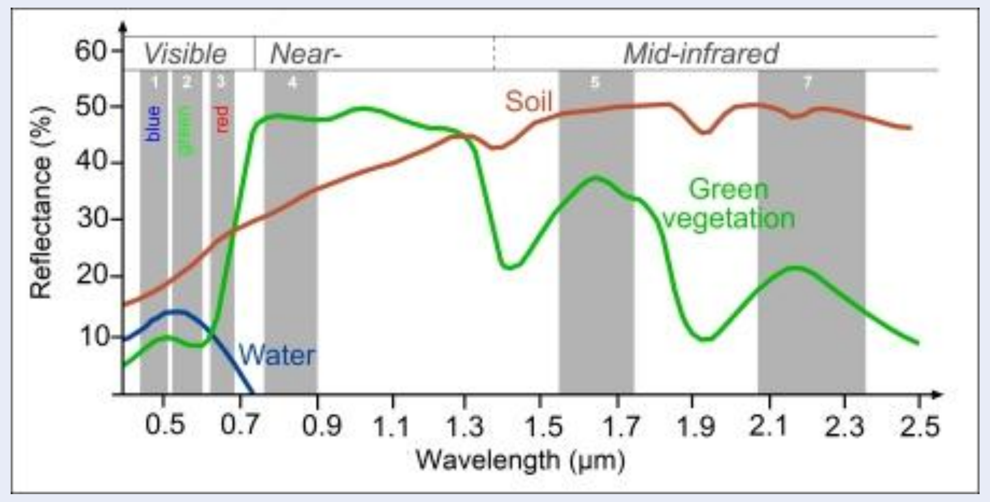

Figure 1: *A visual representation of light wavelengths captured by satellites.* Sourced from http://www.seos-project.eu/modules/classification/classification-c01-p05.html_.

The Smithsonian Environmental Research Center and the McLachlan Lab at Notre Dame are interested in the changing abundances of salt marsh plant species within a research wetland that has been studied for 30 years. A remote sensing reconstruction of such changes in populations would complement controlled experimental work in this area.

Figure 2: *Image of the research center's marshland.* Sourced from https://serc.si.edu/research/research-topics/global-change.

Importance:

1) By studying past marsh plant populations, **other scientists can model how future marshland ecosystems may change** due to shifts in certain marsh plant species abundance and changes in environmental factors, like CO2.
2) The success of this analysis would mean that, in some cases, **Random Forest classifiers can be used for differentiating low-spectrum light reflectance band differences** in objects observed from space.
3) **Analyses of population change through time would complement the previous experimental work** conducted by the Smithsonian Environmental Research Center and the Mclachlan lab.

Data Processing

The initial dataset included **514 1x1 meter plots, which were mapped every 20 meters** within the research center's marshland. Each plot had visually assigned **ordinal scores for 6 species**, including Phragmites and *Schoenoplectus americanus*.

1) Plots were **assigned Landsat OLI resolution spaces** based on their UTM coordinate overlap.

2) **Calculated covariates,** such as

- Normalized Difference Vegetation Index (NDVI)
- Soil Adjusted Vegetation Index (SAVI)
- Normalized Difference Water Index (NDWI)

