

Untitled

Ericka B. Smith

2/27/2021

Introduction

Background

Urban Heat Islands,

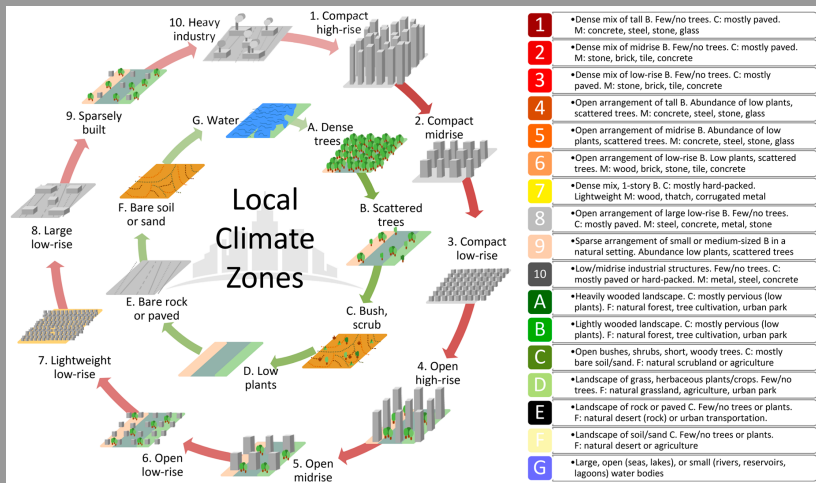


Figure 1: Local Climate Zone classes. Originally from Stewart and

Objective

The goal of this

All code and higher resolution images for this project can be found on GitHub at <https://github.com/erickabsmith/masters-project-lcz-classification>.

Methods

Data

The LCZ reference data

The Landsat 8 data

Table 1: Acquisition Dates of Each Landsat 8 Scene

Scene	Date
1	29-Nov-2013
2	15-Oct-2014
3	16-Nov-2014
4	18-Oct-2015

dat

All 9 available bands of all 4 Landsat scenes amounted to 36 input variables. Each pixel is an observation,

table

Table 2: Delineation of training and test data by polygon and pixel.

Local Climate Zone	Train	Test
Class 1: Compact high-rise	13 (295)	13 (336)
Class 2: Compact mid-rise	6 (117)	5 (62)
Class 3: Compact low-rise	7 (185)	7 (141)
Class 4: Open high-rise	10 (275)	9 (398)
Class 5: Open mid-rise	4 (79)	4 (47)
Class 6: Open low-rise	6 (60)	7 (60)
Class 7: Lightweight low-rise	0 (0)	0 (0)
Class 8: Large low-rise	4 (90)	5 (47)
Class 9: Sparsely built	0 (0)	0 (0)
Class 10: Heavy Industry	4 (107)	5 (112)
Class 11: Dense trees	7 (762)	7 (854)
Class 12: Scattered trees	6 (194)	7 (213)
Class 13: Bush, scrub	4 (459)	5 (232)
Class 14: Low plants	6 (346)	6 (222)
Class 15: Bare rock or paved	0 (0)	0 (0)

Random Forests

Random forests consist of many decision trees.

Evaluate

Splits are typically evaluated by Gini impurity or entropy:

$$\text{Gini Impurity} = I_G(t) = 1 - \sum_{i=1}^C p(i|t)^2$$

$$\text{Entropy} = I_H(t) = - \sum_{i=1}^C p(i|t) \log_2 p(i|t)$$

Where i is a class in the predictor variable, ranging from 1 to C . C is the total number of classes represented for a particular node, t . $p(i|t)$ is the proportion of samples that belong to each i , for a particular node t .

Accuracy Assessment

In line with the methods used in our reference paper and the remote sensing field, accuracy metrics will include the following:

$$\text{Overall Accuracy} = OA = \frac{\text{number of correctly classified reference sites}}{\text{total number of reference sites}}$$

OA_{urb} and OA_{nat} will be used, which are the same as overall OA but only includes the urban and natural classes, respectively.

$$UA(z) = \frac{\text{number of correctly identified pixels in class } z}{\text{total number of pixels identified as class } z}$$

$$PA(z) = \frac{\text{number of correctly identified pixels in class } z}{\text{number of pixels truly in class } z}$$

Results

Varying the Parameter for Number of Trees

The parameter for the number of trees was initially varied between 5 and 500 at intervals of 5. The resulting overall accuracy metrics indicate a leveling off around 125 trees (Figure 2). There's also a clear distinction between accuracy in urban vs. natural classes, with natural classes having a much higher overall accuracy.

