Analysis of Eucalyptus Growth in the San Francisco Bay Area, CA Using Machine Learning Algorithms

Jason Andrews, Rory Dickinson, Tanner Honnef, Caleb Kluchman

Introduction

Eucalyptus globulus, or Tasmanian blue gum eucalyptus, is a large, fast-growing tree that has been planted in many Mediterranean climate biomes around the world, including Northern California. One simple question that has not been answered is whether the tree is continuing to spread and colonize new environments, or if the extent of existing stands is static. Eucalyptus was typically planted near towns as a windbreak or to stabilize loose or sandy soil.

Research Objectives

The goal of this study is to:

1) research how machine learning methods can be used to analyze the extent of eucalyptus based on spectral and other raster data inputs.

2) understand which factors have the greatest importance in predicting eucalyptus distribution

Study Area

This study area consists of 8 counties in the San Francisco Bay Area: Alameda, Contra Costa, Marin, Napa, San Mateo, Santa Clara, Santa Cruz, and Sonoma counties. San Francisco County is not included in the analysis because of the lack of eucalyptus.

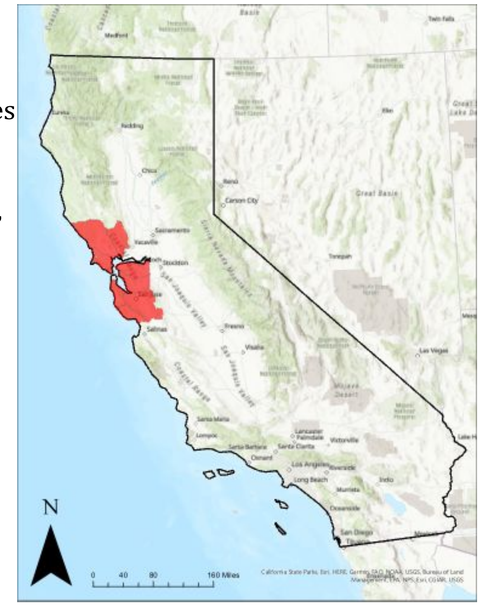

Figure 1: Outlines Study Area

Data

The variables used in this analysis include distance from built, precipitation, Soil Adjusted Vegetation Index, Land Cover, all Sentinel 2 bands, and Elevation. These were obtained from SRTM, Pacific Veg Map, Living Atlas, Sentinel 2, and UCSB CHIRPS.

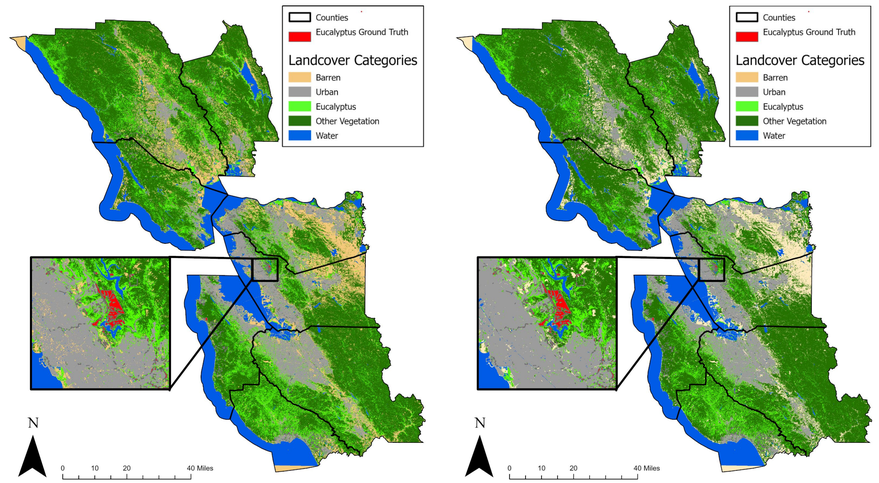

Figure 2: Random Forest Land Cover Classification with 10 trees (left) and with 100 trees (right)

