Cotton-top tamarin suitable habitat identification and classification

Xiaohe Yu

Background

The cotton-top tamarin, a New World primate endemic to the forests of Northwest Colombia, is listed as critically endangered by the International Union for Conservation of Nature (IUCN) with roughly 6,000 individuals remaining. Their habitat continues to be threatened by various human activities such as deforestation relating to agriculture and urbanization.

Major threats to these tamarins include deforestation from the logging industry, agriculture, and urbanization.

Purpose

This project utilized Landsat Landsat 8 OLI at the year of 2018, and integrated ground imagery from data within the historic range. With this data the a current, broad-scale land cover map covering the entire historic range of the cotton-top tamarin will be created attempting to distinguish suitable habitat. With the classification results, the local conservation agency will have a better understanding of the condition of forest habitat which will aid in future conservation policy and reforestation initiatives.

Instead of utilizing ISODATA and K-Means classification method, the project focuses on exploring more efficient classification methods to identify suitable habitat area. Especially to differiciate the tropical forest and the costal vegetation. Thus, the SOM unsupervised is applied to get classes information which then used as the training data to fit the ensemble classification trees.

Algorithm

The unsupervised classification algorithm Self-organizing map (SOM) will be employed to reduce the dimensionality of the multispectral imagery in order to produce a two dimensional classification map. The function we used in matlab is Selforgmap(…). After the supervised classification, the Non-Parametric ensemble decision tree will be implemented for supervised classification. The function we used in matlab is the Fitcensemble(…).

Data input and output

Landsat 8 OLI images were acquired from the USGS. More specifically, Band1 to Band 6 are used as inputs for the SOM algorithm, which is listed in the table below. Data size is around 800mb.

Fig. 1 Landsat8 imagery of the study area

Table. 1 The input spectral bands of SOM

Ultra Blue	Blue	Green	Red	NIR	SWIR1
1,666,560 pixels	1,666,560 pixels	1,666,560 pixels	1,666,560 pixels	1,666,560 pixels	1,666,560 pixels

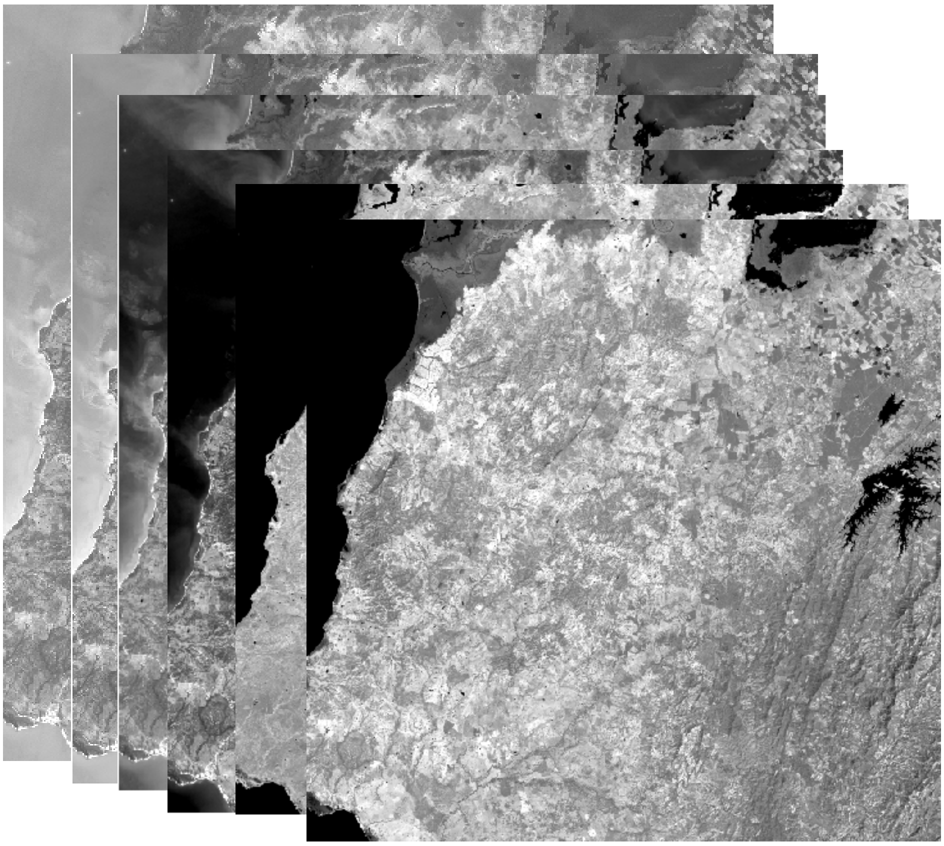


Fig. 2 Landsat 8 bands 1 to bands 6

The output of SOM is a landcover classification map in the target area. Because our purpose is to distinguish the forest area and non-forest area, all the classified results are merged into two classes. After that, the six bands as well as the binary classification results from the SOM will be used as input data for the ensemble classification method.