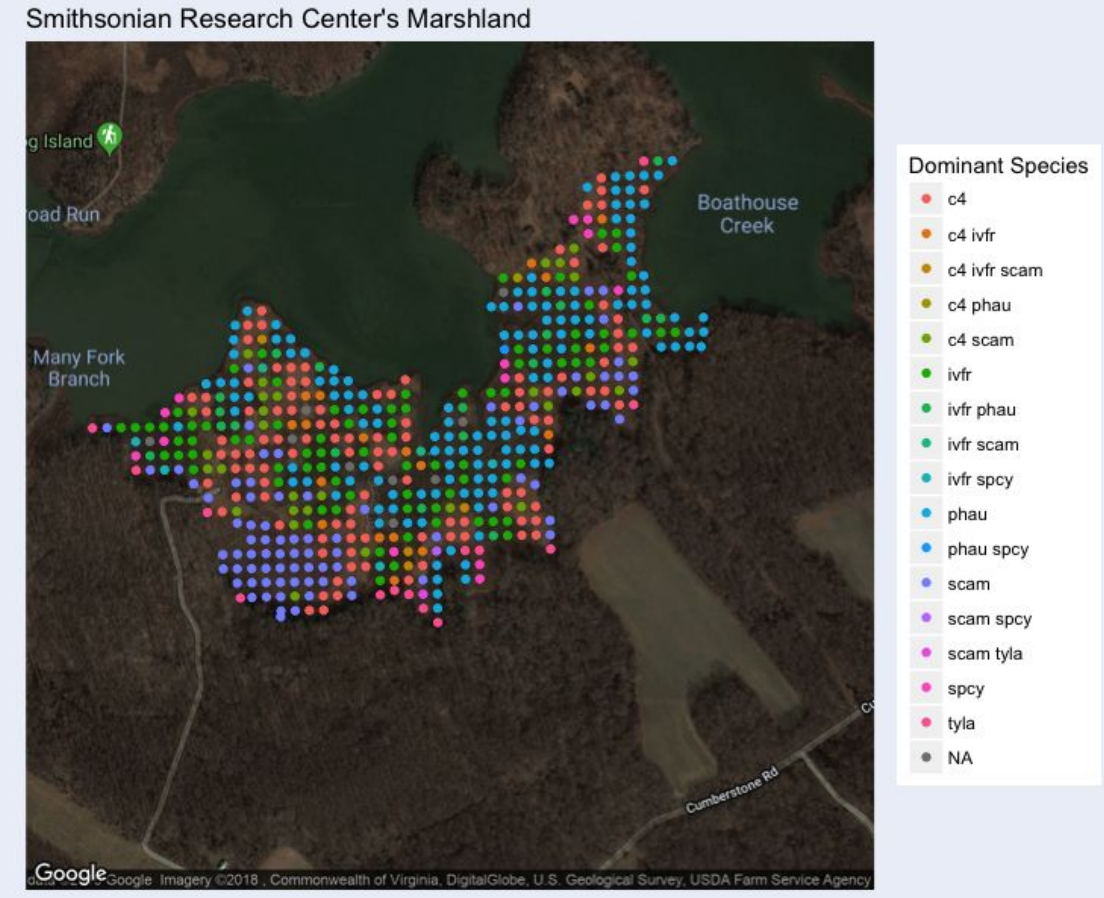

Figure 3: *The research center's most dominant species per 1x1 meter plot overlaid on a Google satellite image.*

Random Forest Model

Algorithm: model = randomForest(plant_species ~ NDVI, SAVI, NDWI, band2 (blue), band3 (green), band4 (red), band5 (NIR), data = training_set, keep.forest = TRUE)

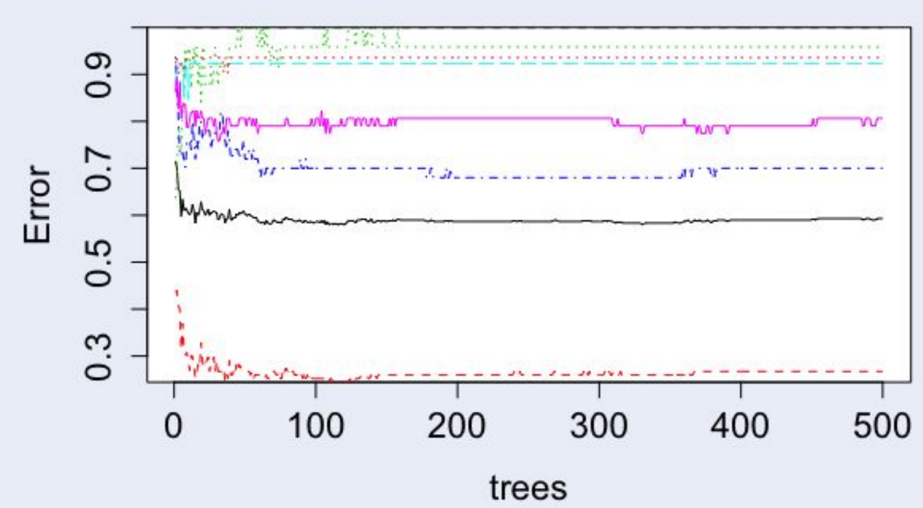

Figure 4: *The compounded error per ordinal class as every additional tree was added to the Phragmites model.*

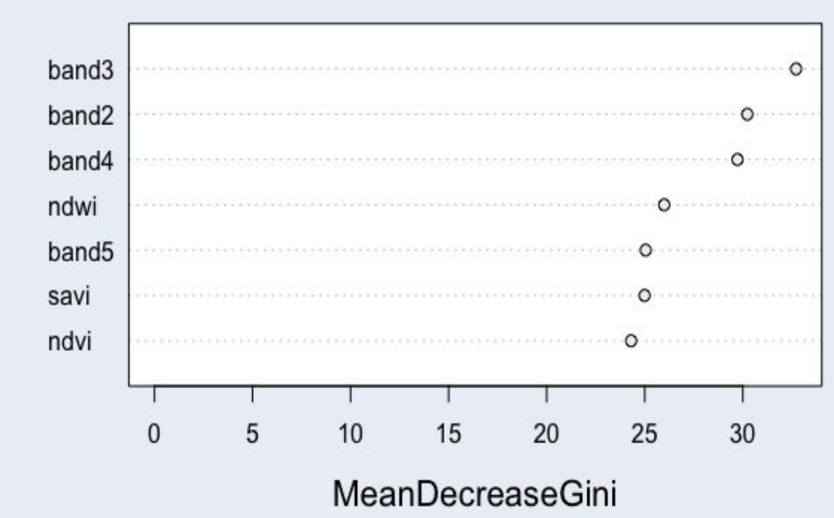

Figure 5: *The variable importance for every covariate in the Phragmites random forest model.*