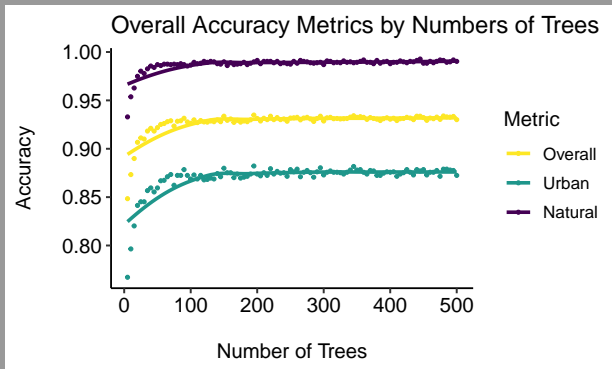
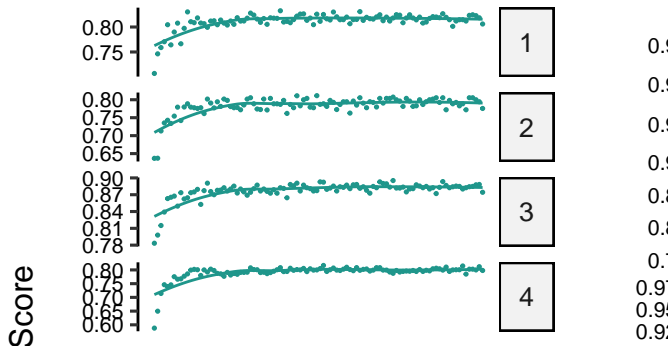


Figure 2: The increase in OA metrics levels off around 125 trees. Urban classes (1-10) have much lower accuracy than natural classes (11-17). These metrics were calculated based on the out-of-bag dataset.

another plot

F-1 Score by Class for 5 to 500



Type

Natural

Predicting on the Test Dataset

Validation Metrics

OA and F-1 metrics dropped dramatically upon applying the random forest to the test data (Figure 4).

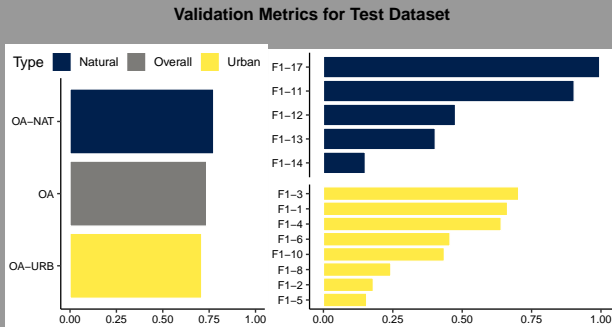


Figure 4: Accuracy among random forest predictions for the test dataset varied widely, but was lower than expected for F-1 scores,

Importance Measures

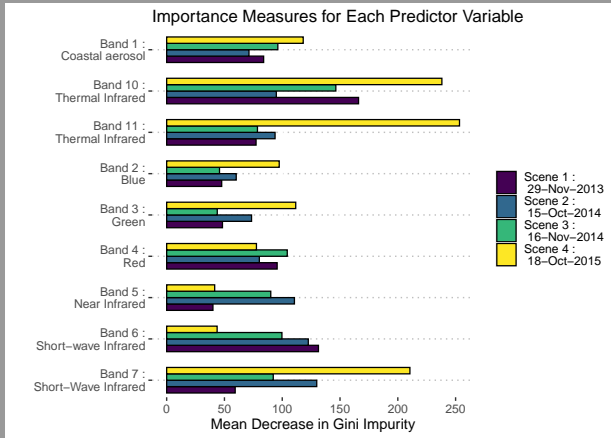


Figure 5: There is not a clear pattern in Mean Decrease for Gini Impurity between the different bands and scenes, though there is some indication that bands in scene 4 were particularly effective as

A Full Prediction

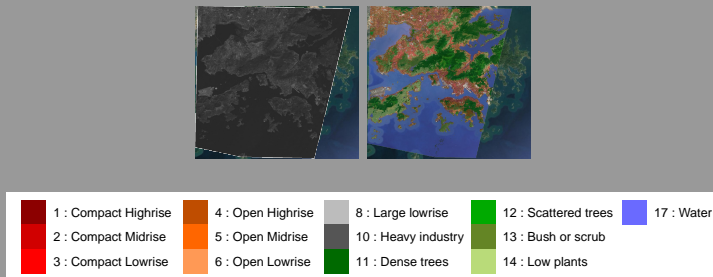


Figure 6: Imagery of the area of interest. Each has a basemap of satellite reference imagery. Top Left: Only satellite reference. Top Right: One Landsat 8 Scene. Bottom: A fully predicted LCZ map.

Discussion

The results of these analyses point to two primary issues