CS486C – Senior Capstone Design in Computer Science

Project Description

Project Title:	**An Application for Characterizing 3D Vegetation Structure in Tropical Ecosystems**
Sponsor Information: NORTHERN ARIZONA UNIVERSITY School of Informatics, Computing, and Cyber Systems	**Patrick Jantz** Assistant Research Professor School of Informatics, Computing, and Cyber Systems Northern Arizona University Patrick.Jantz@nau.edu (928)523-9225

Project Overview:

Climate and land use change are profoundly affecting ecosystem processes and delivery of ecosystem services. For example, drought, high temperatures, and agricultural activities are causing large fires in the Amazon rainforest, one of the largest repositories of biodiversity on Earth. In response to these trends, the Group on Earth Observations Biodiversity Observation Network (GEO BON) has developed a minimum set of variables, known as Essential Biodiversity Variables (EBVs), required to monitor the status of biodiversity globally. EBVs have been designed to leverage large streams of data from Earth observing satellites and other space based sensors. Translating these data streams into products that can enhance decision-making is a major challenge for the EBV program. Specifically, satellite systems can produce data streams of hundreds of thousands of image files. How can one actually make use of this image data in the context of GEO-BON and the targeted set of EBVs to observe and assess? What is needed is a flexible, accurate tool for rapidly classifying satellite images and vector data based on assessing patterns in the data. But how can we do this?

Vegetation structure, the three dimensional distribution of stems, branches, and leaves, is key to understanding ecosystem processes like carbon sequestration, fire dynamics, and animal movement. Until recently, information on vegetation structure came from laborious field measurements or from aircraft lidar (light detection and ranging) scanning of relatively small geographic areas. Recently, NASA funded the Global Ecosystem Dynamics Investigation (GEDI), an Earth Ventures mission that installed a lidar on the International Space Station in December 2018. Over the course of its two-year mission, GEDI will acquire billions of vegetation profiles (Figure 1) across the Earth's temperate and tropical ecosystems, allowing us to characterize vegetation structure in unprecedented detail.

The Vegetation Structure as an Essential Biodiversity Variable (VSEBV) project, based at NAU, aims to use GEDI data to develop a vegetation structure EBV that can be used by policy makers and scientists to improve land use decisions and guide priorities for conservation of biodiversity in tropical landscapes. The problem is that currently our processing workflow is awkward and highly manual, and will quickly be overwhelmed by the high volume of incoming GEDI data. The process starts with the image and vector data to be processed being manually downloaded from NASA Distributed Active Archive Centers (DAACs). It is then analyzed, again manually, using algorithms in R to perform unsupervised classification of GEDI vegetation profiles, called profile classes, and to provide statistical and visual summaries of profile classes. Results are then posted back to shared folders for other scientists to access.

A Solution Vision

The goal of the proposed project is creation of a web application that integrates GEDI data access, data analysis, and data presentation steps, generating easy-to-interpret statistical and visual summaries for end users like the GEO BON program, the United Nations Development Program, or national environment ministries of tropical countries, all of which collaborate with the VSEBV project. In particular, we envision a secure web application that supports two distinct but related categories of end users and usage scenarios:

- **Data Ingestion and Analysis.** The tool must provide a set of graphical interfaces that streamline the data acquisition and integration process. In particular, it must support configuring data sources (e.g. NASA's DAACs) to pull in new data automatically as it appears and feed it into a simple processing pipeline where data are (potentially) cleaned up, then fed through the existing R algorithms to generate the VSEBV profiles, which are then added to the growing datasets available for viewing and downloading.
- **Data Viewing and Access.** The web app must provide clear interfaces for browsing, selecting, graphing and otherwise visualizing the growing data available in the system; a basic interface for downloading selected data as data files for further outside analysis must be supported as well. The interface should allow different access levels and scales of detail: a public "guest" user may be presented with broad aggregate overviews (e.g. a colored map) of conditions, while scientist users will have detailed tools for zooming in to select/analyze/download more specific areas and time periods.