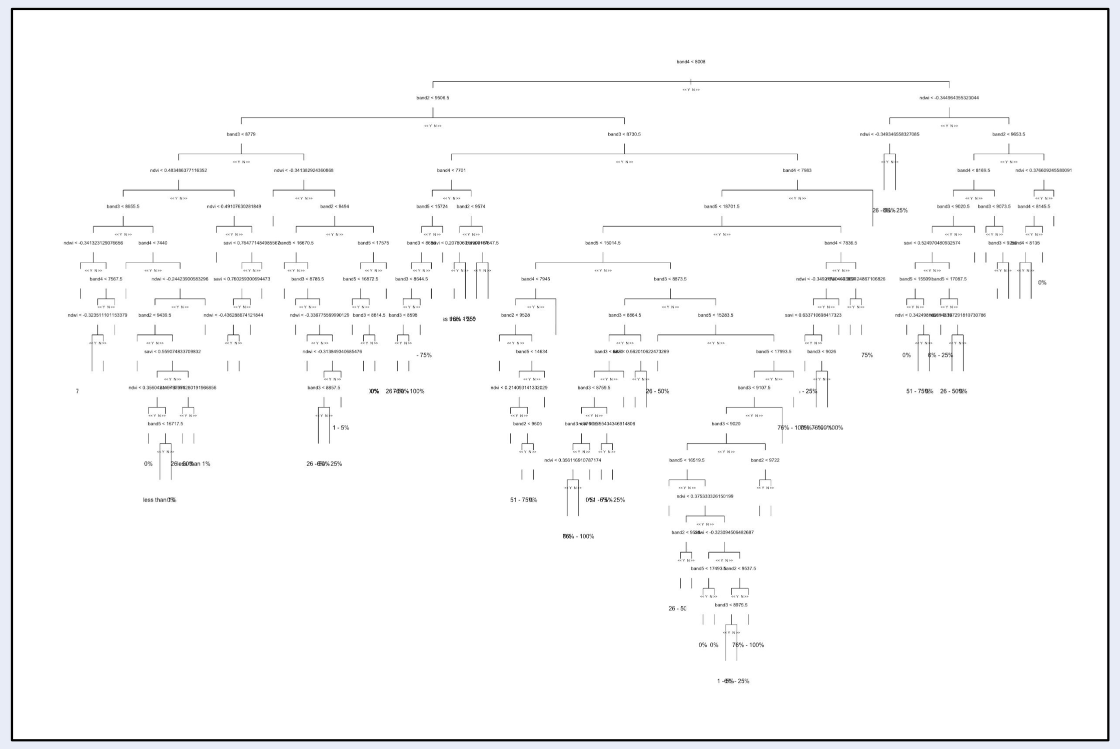

Figure 6: *One of 500 decision trees produced in the Phragmites Random Forest model.* This diagram shows the complexity of distinguishing low-spectrum differences within light bands.

- Created **500 decision trees** and averaged them together to prevent overfitting.
- **Trained on 60%** of the Landsat OLI resolution spaces for **August, 2016**.
- **Dependent on** calculated covariates **NDWI**, **NDVI**, and **SAVI**, as well as individual **light bands 2-5**.
- **Class errors were fairly high except for 0% cover** (the red line in Figure 4), which was around 15% to 20 % error.