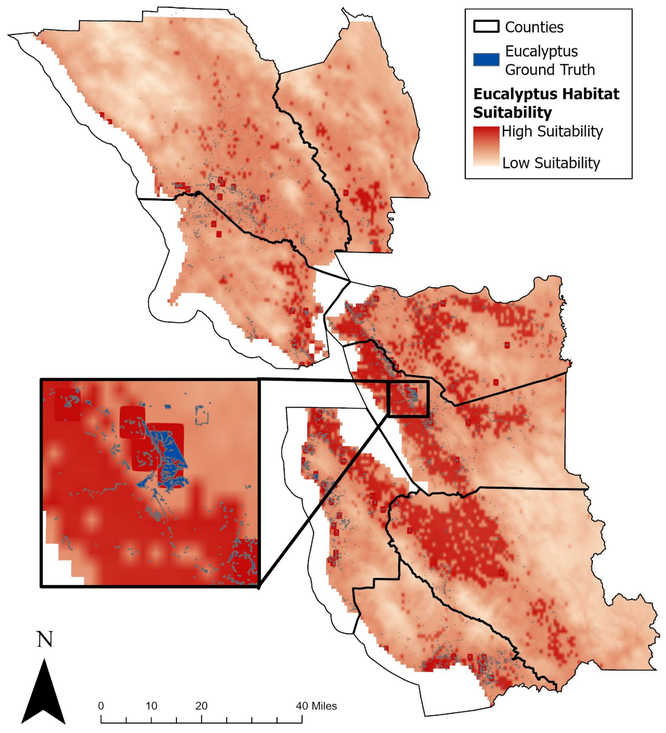

Figure 3: MaxEnt Eucalyptus Suitability Map

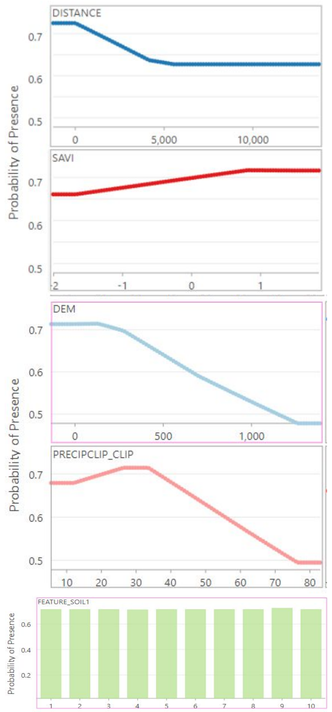

Figure 4: Responses of variables used in MaxEnt. From top to bottom: Distance to Built, SAVI, Elevation, Precipitation, Soil Type

Variable	Random Forest	MaxEnt
Elevation	X	X
Distance to Built	X	X
Spectral Bands	X	
SAVI		X
Soil Type		X
Precipitation	X	X
Land Cover		X

Table 1: Variables Used in Analysis

Random Forest Results

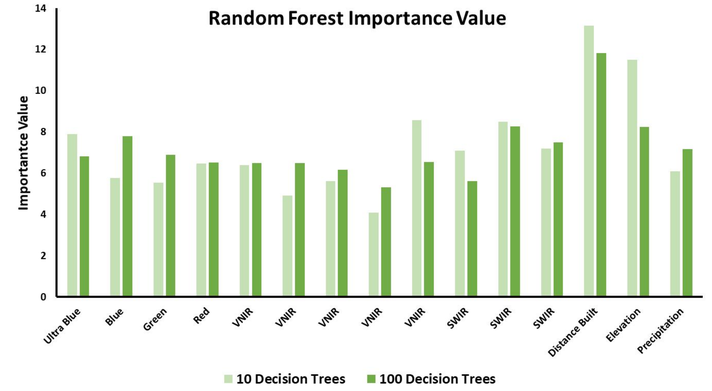

Figure 5: Random Forest Importance Values by Band

Discussion

The MaxEnt result is a suitability map showing where eucalyptus could grow if it was planted, according to where it has grown in our training data. The Random Forest result shows where the model thinks that Eucalyptus exists within the study area.

The two most responsive variables in the analysis in MaxEnt are elevation and distance from built. SAVI and soil type did not have a large impact on MaxEnt. Random Forest labeled distance to built and elevation with the highest importance. The Random Forest with 100 decision trees had a 2% higher validation accuracy than the run with 10 decision trees. Model tuning could increase the validation accuracy particularly setting prior probability parameters. Eucalyptus is frequently planted near built areas which is why the importance of the distance to built variable is so high.