

Project Proposal

Samuel Pease, Eva Arroyo, Nikita Zemlevskiy

3/9/2017

Questions:

(NOTE in the datasets section what kind of data we need for each of these comparisons USE defn of old growth forest from TERBOUGH ET AL)

Compare persistent homology of LiDAR of disturbed versus old growth forests

Try and predict forest types (category of forest) by looking at persistent homology of LiDAR of different forest types

Use persistent homologies of forest types to predict biomass and net primary productivity (NPP), compare this to traditionally used methods.

The advent of LiDAR has allowed ecologists to remotely examine forest canopy heights . LiDAR functions by sending a burst of infrared light and recording the time it takes for light to return to the source, then using this to determine the distance the source is away from the detector. They can take several hundred samples in less than a second, making them extremely effective (CITE) By using this technology in planes flying over land surface, ecologists can obtain a very precise visualization of the maximum canopy height across a forest, improving previous methods of assessing forest structure and size from a ground transect. This allows a more precise picture of the complex predictors of biomass and net primary productivity, crucial measures for understanding ecosystem functioning and carbon sequestration. In addition, these can allow for large scale ecosystem classification which can improve our ability to make planetary and local carbon budgets (CITE). However, many methods using LiDAR are unable to make use of the topological data, simplifying it into maximum height and canopy size distribution across a forest. A hypothesis in forest ecology is that in old growth forests canopy height will locally uneven, the height of each tree is determined by independent random processes (such as local soil quality, hurricanes, lightning strikes). By comparison forests disturbed by human activities such as logging, and then regrown will have much more even canopies, because the entire forest stand should be the same age, and factor to the same events (CITE). For our first exploratory question we will see if this ecological hypothesis holds true. In addition, different forest types under different constraints may have different structures of persistence, or more interestingly there may be a general pattern to all forest types across ecosystems. This will be our second question, if it is possible to differentiate forest type using persistent homology. One of the more important questions for ecologists with LiDAR data is the ability to predict biomass and in turn carbon sequestration of a forest remotely. We will also see if using

persistent homology to predict biomass of a forest is more effective than traditional methods. Theoretical ecology predicts that with more uneven canopy there are likely more trees per hectare, resulting in higher biomass and NPP with more uneven canopies in the same ecological system. Therefore persistent homology may be able to capture more of the true factors surrounding biomass than mean stand height, the traditional method.

DataSets:

We will be looking at LiDAR data in order to get the shape of forest tops. There is LiDAR data openly published by the US department of Agriculture Forest Services and by the group OpenTopography. This data is both in the form of point clouds and as direct shape files. We also emailed John Poulson at the Duke Nicholas School of the Environment asking for access to some of his datasets. He is an ecologist with broad interests in the maintenance and regeneration of tropical forests and conservation of biodiversity with a focused on the effects of anthropogenic disturbance, such as logging and hunting, on forest structure and diversity. He should be able to help provide us with data to compare pristine forests in comparison to those which have been disturbed. US forest services also have information on the biomass and net primary productivity of forests in the US.

<https://data.fs.usda.gov/geodata/edw/datasets.php>

<http://opentopo.sdsc.edu/datasets?listAll=true>

<https://nicholas.duke.edu/people/faculty/poulsen>

Mathematical Methods:

We will be using different topological and learning methods to find answers to the questions we have posed. We will use persistent homology to analyze the shape of the profiles of different forests around the world. We will make use of the useful topological toolkits developed by Dr. Harer and his co-workers such as Ripser and TdaTools. We are looking to identify commonalities between forests that have been logged and those that have not been logged. Thus, the learning methods we will be using will most likely be mainly supervised, as we will know apriori whether the forest in question had been logged or not. We will learn a model to predict whether the forest was logged or not based on the shape and on the persistent homology of the canopy. In addition, we may be able to learn models to predict the type of forest given the shape of its treetops. Once these models have been

developed, they can be applied to LiDAR data on other places in the world, where it is not known whether logging has been taking place. To do this, we can use the k nearest neighbors classification, or we could use the SVM model. We can attempt to classify those forests based on the shapes of their canopies and other relevant information. There are many places where logging is illegal, or special laws in place such as a requirement to plant two trees for every tree that is cut down, so there is a legitimate application of this research.