Classifier Results

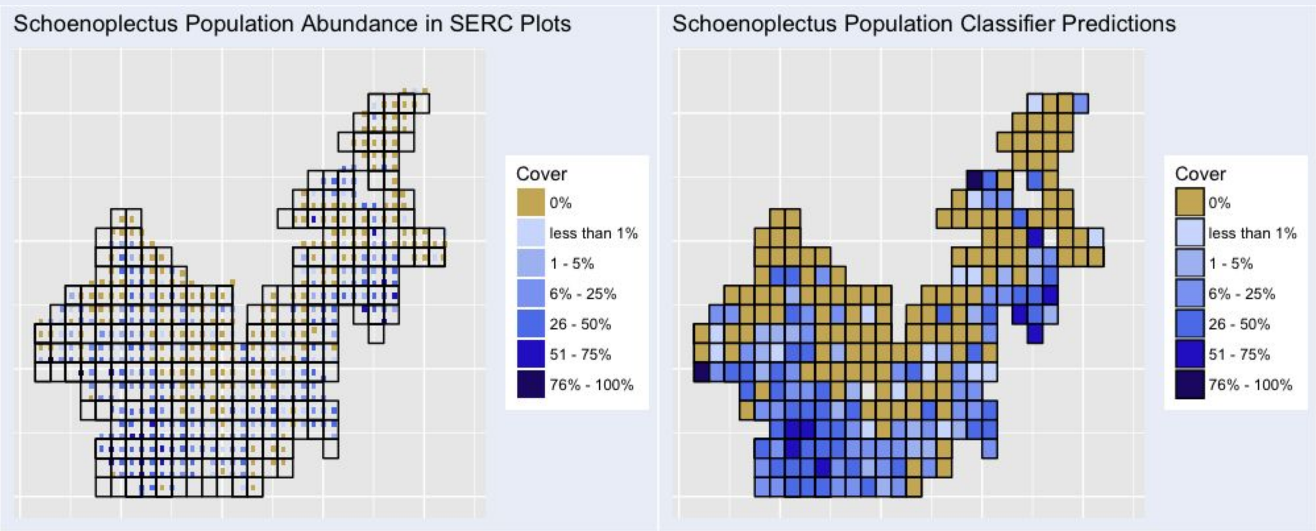

Figure 7: *A comparison of the ordinal data and the classifier's full prediction for Schoenoplectus americanus in the research center's mashland.*

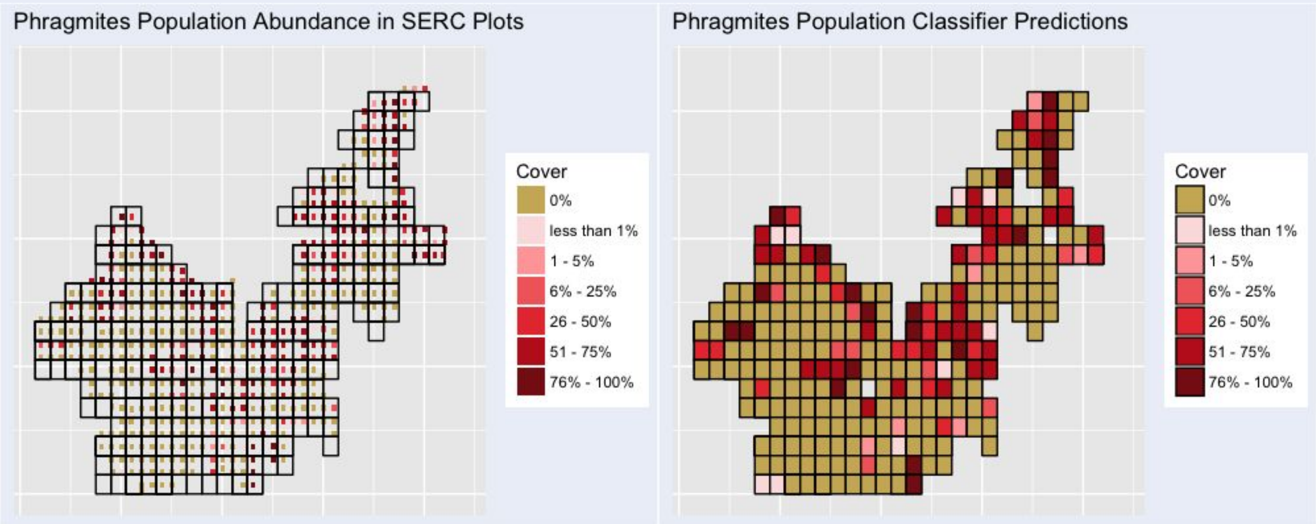

Figure 8: *A comparison of the ordinal data and the classifier's full prediction for Phragmites in the research center's mashland.*

The overall species models had **low out-of-bag errors of around 45 to 60% accuracy.** However, this was due to uncertainty when predicting ordinal scores. The classifiers can accurately predict Phragmites presence/absence in a pixel with **85% accuracy.** The classifiers also predict *Schoenoplectus americanus* with **75% accuracy.**

Future Work

The model will be tested on previous years of Landsat image data. If accurate, running the model on the 40+ years of available Landsat images would allow scientists to study how these populations have shifted over time.

If these population changes correlate with changes in environmental factors or the rise in Phragmites populations, the research center has additional support for their work in a physical setting.

If environmental factors and the model's population predictions do not correlate, the model would fail to support the controlled work that the McLachlan Lab and the research center have completed, which may add some level of uncertainty to the results of the controlled tests.

Acknowledgments

We thank the Smithsonian Environmental Research Center for letting us use their dataset. This work was supported by National Science Foundation grant IIS-1560363: "REU Site: Data Intensive Scientific Computing".