Predicting canopy height using Landsat spectral information through integration of GEDI data for a localized site

Arnav Goel

Geospatial Data Science Lab

Civil Engineering

PURDUE UNIVERSITY®

Problem

- Lidar satellite data (e.g. from GEDI) is expensive and sparse. Spectral information is not, satellites like LANDSAT have a continuously updated ubiquitous coverage over the entire planet.
- Can we predict structural information of a location, e.g., RH95 height metric, using spectral data from satellite imagery ? Is there a strong correlation between them that we can exploit ?
- Training on the limited GEDI sensor's coverage, the objective is to create a model that finds a relationship between this height value and LANDSAT's spectral band values, for a localized site (e.g., Tippecanoe County)

inPUT

- Landsat 8 ARD Collection 2; spectral bands that shall be our features
- GEDI L2A data product; RH95 height metric that shall be our target
- USGS's NLCD dataset; For masking out urban areas
- USGS's 3DEP point cloud dataset; For validation and accuracy assessment.

outPUT

- RH95 height map for the Tippecanoe county; primary focus on non-urban areas(forests and agricultural land)

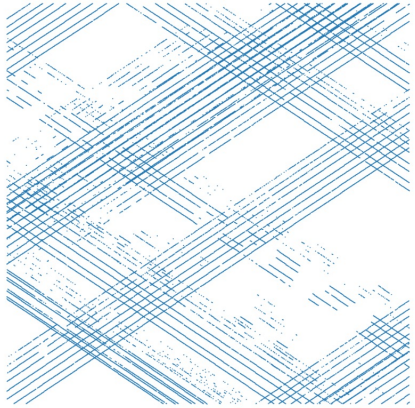
Solution

- Broadly speaking, the steps were as follows:
  - Creating a base raster for the region of interest.
  - Find all the pixel locations whose GEDI RH95 height is available.
  - Build a training dataset all the pixels in the ROI, extracting the band 1-7 values from LANDSAT data of that region(year 2020-22), including their pixel flag.
  - Train a simple neural network on the dataset and use that to predict height values for all the pixels in the base raster -> Build the RH95 NDHM.

dataset

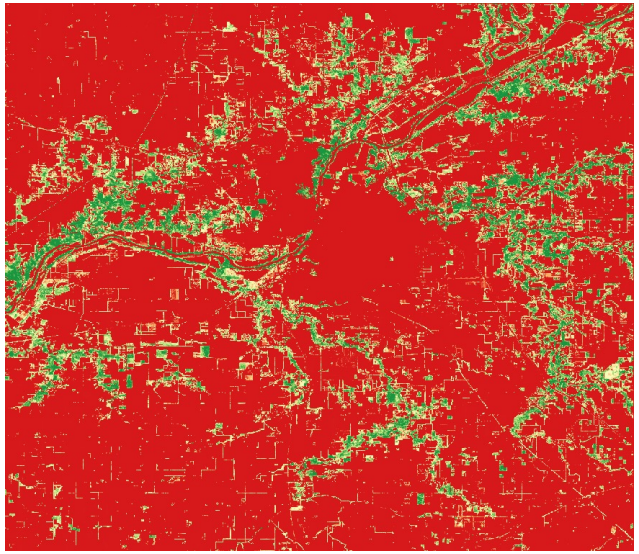
	2	3	4	5	6	7	8	9	11	12
0	0.529774	0.503337	0.458486	0.428414	0.370525	0.322473	0.273409	0.0	0.458839	0.463086
1	0.545415	0.518323	0.469377	0.436886	0.378307	0.322803	0.273282	0.0	0.410340	0.413028
2	0.651413	0.633141	0.606277	0.584448	0.539092	0.529459	0.483730	0.0	0.424383	0.424519
3	0.495355	0.472525	0.422256	0.395484	0.346585	0.328387	0.305525	0.0	0.483723	0.484243
4	0.571846	0.548772	0.501869	0.474981	0.423190	0.386791	0.349886	0.0	0.459473	0.464251
...	...	...	...	...	...	...	...	...	...	...
27523	0.640979	0.621320	0.590704	0.567973	0.528495	0.497612	0.473858	1.0	0.352522	0.355917
27524	0.684866	0.661303	0.617952	0.587694	0.534491	0.461872	0.408313	1.0	0.264114	0.265277
27525	0.681394	0.663591	0.630290	0.604046	0.559027	0.505326	0.450979	1.0	0.248518	0.252873
27526	0.559163	0.538114	0.490335	0.463712	0.415225	0.318469	0.270274	0.0	0.391765	0.396216
27527	0.608178	0.593941	0.556002	0.537880	0.500710	0.494196	0.490170	1.0	0.410087	0.406542

27528 rows × 522 columns

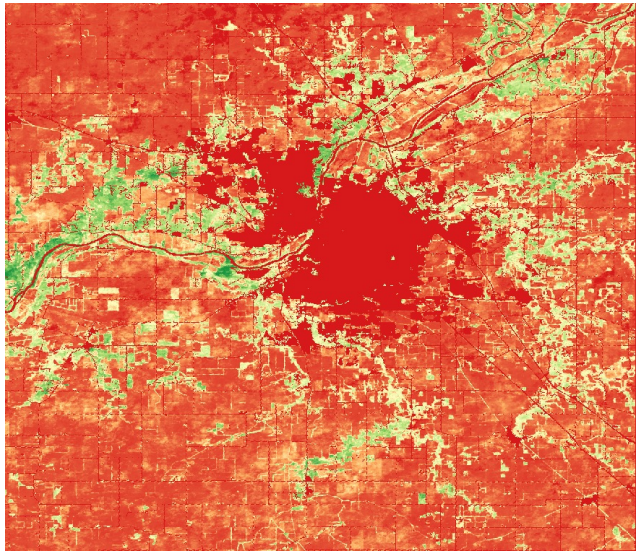


GEDI footprint over Tippecanoe County

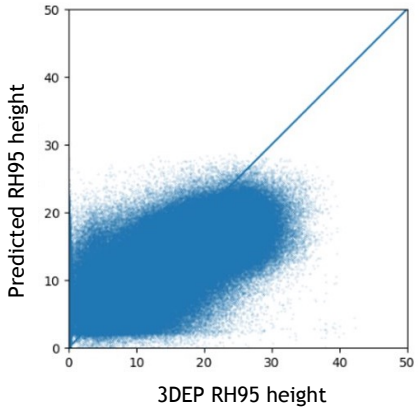
Results



3DEP derived RH95 NDHM



Landsat predicted RH95 NDHM



For each pixel in Tippecanoe county, the graph on the left shows the comparison between the heights predicted solely from Landsat information and their heights derived from 3DEP's point cloud dataset.

The root mean square error between the two heights was 5.028m

Challenges and Next Steps

- Currently, we are trying to experiment with different models and hyper parameters, to improve our accuracy.
- **Challenges** include, overcoming the shortcomings of Root Mean Squared Error as our validation metric. Is it the most appropriate one for this problem?
- How much value would temperature bring into the equation ? Our **next step** is to add ECOSTRESS's land surface temperature as our features and see how it affects our results.