Forest biomass estimation using kNN in the Highlands of New Jersey

Inga Parker La Puma and Rick Lathrop
Center for Remote Sensing and Spatial Analysis
Department of Ecology and Evolution
Rutgers University
14 College Farm Rd.
New Brunswick, NJ 08901
iparker@crssa.rutgers.edu
lathrop@crssa.rutgers.edu

Abstract: The Highlands region of northern New Jersey is composed of a temperate deciduous and mixed coniferous forest matrix. This area has recently been experiencing high development pressure with an accelerating rate of forest conversion to urban land uses. The purpose of our ongoing project in the Highlands is to estimate the loss of forest biomass associated with urban land use conversion. The objective of this portion of the study was to examine the efficacy of using kNN techniques using satellite remote sensing and a ground-based forest inventory in estimating forest biomass at a single point in time. 174 forest inventory points were sampled during the 2003 growing season and biomass calculated via diameter-based regressions for North American tree species (Jenkins et al 2003). We acquired Landsat5 Thematic Mapper images from August 25th, 2003 and September 10th, 2003 of the Highlands area to correspond with data collection dates. The kNN procedure was utilized via a C++ program which uses a leave-one-out validation module to estimate root mean squared error (RMSE). The number of neighbors (k) as well as distance decomposition (d) functions were compared with differing baseline datasets. Different datasets were based on varying cloud cover amounts in each image as well as the fact that some data points were located far south of the majority of the Highlands points in a mixed-use estate area (Duke Farms). Therefore several combinations of points were tested for the best RMSE and k combination. Choosing the correct k, along with the appropriate distance function was based on a method discussed by McRoberts et al (2002) whom describe the process as: 1) selecting the lowest RMSE over a range of k and d combinations (RMSEmin), then 2) choosing the lowest k that corresponds to an RMSE that is within an a priori percentage of RMSEmin (e.g. <.5 %). Our results show that the best data set for predicting biomass in the area contained both dates of imagery but excluded sites that contained high thin cloud cover and the more distant points of Duke Farms. The result of these exclusions gave us a total sample size of n=43, with k=17, d=1 and a RMSE of 71 Mg/ha (63412.02 lbs/acre). The RMSE selected represented approximately +-28% of the observed sample mean. Using only the September image and excluding Duke sites resulted in a sample size of n=150 with k=11, d=1 and a RMSE of 111.51 Mg/ha (99574.25 lbs/acre) representing +-54% of the observed sample mean. Further research will be focused on assessing the accuracy of the best fit model between image dates. Once the best fit model has been tested, we intend to apply it to a summer image from 1995 to understand how land cover change has affected biomass in the Highlands region over the last decade.

Jenkins JC, Chojnacky DC, Heath LS, Birdsey RA (2003) Comprehensive database of diameter-based biomass regressions for North American tree species. General Technical Report NE-319, Northeastern Research Station USDA Newtown Square, PA.

McRoberts RE, Nelson MD, Wendt DG (2002) Stratified estimation of forest area using satellite imagery, inventory data, and the k-Nearest Neighbors technique. Remote Sensing of Environment 82:457-468.