IMPROVING TECHNIQUES FOR HISTORIC URBAN TREE COVER MAPPING WITH ARCHIVAL MODERATE RESOLUTION REMOTE SENSING DATA

Andrew Johnston
Center for Earth and Planetary Studies
Smithsonian National Air and Space Museum
Washington DC 20013-7012
johnstona@si.edu

ABSTRACT

Urban forests are a central element of the urban environment. Improved observations of historic tree cover dynamics are required to better understand how the urban environment changes though time. In this study, satellite remote sensing techniques were applied to observe past variability in tree cover area within the District of Columbia using highly calibrated Landsat data. Validation was performed with data from field surveys and public geospatial data on standing tree cover. Testing of alternate methodologies demonstrated that an approach utilizing support vector regression produced results with greater accuracy across the city when compared to linear spectral mixture analysis. Per-pixel uncertainty remained high using both techniques. Spectral mixture analysis overestimated tree cover in low population density areas and underestimated tree cover in the urban core, while support vector regression provided consistent accuracy across land use types. The consistent reliability allowed results from support vector regression to be used for observing tree cover changes between different land use zones. This will make it possible to identify past tree cover changes in low density residential zones within the District of Columbia. These results provide useful background information for maintenance and resource management as part of efforts to monitor and expand urban tree cover. Further development of these methods may enable their application with archival moderate resolution satellite remote sensing data for other study areas.

Keywords: Urban Forests, Land Cover/Land Use, Subpixel Land Cover

INTRODUCTION

Remote Sensing of Urban Land Cover

Several methods have been tested for mapping of proportional tree cover, defined as the ratio of tree canopy area to land surface area. This ratio is utilized in studies of the urban environment and by management authorities for setting canopy expansion goals. Aerial photography is useful for mapping urban forests because the imagery scale allows individual trees to be visible (Walton et al. 2008). Randomly distributed plots can be observed and scaled up to the entire study area (Nowak et al. 1996; Walton 2008a). Using this approach, trees are counted at points on airphotos, which can be scaled up to an entire jurisdiction, resulting in city-wide estimates of tree cover. This approach has been applied for estimating urban tree cover in cities for validation of satellite remote sensing (Walton 2008a) and for estimating urban tree cover change (Nowak and Greenfield 2012). Spectral mixture analysis (SMA) techniques are used to determine the proportional amount of land cover types within single pixels (Settle and Drake 1993). Each pixel in SMA is assumed to consist of a combination of different land cover types with distinct spectral responses. Spectral data are transformed to estimate the prevalence of those land cover types within each pixel.

Support Vector Machines

Support Vector Machines are a group of applications for classification developed from machine learning theory. When applied to remote sensing data, classification techniques aim to determine land cover type of each pixel. Data points near the margins of each class ("support vectors") are used to determine class boundaries (Vapnik 1995) with maximal distance in feature space between support vectors. Support vector regression (SVR) is an application of support vector machines for regression useful for estimating areas within pixels (Smola and Scholkopf 2004). SVR requires that spectral data be transformed into high-dimension feature space. As with support vector machine classification, SVR finds the optimal solution using data points (“support vectors”) to define the class or land cover type. The target data must be transformed into feature space before application of support vector regression or classification. The Radial Basis Function (RBF) kernel can provide higher and more stable SVM classification accuracy than the polynomial kernel, especially using a low order polynomial (Huang et al. 2002). The RBF kernel was selected for application of SVR in a previous study of urban forest cover (Walton 2008b). In implementing parameter cross-validation, values for each parameter are tested iteratively within a specified range. This procedure is run iteratively, with the parameter values altered after each run until the best accuracy is found.

District of Columbia Study Area

The District of Columbia is the capital of the United States, home to 617,000 residents (U.S. Census 2011). The District of Columbia is an urban jurisdiction containing a full range of population densities and urban land use patterns. During the study period of 1984-2004, the city experienced significant population loss and demographic change.

DATA AND METHODOLOGY

Tree Data

Tree data from the District of Columbia were acquired from the District government's GIS data online access system (D.C. Government 2007). No tree attributes such as crown size were collected. The same source also includes a data layer indicating polygon coverage of closed-canopy forests. These were supplemented with detailed data on District of Columbia street trees acquired by the nongovernmental organization Casey Trees during summer 2000, and include the location and size of every street tree in DC (Casey Trees 2003). In contrast to the DC GIS data, the Casey data include only street trees and not trees on private property.