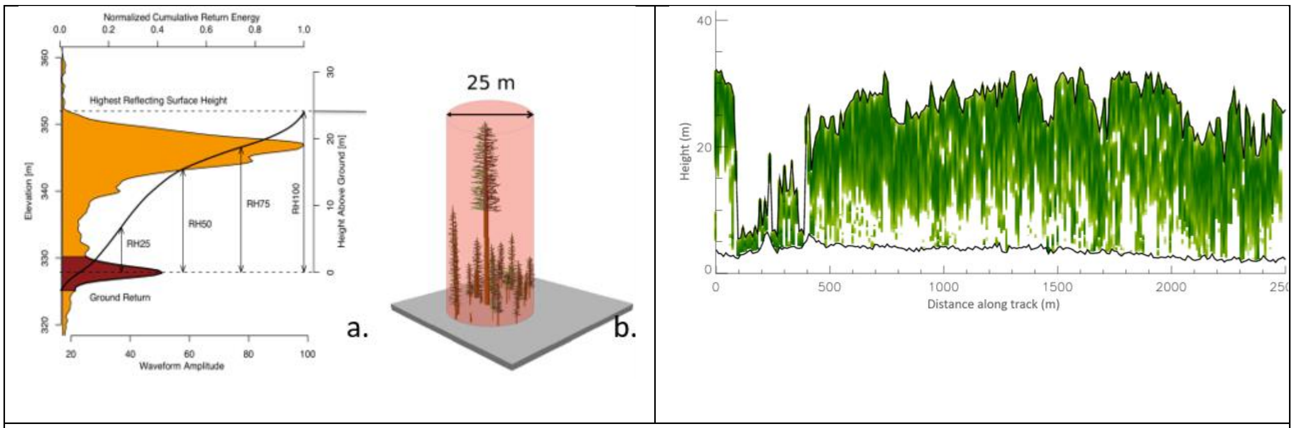

Figure 1. [left] Sample GEDI lidar waveform (a). The light brown area under the curve represents return energy from the canopy, while the dark brown area signifies the return from the underlying topography. The black line is the cumulative return energy, starting from the bottom of the ground return (normalized to 0) to the top of the canopy (normalized to 1). The diagram (b) shows the distribution of trees that produced the waveform in (a). [right] Along track lidar return energy showing vertical distribution of vegetation from a sample GEDI orbit.

The envisioned web app will provide the cornerstone for a world-wide community of scientists, public officials, and policy holders, providing them with accurate, timely access to the VSEBV data extracted from the GEDI data streams. The many functions that this application can have can be organized and prioritized as follows:

Core functionality: A minimum viable product includes

- A well designed application that streamlines access to NASA scientific data and integrates well with modern scientific computing software. The nature of the application (standalone, web-based, etc.) will be determined via a client/team consultation.
- Ability to accept and output tabular and geospatial data in common formats (e.g. shapefile, GeoJSON).
- Statistical and visual summaries of vegetation structure for user-defined areas of interest.

Augmented functionality: A truly useful product includes

- Ability to integrate multiple spatial data types, such as raster and vector data, to improve characterization of vegetation structure.
- Ability to scale analysis to large numbers of GEDI observations.
- Ability to output spatial data for input into other spatial decision support systems used by partner organizations.

Stretch goals: Supporting scientific discussion and community functions

- Attachment of discussion forums to particular data/result sets, to allow registered community members to add information or commentary on the significance or limitations of the result.
- Ability to link various analyses (VSEBV won't be the only analysis done!) on image data sets, to help viewers put together various analysis and evidence regarding an area.
- Ability to link scientific publications to the data sets that drive them. Conversely, ability to group images into named "datasets", with ability to link to these externally, e.g., so that authors could provide a direct link to supporting data in their publications.

Knowledge, skills, and expertise required for this project:

- Familiarity with modern, secure Web2.0 design (e.g. REST principles).
- Familiarity or interest in statistical and scientific computing.
- Familiarity or interest in unsupervised learning approaches.
- Familiarity with graphical interface design, usability, and scientific visualization

Equipment Requirements:

- Ordinary development platform and software/tools freely available online.
- Client will provide access to datasets and models as needed.

Software and other Deliverables:

- The software application as outlined above, tested and refined with project partners/end-users.
- A strong as-built report detailing the design and implementation of the product in a complete, clear and professional manner. This document should provide a strong basis for future development of the product.
- Complete professionally-documented codebase, delivered both as a repositoty in GitHub, BitBucket, or some other version control repository; and as a physical archive on a USB drive.