Table. 2 The input data for ensemble classification method

Ultra Blue	Blue	Green	Red	NIR	SWIR1	Label
X1 …. …. …. X1,666,560	X1 …. …. …. X1,666,560	X1 …. …. …. X1,666,560	X1 …. …. …. X1,666,560	X1 …. …. …. X1,666,560	X1 …. …. …. X1,666,560	C1 …. …. …. X1,666,560

A confusion matrix and a predictor importance rank map will be created for the Fitcensemble method.

Classification Results

In order to differiciate the forest and vegetation, totally 25 neurons are created to implement the SOM. The 25 classes represent 25 landcover types.

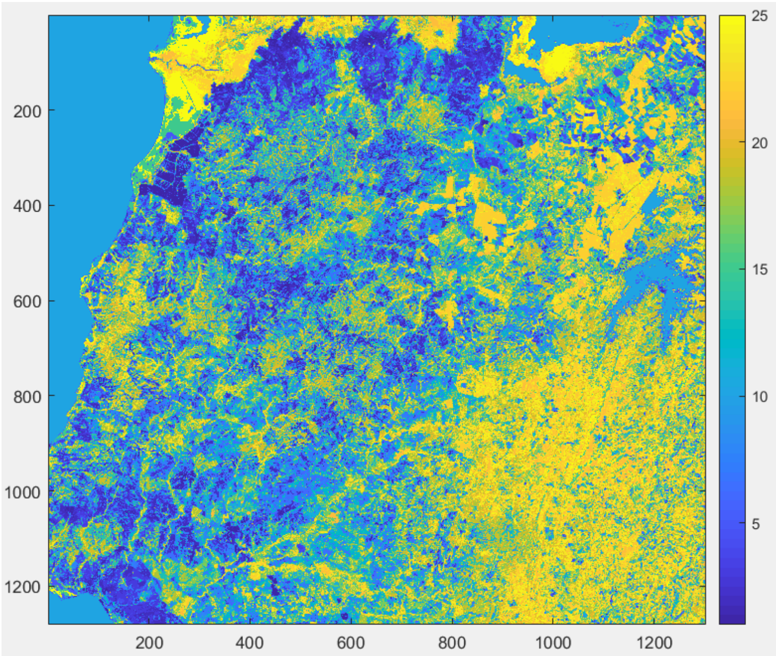


Fig. 3 The SOM classification results with 25 classes

Referring to the ground truth imagery, all classes representing forest are merged together and coded as 1. The rest of classes are merged and coded as 0. Based on visual check, the class 12, 16,18,19 and 20 are merged together to represent the forest area. The forest classification result is then represented on the top of the ground true imagery in the map below.

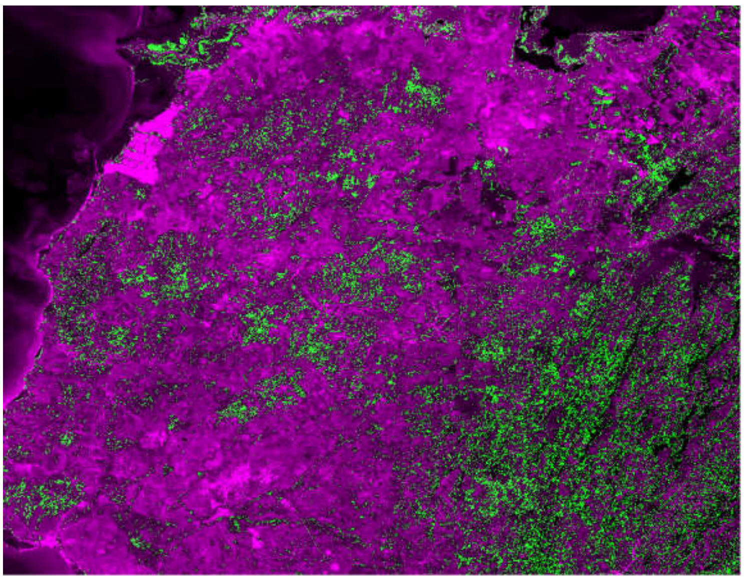


Fig. 4 The forest class based on SOM

Now we use the results from SOM as the labels and band 1 to band 6 as predictor variables to train and classification tree by using Fitcensemble method. We split the all data set into training group and testing group which account 90% and 10% respectively. A forest map is plot based on the classification tree model and the result is shown below.

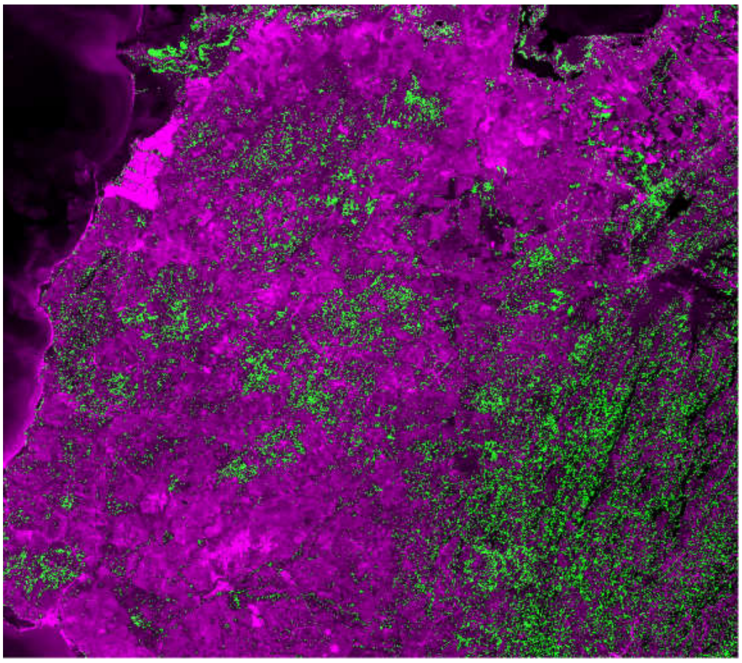


Fig. 5 The forest class from ensemble supervised method

A relative importance chart is plotted illustrating the relative importance of each spectral bands in distinguishing forest and non-forest areas. In this case, band 3 band 1 and band 2 are the three most important band to differentiate forest and non-forest areas. The importance of band 5 and band 6 are also very important. Band 4, however, is not very effective in differiciate the forest and non-forest areas in our case.

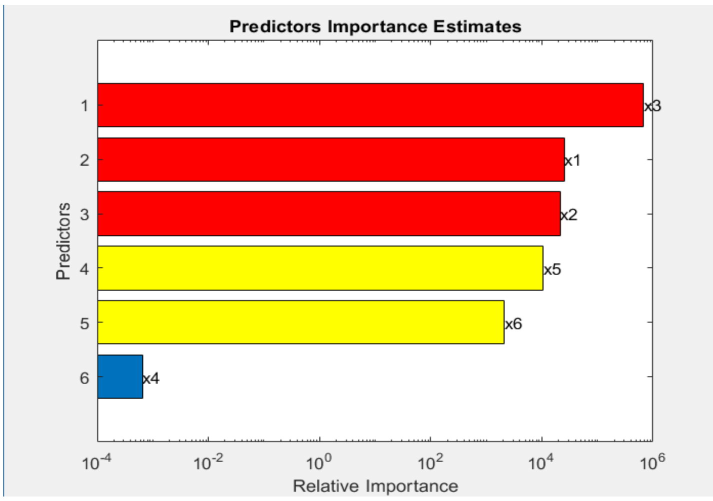
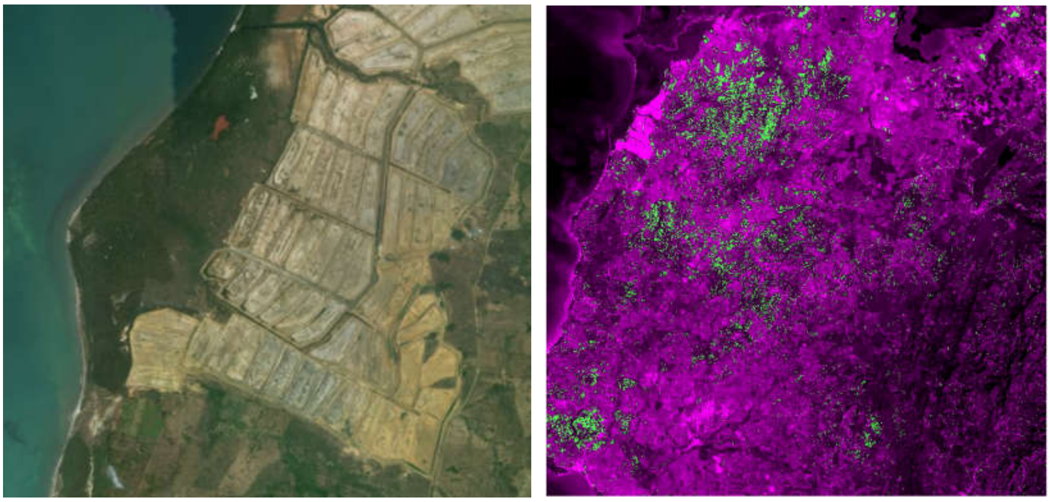


Fig. 6 The bands importance rank

Quality Assessment

For SOM unsupervised method, the quality assessment is mainly based on visual check. By overlapping the classified images with the high resolution ground imagery, the classified results could be properly labeled. In order to implement a more accurately quality assessment, random sample point and ground truth data is required, which is not included in this project.



(a) (b)

Fig. 7 (a) The high resolution ground imagery, (b) the overlayed image for visual check

For the supervised ensemble classification method, the data set is splitted into training and testing group. Tarining group is used to traing the ensemble classification model, and the testing group is used to conduct the accuracy assessment. There are 1,499,904 pixels in the training group and 166,656 pixels in the testing group. An confusion matrix is produced as the picture below. The overall prediction accuracy is 99.7%. Thus, the model is accurate enough in this case.

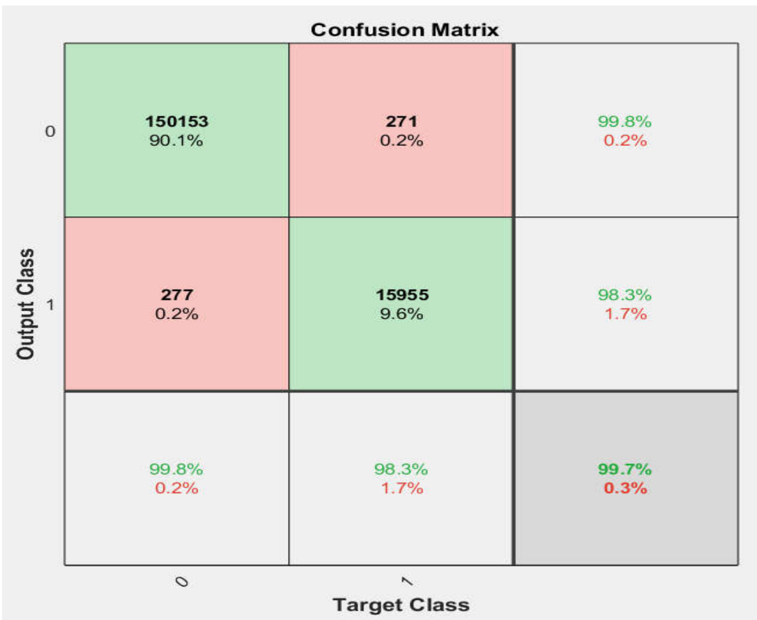


Fig. 8 The confusion matrix of ensemble model

Reference

USGS website https://earthexplorer.usgs.gov/

Goncalves, M. L., M. L. A. Netto, J. A. F. Costa, and J. Zullo Junior. "An unsupervised method of classifying remotely sensed images using Kohonen self-organizing maps and agglomerative hierarchical clustering methods." International Journal of Remote Sensing 29, no. 11 (2008): 3171-3207.

Filippi, Anthony, Iliyana Dobreva, Andrew G. Klein, and John R. Jensen. "Self-organizing map-based applications in remote sensing." In *Self-Organizing Maps*. InTech, 2010.

Lary, David J., Gebreab K. Zewdie, Xun Liu, Daji Wu, Estelle Levetin, Rebecca J. Allee, Nabin Malakar et al. "Machine Learning Applications for Earth Observation." In *Earth Observation Open Science and Innovation*, pp. 165-218. Springer, Cham, 2018.