**Predicting urban population using GEDI-L2A tree coverage data for New York State**

Frank Seidl, Luke Kiernan, Keavin Moore, Nicholas Barvinok, Noah Rahman

**Overview**

The tree coverage of a given region of Earth’s surface can provide important implications for researchers and developers concerned with climatological trends over time. This can provide information to describe a link between important human factors, such as groundwater and population density for, example, the state of New York. We hence propose a model that can **predict urban proportion from laser measurements of tree foliage profiles**based on orbital laser measurements onboard the International Space Station (ISS). We then hope to use this training set to predict the relation for another region of interest.

**Data Collection & Modelling**

We plan to use the **GEDI L2A Vector Canopy Top Height** dataset[1] which provides orbital laser readings covering a 25-m beam footprint following the orbit of the ISS. We can use an API to download the relevant data files for our chosen region-of-interest (**New York state**) and clean/preprocess the data based on various flags such as ‘*degrade\_flag’* and ‘*quality\_flag’*. We can then predict ‘**urban\_proportion**’ using other relevant columns in the dataset, such as ‘*landsat\_treecover*’ and ‘*landsat\_water\_persistence*’, but especially using the relative height ‘***rh\_n*’** metrics. These metrics provide the *n*-th percentile of returned beam energy, with these 101 metrics allowing mapping of both the tree canopy height and its vertical structure[2]. We will find relevant correlations between relative height metrics and other relevant features, and test various regression models before deciding on a best-fit model. We choose New York as a test case due to it containing forests, plains, lakes/ponds, and urban developments; the GEDI L2A dataset has great coverage of New York, passing over these various geographical features as the ISS orbits. With our best-fit model, we can then use this tool to try to predict trends for regions other than New York, such as highly-developed areas (e.g., California) or highly-forested areas (e.g., the Amazon, northern Europe).

**Stakeholders & KPIs**

* Real estate developers looking for an appropriate region for housing developments (***not great for climate if they’re cutting down trees, but does give an actual, tangible business scenario for our training***)
* Groundwater researchers (***important implications based on the links Noah sent?****)*
* Vegetation researchers ***(canopy height and vertical structure can be linked to the above, and with urban proportion prediction, relevant regions for a given study can be selected using model)***
* Popular science writers ***(what is the link between tree coverage and population for a given region, e.g., the Amazon rainforest, California?****)*

[1]: <https://developers.google.com/earth-engine/datasets/catalog/LARSE_GEDI_GEDI02_A_002#description>

[2]: https://daac.ornl.gov/GEDI/guides/GEDI\_ICESAT2\_Global\_Veg\_Height.html