COMBINING MULTISPECTRAL IMAGERY WITH HABITAT SPECIFIC ELEMENTS TO LOCATE EASTERN HEMLOCK

Jarrod Doucette, Research Assistant
Dr. William Stiteler
Dr. Lindi Quackenbush
SUNY Environmental Science and Forestry
Syracuse, NY USA
jsdoucet@syr.edu
wmstitel@esf.edu
ljquack@esf.edu

ABSTRACT

Eastern Hemlock (Tsuga Canadensis) fills an important role in the forests of the Eastern United States. This hemlock is threatened by the spread of an invasive species called the Hemlock Woolly Adelgid, and requires human assistance to survive this pest. The goal of this work is to help forest managers direct their management efforts by developing accessible methodology for locating Eastern Hemlock using readily available and affordable data. Leaf on, leaf off, and spring Landsat ETM+ imagery and a 10m Digital Elevation model are the foundation of the model, and several layers are derived from topographic data to provide information on hemlock habitat. The Classification and Regression Tree technique was chosen to unlock the complex relationships and interactions among the datasets that affect the growth and prosperity of hemlock. Ground reference plots collected by the Forestry Organization Remote Sensing Technology Project were used to train and assess the accuracy of models created by the Cubist data mining software.

INTRODUCTION

The accurate location hemlock is important to the efforts directed at monitoring and controlling the spread of the Hemlock Wooly Adelgid. Prioritizing resources becomes an issue when hemlock stands throughout the Eastern United States are threatened, and the quantity of hemlock becomes invaluable in assigning those resources. Classification and Regression Trees (CART) allow the incorporation of hemlock habitat elements and hyperspectral imagery into a predictive model. With the use of ground reference data, and minimal investment in data and time this methodology can methodology can be used to map the amount of hemlock across a region.

Eastern Hemlock and the Hemlock Wooly Adelgid

Eastern Hemlock (*Tsuga Canadensis*), commonly referred to as Canada Hemlock or Hemlock Spruce (USDA, 1965), is a member of the pine family. It is an evergreen with a dense conical crown, and drooping branches. Hemlocks have short dark green needles, and small ovoid shaped dark brown cones. Hemlocks do well under shade, and commonly spend their adolescence growing slowly in the understory. Hemlock is typically found in cool and moist areas with well drained soils. The native range of hemlock extends from northern Georgia in the South, north to Nova Scotia, and west through the Great Lakes.

The Hemlock Woolly Adelgid is native to Asia where host resistance and natural predators allow it to coexist with hemlock. Limited only by its dependence on animals and humans (McClure 1990) the adelgid has spread from its initial 1951 introduction in Virginia to stands from Northern Georgia to Southern Maine. Adelgids spend the majority of their life at the base of a hemlock needle. They feed by injecting stylets through the base of the needle into the xylem rays of the branch where they harvest the trees nutrients. Within a few months of infestation the needles loose their color, and eventually drop as a result of nutrient depletion and the adelgid's toxic saliva. After several years the adelgid population becomes established and extensive needle drop results in branch dieback. While hemlocks have been observed to persist for several years after heavy adelgid infestation, their increased susceptibility to drought and other damaging agents often lead to mortality within 3 – 10 years (Mayer *et al*., 2002). The lack of natural resistance and predators requires human intervention on protect Eastern Hemlock. Remote detection of the tiny Hemlock Wooly Adelgid is not an option, and physically tracking and assessing infestations is a costly method. These costs can be dramatically decreased by limiting efforts to key areas with significant hemlock population.

Catskill Study Are and Reference Plot Data

The Catskills Park is a mountainous region located across the counties of Ulster, Greene, Delaware, and Sullivan in southeastern New York. It is part of the Allegheny Plateau physiographic region, and the Appalachian Mountain system. The park, comprised of public and private land, is home to almost 300,000 acres of forest preserve and ninety-eight peaks over 3,000 feet high. It is home to a handful of reservoirs that provide drinking water for millions of people in the Hudson Valley and New York City.

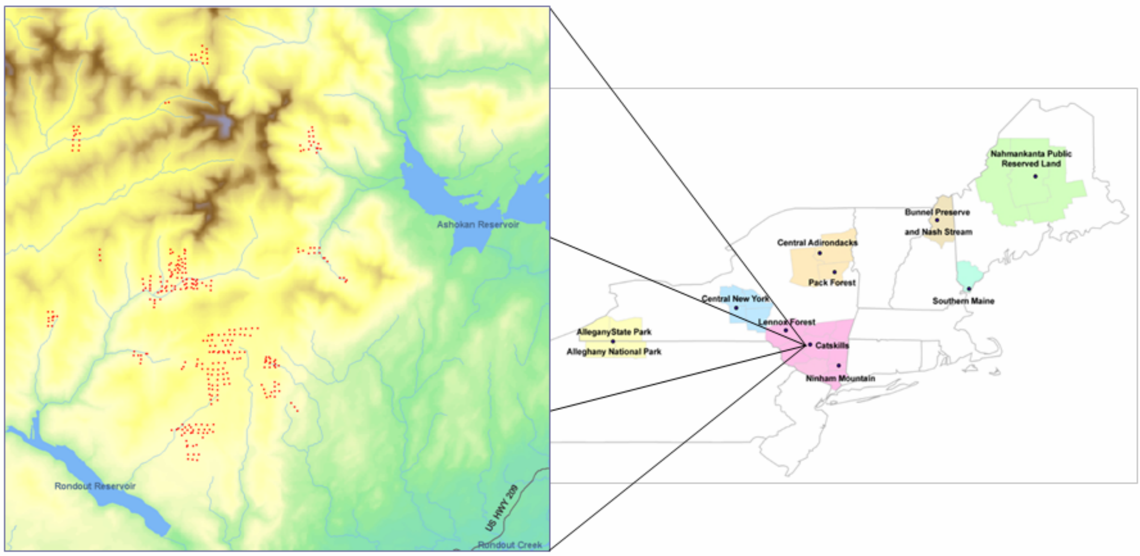

Figure 1. Map of Catskills Study Area with forest plots.

Forest plots were primarily located in western Ulster County near the Ashokan and Rondout Reservoirs (Figure 1). A plot consists of all trees greater then 3" diameter at breast height (DBH) within a 15m radius of the plot center. Trees were located based on their azimuth and distance from plot center, and characteristics recorded for each tree include species, DBH, crown position, and general health. Plot data was collected by the Forestry Organization Remote Sensing Technology (FOReST) Project, located at the State University of New York (SUNY) College of Environmental Science and Forestry (ESF) (http://forest.esf.edu/, 2005). In addition to the Catskills similar forest plot data was recorded in seven other forested areas throughout the Northeastern United States to server as ground reference for remote sensing applications.. Plots in the Catskills were collected during the summers of 2001 through 2005. Systematic random sampling was used to locate plots on an approximately 300m grid. The center of each plot was record using a survey grade GPS instrument.

Northern Hardwoods (37%) and Oak/Hickory (34%) are the dominant forest types in the Catskills. Eastern Hemlock is the fourth most common tree species in the Catskills behind Red Maple, Sugar Maple, and White pine. Of trees greater then 11" DBH only White Pine and Sugar maple are more common (Alerich 1995). The large presence of hemlock makes the Catskills a prime location to develop a model, but adds challenges to assessing model accuracy. The reference plots closely reflect the regional forest composition except for a lack of non-hemlock conifers (Figure 2). Roughly 2/5 of the plots are pure deciduous while the remaining 3/5 are composed of a mixture of hardwood and softwood. In the 148 out of 210 cases where the coniferous component of the mixed forest is purely hemlock a model that accurately predicts the amount of coniferous forest would also correctly predict the amount of hemlock. With this in mind the sampling strategy was modified to extract the most information from the remaining 37 mixed coniferous that contained hemlock, and 25 coniferous plots with no hemlock.

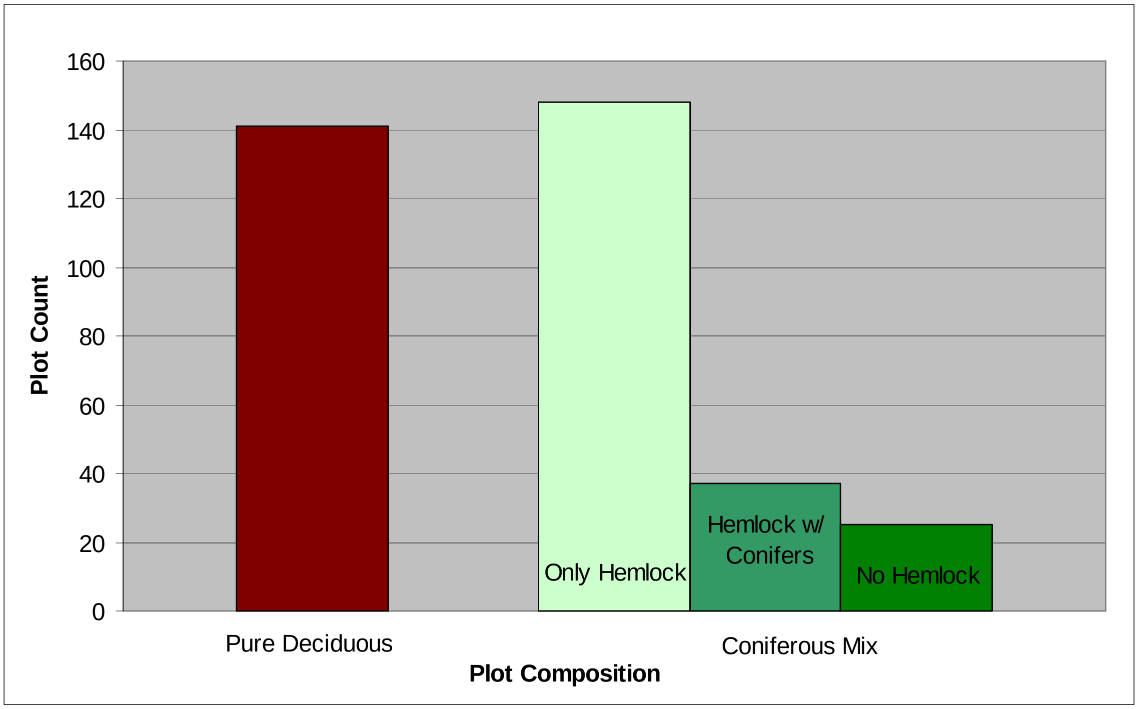

Figure 2. Reference plot composition relative to Eastern Hemlock.

CART and Cubist

Classification and Regression Trees analysis is an ideal tool for locating Eastern Hemlock. They can be used to explore relationships between variables, and to categorize and quantify a single response variable. They have been successfully used to map impervious surfaces and forest canopy (Herold 2003), forest groups and basal area (Xian 2002), and hemlock distribution in the Great Smokey Mountains (Koch, Pers. Corrs., 2005). Advantages of CART for mapping hemlock lie in its ability to accept various types of data that are non-normally distributed, and the ease of both developing and applying the rules across the study area. Additionally the ability of CART to selectively choose the most appropriate datasets for modeling allows for the inclusion of maximum range of independent variables without worry of diluting the strength of the model.

Cubist version 2.02, developed by Rulequest Research (www.cubist.com), was chosen to develop the regression tree models for locating hemlock. Cubist is a modified multivariate regression-based data mining program that creates rule-based predictive models. Rules are created from a combination of conditions associated with a linear expression. A tree is composed of a series of rules that cover the entire range of input data. Cubist will average the value of cases that are covered by multiple rules, and will not predict values outside the range of initial response variables by default. In this case the percentage of hemlock basal area in a plot is the response variable, and the Cubist model is then used to predict the percentage of hemlock per pixel across the study area.

METHODOLOGY AND PROCEDURES

There are practical considerations when choosing datasets for locating a tree species that is found throughout the eastern United States. Spectral composition, spatial extent, ground resolution, and availability are at the top of the list. Extent and availability are particular concerns of state agencies with large areas of responsibilities and limited budgets. The near infrared portion of the spectrum contains the majority of information on plants, thus eliminating simple color and color infrared imagery. At the other end of the spectrum is hyperspectral imagery that would provide more detailed information, but the combination of the remaining three concerns makes it impractical. The spatial extent of Eastern Hemlock makes airborne image acquisition a costly endeavor, and the thus multispectral satellite imagery is the most practical choice. With its moderate spatial resolution, decades of continuous global coverage and accessible price the Landsat sensor became the primary target.

Differentiating pure stands would be difficult with multispectral imagery, and Eastern Hemlock is rarely found by itself. To aid the classification additional datasets representing the habitat characteristics hemlock prefers were included. A 10m Digital Elevation Model (DEM) was used to model topography elements the correlate with environmental conditions. The cool moist climate that hemlocks prefer was characterized by heat and wetness indexes derived from the DEM. More basic measures of slope, aspect, and plane convexity created to provide a baseline for the inclusion of topographic variables.

Data Preparation

The Landsat ETM+ imagery was provided by the USDA Forest Service Northeastern Research Station's Urban Forest Research Unit. Radiometric calibration along with geometric and terrain corrections were performed at the EROS Data Center of the United States Geological Survey using established methods (Irish, 2000). Methods developed by Huang (2001) were then employed to convert the reflectance digital number to at-sensor reflectance. The high gain thermal band was converted to at sensor temperature and resampled to match the 30m resolution of the 6 reflectance bands. Finally the Landsat imagery was resampled to 10m using nearest neighbor approach, such that each original pixel was now represented by a 3 X 3 group of pixels.

A 10m DEM was used to create topographic datasets depicting slope, aspect, and plane convexity at two scales. Plot topography was calculated on a 5 X 5 window while local topography was calculated using a 51 X 51 window. To account for hemlocks' preference for moist soil a topographically derived moisture index was created (Equation 1). The index uses the relationship between the catchment area draining across a cell (A_s) and the slope ( ) at that cell (Moore 1993).

Equation 1: Wetness Index

$$Wetness = \ln(A_s/\beta)$$

The DEM also served as the basis for depicting the relative change in temperature across the study area. A modified heat load index (McClure 2002) was created to highlight the cool habitat preferred by hemlock. Heat load index is similar to potential direct incident radiation, but traditional aspect is folded about a northeast-southwest axis to better reflect the relationship between solar radiation and temperature. The heat load index was further modified to account for variations in latitude and the effect of steepness of slope on temperature. Temperature in the form of a unitless index was determined based on the following equation where L = latitude, S = slope, and A = Folded Aspect.

Equation 2 : Heat Load

$$\text{Heat Load} = 0.339 + 0.808 \bullet \text{COS(L)} \bullet \text{COS(S)} - 0.196 \bullet \text{SIN(L)} \bullet \text{SIN(S)} - 0.482 \bullet \text{COS(A)} \bullet \text{SIN(S)}$$

The percentage of hemlock in a plot was determined as the sum basal area of hemlock versus the total basal area of trees in the plot. Sampling of reference data and accuracy assessment were chosen to maximize the discrimination of the models ability to predict the amount of hemlock present versus simply predicting the amount of coniferous trees present. First the plots were stratified into the pure deciduous, or one of the three types of mixed deciduous/coniferous. Within each stratum the plots were randomly divided into five groups. This partitioning of data allowed for a 5 fold cross validation in which 80% of the data (280 plots) were available for training and the remaining 20% (70 plots) were set aside for model assessment.

Data Combinations

Four combinations of data, each adding datasets to the previous, were used to examine the effects of additional information on model accuracy (

Figure 3). A leaf-off Landsat ETM+ image served as the basis of comparison. Leaf-on and spring imagery were then added to provide information on the change in forest reflectance throughout the year. Simple topographic indices in the form of elevation, slope, aspect, and plane convexity were added next. Finally more complex topographic indices designed to characterize and emulate ecosystem characteristics affecting hemlock.

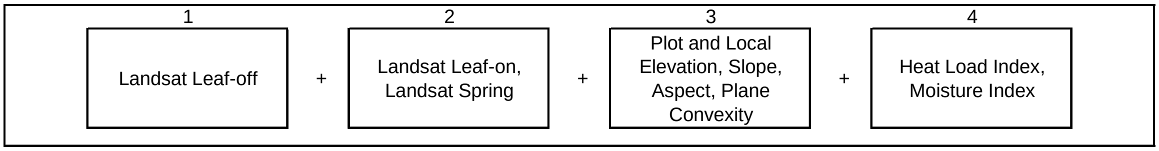

Figure 3. Overview of datasets.

DISCUSSION

Preliminary results with similar data (percent hemlock based on DBH) using Cubist's built in cross validation tool are promising overall, but difficult to interpret. The mean error for each of the four models falls in the 10% to 11% range and correlation coefficients between 0.60 and 0.68. Analysis of one output shows that a 10% average error results in the prediction of the percentage of hemlock per-pixel 49% of the time to within +_5%, and 78% of the time to within +_ 15%. Work is currently underway to determine the forests types that the models have difficulty predicting the amount of hemlock, and to examine the similarities and differences between the models created with different datasets. Even at the current indicators suggest that the methodology will provide useful information on hemlock abundance.

The goal of this work was to develop a methodology for locating Eastern Hemlock to aid forest managers in combating the Hemlock Woolly Adelgid threat. Using data readily accessible to forest managers was a primary goal, which has been successful. Upon completion of the accuracy assessment work will focus on applying the methodology to other areas and the incorporation of additional datasets. Of particular note is the digital version of the county based Soil Survey Geographic (SSURGO) data that is currently available for a majority of counties, and scheduled for national completion in 2008. Drainage characteristics and soil composition should aid the Cubist data mining process. The Forest Project has collected reference plot data in the Adirondacks, Central, and Western New York regions as well as New Hampshire and Maine that will provide a good test of the methods outlined above.

REFERENCES

Alerich, C.L., Drake, D.A. (1995) *Forest Statistics for New York*: 1980 and 1993. USDA Forest Service, Radnor, PA, pp. 191-206.

Herold, N., Koeln, G., Cunnigham, D. (2003) Mapping Impervious Surfaces and Forest Canopy using Classification and Regression Tree (CART) Analysis, ASPRS 2003 Annual Conference Proceedings, May 2003.

Huang, C., Yang, L., Homer, C., Wylie, B., Vogelman, J., DeFelice, T. (2001) At-Satellite Reflectance: A First Order Normalization of Landsat and ETM+ Images, http://landcover.usgs.gov/pdf/huang2.pdf

Irish, R. (2000) *Landsat 7 science data user's handbook*, Report 430-15-01-003-0, National Aeronautics and Space Administration, http://ltpwww.gsfc.nasa.gov/IAS/handbook/handbook_toc.html

McClure, M. S. (1990). Role of wind, birds, deer, and humans in the dispersal of Hemlock Woolly Adelgid (Homoptera: Adelgidae), *Environmental Entomology*, 19:36-43

Moore, I.D., Gessler, P.E., Nielsen, G. A., Peterson, G.A. (1993) Soil Attribute Prediction Using Terrain Analysis, *Soil Science Society of America Journal*, 57(2) pp. 443

U.S. Department of Agriculture (USDA), Forest Service. 1965. Silvics of forest trees of the United States. H. A. Fowells, comp. *U.S. Department of Agriculture, Agriculture Handbook 271*. Washington, DC. 762 p.

Xian, G., Hoppus, M., Fleming, M. (2002) Application of Decision-Tree Techniques to Forest Group and Basal Area Mapping Using Satellite Imagery and Forest Inventory Data, Pecora 15/Land Satellite Information IV/ISPRS Commission I/FIEOS 2002 Conference Proceedings, November 10-15, 2003.