Tree Species Composition in Urban Forests

Due by Friday December 22 2017

Background
As of 2010, 82% of the U.S. population lived in urban areas and 50.5% worldwide. As urban lands and population continue to increase, urban areas are sources of greenhouse gas emission, increased streamflow, nutrient export, and decreased biodiversity. However, urban forests provide a range of ecosystem services to offset these negative environmental effects of urbanization. Urban trees sequester carbon through photosynthesis, decrease stormwater runoff by intercepting rainwater, reduce nutrient export through root system, and increase biodiversity by providing habitat and food to a variety of wildlife. Information on tree species composition and stand structure are necessary for understanding urban forest status and functions. It can be also used to assess the impacts of insect disturbance (e.g. Emerald ash borer, and Dutch elm disease), and help city foresters make management strategies.

Rapid advances in remote sensing technology enable forest survey over a large area in an efficient way. Generally, remote sensing is to acquire information of objects on the earth through sensor instruments put on a platform like airplane or satellite. Different sensors are designed for a wide range of purposes and applications. For example, multi-spectral remote sensing has been widely used in global mapping of land cover and land use change. For the purpose of this project, two types of sensor (Lidar and AVIRIS hyperspectral sensors) are of particular interest (see Data section for more details).

Objectives
The overarching goal of this research is to map the tree species composition for forests in the Madison, Wisconsin area, using remote sensing data. The specific goal is to build predictive models for the proportion of abundance for a given species based on remote sensing data.

Field Data
There are two types of data: field data and remote sensing data. For the field data, each observation is a measurement from a field plot sampled by expert forestry students. At each plot, the researchers have calculated the basal area (BA, trunk cross sectional area) of each species present at the plot. Composition can therefore be described two ways: basal area by species, or relative basal area by species. That is, for each species on a plot, we can divide its basal area by the total plot basal area to get an estimate of relative basal area. The relative basal area (RBA) by species is what we want to predict. At each location we are really only interested in correctly estimating species with RBA greater than about 10%. Species with less than 5-10% would just be considered “low” or “rare”, compared to other species. All estimates of species need to add to 100% (we cannot have more than 100% relative basal area). This gives us proportional estimates of composition by species. Later we can group these together to estimate forest types.

Remote Sensing Data
The remote sensing data include two sources: (1) Hyperspectral imagery from NASA’s AVIRIS sensor, and (2) Lidar data collected by the city of Madison. Hyperspectral data consist of reflectance measurements every 10 nm between 400 and 2500 nm, excluding wavelengths where there are atmospheric absorption features that interfere with the remote sensing signal. This means that there are on the order of 182 “bands” of data.

Lidar data consist of measurements that come from laser pulses reflected from canopies, branches, or ground, and are summarized into a number of metrics of both height distribution from the lidar return and intensity distribution. Think of it this way: the hyperspectral information tells you the reflectance of the canopy by sunlight, and the lidar data gives you information about the vertical structure of the canopy.

Different species have different reflectance characteristics and different vertical profiles that we can use to identify the composition of the species that occur in a location. The differences in the species may have to do with different leaf shapes (broad vs. needle leaves), or different chemistries, or different canopy shapes (think of how a conical fir tree is different from a more spherical oak). To that end, we have used the hyperspectral imagery and lidar data to map canopy chemistry and vertical structure, independent of species composition. Specifically, we have measures of foliar concentration of carbon, cellulose, fiber, lignin, leaf mass per area (LMA), nitrogen, as well as indices NDVI, PRI, NDWI and NDII capturing vegetation abundance, light use efficiency and water content, all derived from the hyperspectral images. In addition, we have biomass, basal area, average tree diameter, canopy height, crown length, and crown width, all derived from lidar.

Data File

The data set available includes:

1) Field data: plot ID, total basal area, basal area for 10 tree species, and basal area for the rest of species in the field plots;
2) Remote sensing data derived from lidar data: The measures of forest structure derived from lidar for all the listed field plots are shown in the columns AGB, BA, DBH, Ht, CL and CW;
3) Remote sensing data derived from hyperspectral data: The measures of canopy chemistry, LMA and indices derived from hyperspectral imagery for all the listed field plots are shown in the columns of C, Cellulose, Fiber, Lig, LMA, N, NDVI, PRI, NDWI and NDII, as well as standard deviations of these measures within the plot: C_sd, Cellulose_sd, Fiber_sd, Lig_sd, LMA_sd, N_sd, NDVI_sd, PRI_sd, NDWI_sd and NDII_sd .

The datasheet with the "Data" tab lists field data and remote sensing data; and that with the "Description" tab gives more detailed description of each column of data.

Specific Questions

A. Using hyperspectral data, construct a predictive model for the proportion of: (1) white oak (2) white pine and (3) ash.
B. Repeat Question A but this time, use the lidar data.
C. An emerging idea is to combine both types of remote sensing data for the purpose of constructing a predictive model. Explore this possibility with each of the three species above (i.e., (1) white oak (2) white pine and (3) ash) and argue for or against this idea of combining hyperspectral data and lidar data.
D. What are the essential steps to take next for achieving the overarching goal of mapping tree species composition for forests in the Madison, Wisconsin area, using remote sensing data? Provide your recommendations to the researchers and explain your reasoning. You do not need to carry out these steps.

Data Analysis and Report

You are to conduct an analysis of the data for the three questions above and write a report directed to the researcher.

The maximum length for your report is 12 pages; this includes title, abstract, introduction, data, methods, results, and discussion. The report text must be double-spaced and in a font that produces no more than 32 lines of text per page (e.g., 1.5 line spacing, standard width, and various 11pt fonts in Word; or various LaTeX formats using 11pt or 12pt). Integrate key figures and tables (with captions) into the report.

Information critical to major analyses should appear in the main body of the report. A very brief Appendix A for supplementary materials is not necessary but may be included. Include your computer code in Appendix B. Appendixes are not counted toward the 12-page limit. Number all the pages on your report including the appendixes.

Your analysis should be thorough, correct, and as straightforward as possible. Your writing should be thorough, accurate, clear, and concise, while containing the appropriate amount of "hard" statistical information. Finding the balance between too much and too little formal statistical information is a key part of report preparation.