Satellite Remote Sensing Data

Calibrated Landsat data of the District of Columbia from 2000 were processed as part of a time series of Landsat observations within the North American Forest Dynamics (NAFD) project to evaluate the role of forest disturbance in North American carbon dynamics (Goward et al. 2008). The data were developed by applying an automated mapping approach for detecting forest disturbance (Huang et al. 2009). The NAFD Landsat data have been highly processed to address specific limitations of Landsat imagery when utilized in time series analyses, including atmospheric contamination and geographic registration errors. Atmospheric effects were addressed using the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) (Masek et al. 2006). Top of atmosphere reflectance is been calculated by using published calibration coefficients. Effects of atmospheric scattering were removed using a radiative transfer model (Vermote et al. 1997) were atmospheric optical depth was measured by estimating visible reflectance from shortwave infrared reflectance. Geographic registration errors caused by topographic relief displacement were corrected by an orthorectification procedure (Gao et al. 2009).

Methodology

Subpixel proportion of urban tree cover was mapped using spectral mixture analysis (SMA) and support vector regression (SVR) to test these two techniques and identify methods for accurate urban tree cover mapping. These two techniques were selected to test the relative performance of a methodology applied in many previous studies and one that has not been widely utilized in studies of the urban environment.

To validate the Landsat-scale maps of static urban tree cover, a dataset of standing trees was compiled by combining data from field surveys of street trees and airphoto interpretation performed by the DC government. Three data sets were combined to produce a map of standing trees: 1) Casey street tree data, 2) DC GIS tree locations recorded as points, and 3) DC GIS closed canopy polygons. The dataset indicates the location of standing tree canopy corresponding to the 2000 Landsat image. The combined tree data covers the entire District of Columbia, allowing validation to be performed for the entire study area.

Casey street tree data were used to create a data layer that indicated the location and crown size of each tree. Crown radius was recorded in 5 foot intervals. Of the approximately 380,000 trees in the DC GIS point data, approximately 130,000 were duplicated in the Casey street tree data and the remaining 250,000 were non-street trees. The median tree crown radius of duplicated street trees was 15 feet (4.6 meters), and 70% had a crown radius between 10-20 feet. The median crown radius from the trees duplicated in the two data sets was selected to represent all non-street trees.

The vector data layers were combined into one binary raster with a cell size of one meter utilizing ENVI/IDL software. This raster layer was resampled to a Landsat-scale 30 meter cell size. A 30 meter raster layer was created in ENVI with map limits matching the one meter binary raster. The value for each 30 meter cell was calculated as the proportion of one meter tree pixels contained within that cell. The final city-wide raster layer indicates the proportion of each 30 meter pixel occupied by tree cover between 0-100% (Figure 7). The 30 meter cell size was selected to make it possible to compare to full resolution Landsat data.

The subpixel proportion of tree cover was estimated by applying a linear spectral mixture model to the 2000 Landsat scene. Training locations were selected at homogeneous sites of tree, grass, impervious cover high albedo, and low albedo water. Training samples were selected using visual analysis of air photography. SMA was applied as implemented in ENVI software (Exelis 2013), where SMA functions using linear unmixing algorithms (Gong et al. 1991). SMA calculations were constrained so that total land cover equaled 100%.

The LIBSVM program (Chang and Lin 2013) was used for support vector regression. Landsat reflectance data were scaled and converted for input into the LIBSVM package. Processing and data manipulation were performed using ENVI/IDL software (Exelis 2013). IDL routines were created to convert satellite data to proper format for LIBSVM. Training sites for SVR were selected by generating random points utilizing procedures within the ESRI ArcGIS software package (ESRI 2013). An initial set of points were stratified by tree cover proportion in the 2000 validation data to identify locations with the full range of tree cover values. Tree cover was then manually interpreted using aerial photography for the plots, each sized to be equivalent to a 3x3 pixel Landsat window. Images were displayed and tree crowns were manually digitized using ENVI software (Exelis 2013). A set of 62 training data points resulted. RBF kernel was selected and the LIBSVM cross-validation procedure was used to determine optimal parameters for support vector regression. After parameter validation was performed, the final SVR calculation was performed. LIBSVM was then used to calculate SVR estimates for each pixel in the 2000 Landsat data.

For both SMA and SVR, total tree cover was calculated by combining tree cover values across the city. The mean error and error standard deviation were calculated as a function of tree cover proportion. To assess the uncertainty of tree cover estimates, the standard deviation of the error was calculated. The entire error distribution was utilized to calculate the z statistic for each Landsat pixel. This resulted in an error value in standard deviation units for each pixel. Tree cover estimates were combined into 5% bins, and error was calculated for all pixels within that range. The z statistic was utilized to calculate confidence limits ($p<0.05$) for per-pixel tree cover observations. This was performed for each 5% increment to calculate confidence limits as a function of tree cover. Therefore confidence limits could be calculated for different levels of tree cover.

RESULTS

Application of SMA and SVR resulted in maps indicating tree cover proportion between 0-100%. The maps show spatially variable tree cover across the District of Columbia (Figure 1 and Figure 2). In the SMA map, densely forested parks areas such as Rock Creek Park lie in the northern part of the city. Moderate amounts of tree cover are visible in residential zones of the city. The urban core and some corridors in outer sections of the city have the lowest tree cover proportions. The application of SVR resulted in a map of proportional tree cover for the District of Columbia (Figure 2). Fine-scale differences in tree cover are more visible in the SVR results in densely developed zones of the city.

The total SMA tree cover proportion estimate for the District of Columbia was 32.9% land surface area, and RMS error of the SMA estimates was 21.0% land surface area. The total SVR tree cover estimate for the District of Columbia was 26.8% land surface area, compared to 27.0% for the validation data. Root mean square error of the SVR estimates compared to validation data is 7.7% land surface area.

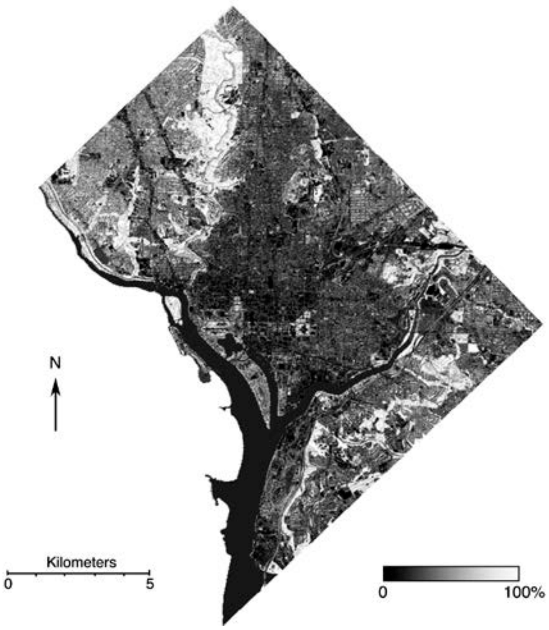

Figure 1. SMA Tree cover proportion

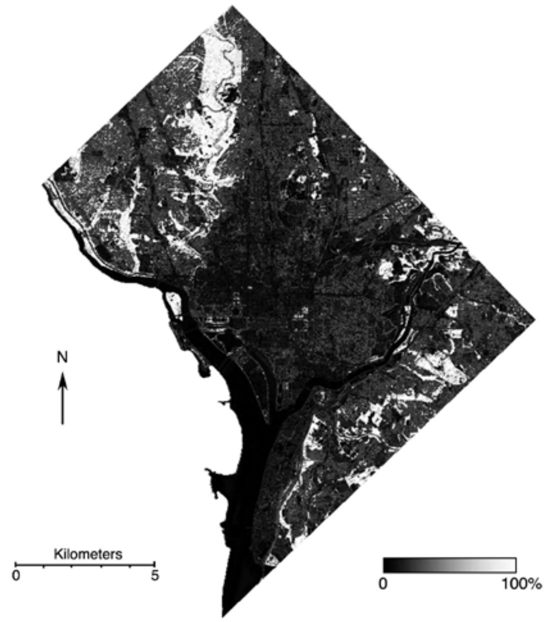

Figure 2. SVR Tree cover proportion

Mean error and confidence limits for SMA and SVR observations were plotted for each 5% bin of estimated tree cover (Figure 3 and Figure 4). SMA error across all tree cover values at the per-pixel scale was highly variable. Lower variability at high values for tree cover estimates is likely due to spectrally similar responses of closed canopy tree cover with training data. SMA overestimates tree cover, especially at intermediate tree cover values. SMA error is greatest when tree cover proportion is approximately 30%-40%. That tree cover range is typical for the low density residential areas of the District of Columbia. SVR produces more linear and consistent estimates of tree cover compared to SMA. The SVR tree cover results better match the validation data, but with a wide distribution of error still evident (Figure 4). RMS error was about one third as high for SVR compared to SMA (7.7% land area, compared to 21%). Fit to validation data was closer with results from SVR (R^2=0.91) than SMA (R^2=0.75). Tree cover estimates from SVR had higher overall accuracy than SMA. The consistent error with the SVR results may allow this technique to be used to discern tree cover differences between sections and land use types within the city.

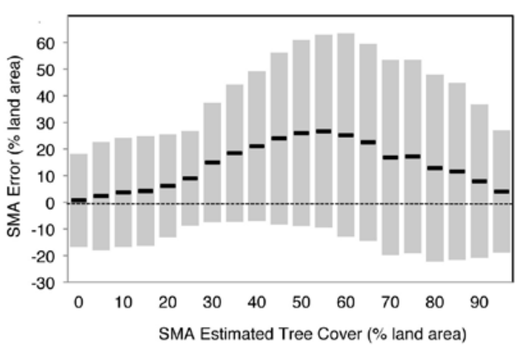

Figure 3. SMA Error

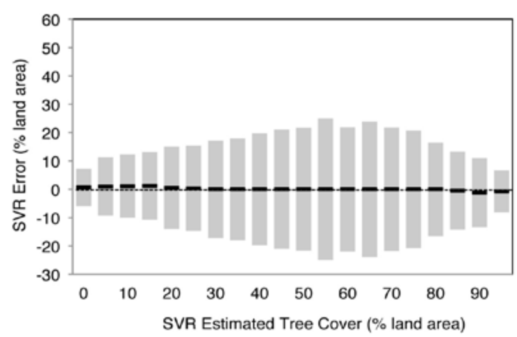

Figure 4. SVR Error

CONCLUSIONS

Support Vector Regression (SVR) produced results with lower and more spatially consistent error than Spectral Mixture Analysis (SMA) for observing urban tree cover. Results from SMA overestimated tree cover in low population density areas. In contrast, accuracy across land use types was more consistent with SVR. The consistent reliability across land use types provides an important advantage, which allows SVR results to be used for identifying tree cover changes between different regions within a city. The SMA overestimation of tree cover in the middle range of values is possibly due to confusion between tree and grass cover. Enhancements of SMA are available to provide higher accuracy for estimating land cover proportions. For instance, it may be possible to better discriminate tree and grass using nonlinear SMA techniques. SMA techniques may not be optimal to address tree/grass spectral confusion because they are trained on homogeneous tree cover. The SVR methodology addresses the problem of tree/grass confusion by incorporating spectral information from a range of tree cover proportions.

REFERENCES

Casey Trees, 2003. State of Our Trees: Status and Health of Street Trees of Washington DC.

Chang, C., & Lin, C., 2013. LIBSVM: A library for support vector machines. Available from http://www.csie.ntu.edu.tw/~cjlin/libsvm.

D.C. Government, 2007. DC GIS Clearinghouse. Database available from http://dcatlasdcgis.dc.gov/catalog

ESRI, 2013. ArcGIS documentation: Create random points.

Exelis, 2013. ENVI/IDL Documentation.

Gao, F., Masek, J., & Wolfe, R., 2009. An automated registration and orthorectification package for Landsat and Landsat-like data processing. *Journal of Applied Remote Sensing, 3.*

Gong, P., Miller, J.R., Freemantle, J., & Chen, B., 1991. Spectral decomposition of Landsat TM data for urban land-cover mapping. In, *14th Canadian Symposium on Remote Sensing*, pp. 458-461.

Goward, S.N., Masek, J., Cohen, W., Moisen, G., Collatz, G.J., Healey, S., Houghton, R.A., Huang, C., Kennedy, R., Law, B., Powell, S., Turner, D., & Wulder, M.A., 2008. Forest Disturbance and North American Carbon Flux. *EOS Trans. AGU, 89*, 105-106.

Huang, C., Davis, L.S., & Townshend, J.R., 2002. An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, *23*, 725-749.

Huang, C., Goward, S.N., Masek, J., Thomas, N., Zhu, Z., & Vogelmann, J.E., 2009. An automated approach for reconstructing recent forest disturbance using dense Landsat time series stacks. *Remote Sensing of Environment, 114*, 183-198.

Masek, J.G., Vermote, E.F., Saleous, N.E., Wolfe, R., Hall, F.G., Huemmrich, K.F., Gao, F., Kutler, J., & Lim, T.K., 2006. A Landsat surface reflectance data set for North America, 1990-2000. *IEEE Geoscience and Remote Sensing Letters, 3*, 68-72.

Nowak, D.J., & Greenfield, E.J., 2012. Tree and impervious cover change in U.S. cities. *Urban Forestry and Urban Greening*, *11*, 21-30.

Nowak, D.J., Rowntree, R.A., McPherson, E.G., Sisinni, S.M., Kerkmann, E.R., & Stevens, J.C., 1996. Measuring and analyzing urban tree cover. *Landscape and Urban Planning, 36*, 49-57.

Settle, J.J., & Drake, N.A., 1993. Linear mixing and the estimation of ground cover proportions. *International Journal of Remote Sensing, 14*, 1159-1177.

Smola, A.J., & Scholkopf, B., 2004. A tutorial on support vector regression. *Statistics and Computing, 14,* 199-222

U.S. Census, 2011. District of Columbia 2010 Census Population Totals.

Vapnik, V.N., 1995. *The Nature of Statistical Learning Theory*. New York: Springer-Verlag.

Vermote, E., Tanre, D., & