Josh Virene
NR426: Programming for GIS I
10 March 2023
Final Lab Report

Using National Agricultural Imagery Program data to classify Land Cover in Big Sky, Montana

Background: Land use / Land cover (LULC) is an important tool in natural resource management that can help determine the type of land across the United States. The ability to classify landcover using aerial imagery is highly valuable, and this lab will demonstrate a workflow to accomplish this task.

Data: The data needed to accomplish this analysis includes:

1. NLCD Dataset- this shows each landcover type and can map landcover across the United States. The year of this data is 2016.
   Source: Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Case, A., Costello, C., Dewitz, J., Fry, J., Funk, M., Grannemann, B., Rigge, M. and G. Xian. 2018, [A New Generation of the United States National Land Cover Database: Requirements, Research Priorities, Design, and Implementation Strategies] (https://www.sciencedirect.com/science/article/abs/pii/S092427161830251X), p. 108-123.
2. NAIP Imagery- This is aerial imagery with four bands: Red, Green, Blue and Near-Infrared. The year of this data is 2019, though in the script the user may specify any year.
   Source: USDA Farm Production and Conservation - Business Center, Geospatial Enterprise Operations
3. Study area: This is the region over which the analysis takes place: This boundary is hand-drawn directly in Google Earth Engine, and coordinates are copied into the Python API.
4. Supporting literature: The majority of this script is adapted from a study that uses the GEE-Python API for landcover classification in the Florida Panhandle: RITIKA PRASAI. Using Google Earth Engine for the complete pipeline of temporal analysis of NDVI in Chitwan National Park of Nepal, 10 May 2022, PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-1633994/v1]

Model: The final landcover map is generated by using a random forest classifier; in the script, the NAIP imagery bands are applied to a set of training points, which are then fed into the classifier.

Importing Modules:
Note: The modules that are needed for this analysis are not pre-installed, there are several steps that are required in order to use the Google Earth Engine functionality in the Jupyter Notebooks environment.

In the terminal window, run the following:

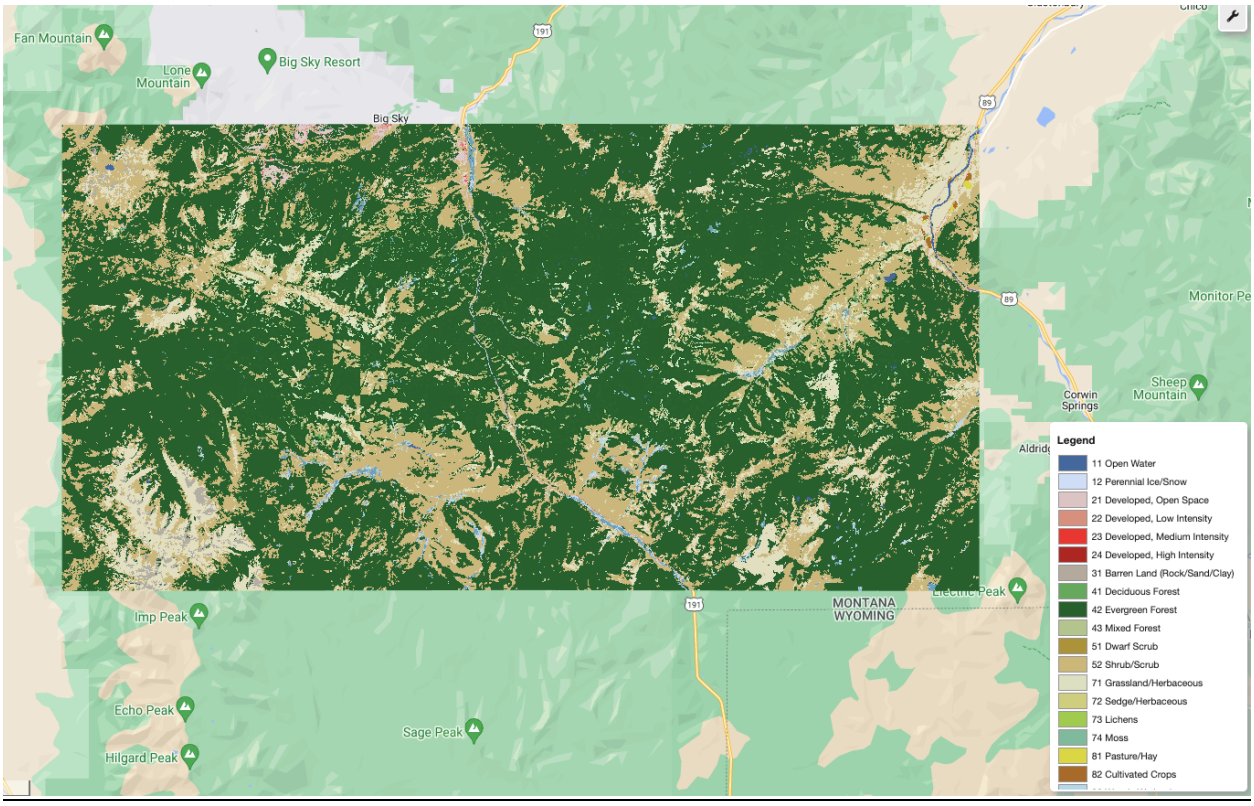
1. Create a new environment:
- conda create name ee (ee = the name of the environment)
- conda activate ee
2. Install necessary packages:
- git install geemap
- conda-forge earth engine-api
- conda-forge earth jupyterlab
3. Run jupyter notebooks (this opens a browser window)
- Jupyter-lab

Terminal window:

```
Last login: Tue Mar  7 18:58:09 on ttys000

The default interactive shell is now zsh.
To update your account to use zsh, please run `chsh -s /bin/zsh`.
For more details, please visit https://support.apple.com/kb/HT208050.
(base) Joshs-MacBook-Air:~ joshvirene$ conda activate geemap
(geemap) Joshs-MacBook-Air:~ joshvirene$ sudo conda install -c conda-forge geemap
Password:
Collecting package metadata (current_repodata.json): done
Solving environment: done

# All requested packages already installed.

(geemap) Joshs-MacBook-Air:~ joshvirene$ conda install -c conda-forge earthengine-api
Collecting package metadata (current_repodata.json): done
Solving environment: done

# All requested packages already installed.

(geemap) Joshs-MacBook-Air:~ joshvirene$ conda install -c conda-forge jupyterlab
Collecting package metadata (current_repodata.json): done
Solving environment: done
```

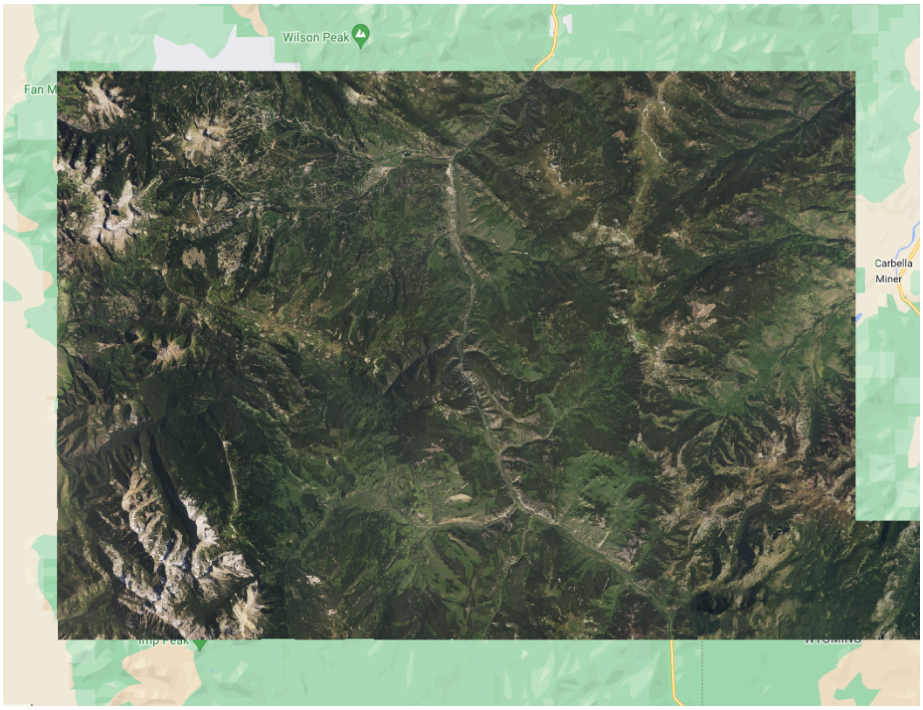
Results:

Image one: NLCD Map for the Big Sky, Montana study area



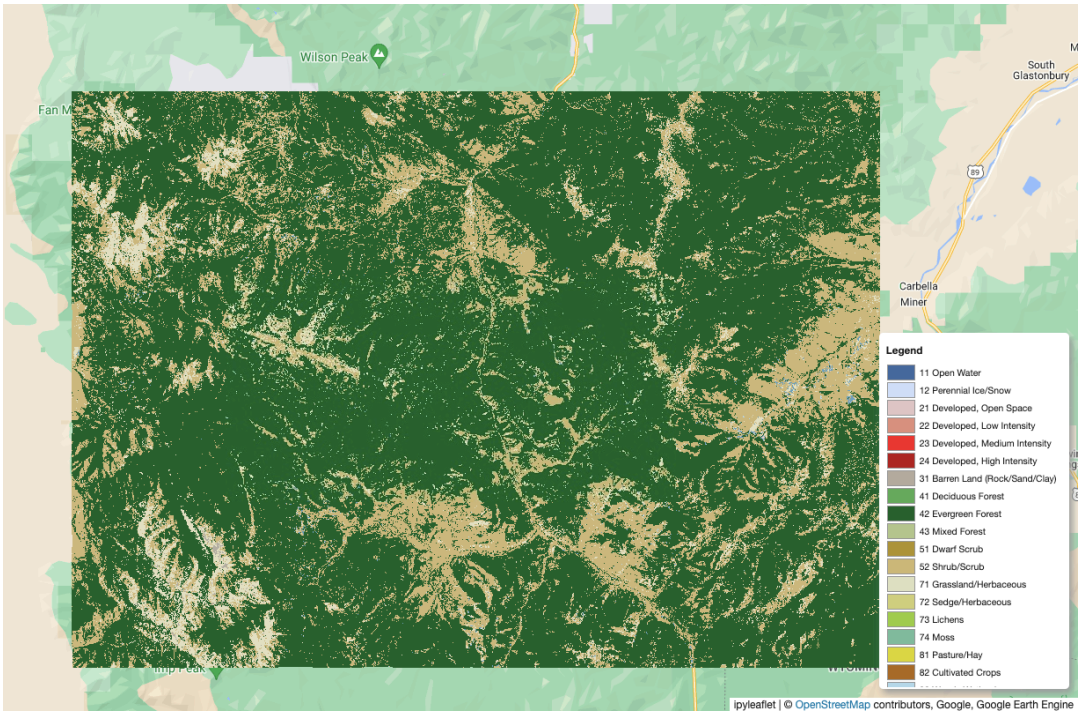
This is the NLCD imagery, showing how the dataset classifies landcover in the study region, which is useful for comparison to the classified imagery the script generates.

Image two: 2019 NAIP Imagery



In this imagery, the bands: Red, Green, Blue, and Near Infrared are used as variables in the random forest classifier model to determine the landcover types in the region.

Image three: Classified Landcover



Discussion:
The accuracy assessment section of the script evaluates model performance; using 5000 training points, the model has an accuracy of 75%. Within the script, users can alter parameters including the imagery source, the bands of imagery used, the number of training points, and many other factors. By changing parameters, the model will either increase or decrease in accuracy.

Comparing these two images the classified landcover image has higher resolution and gives a reasonably accurate description of landcover in the study region and can be validated with the NLCD layer shown in image one. An important consideration is that the NAIP imagery scale is 0.6 meters, while the NLCD layer is at 30 meters per pixel, which helps explain differences in how these images display.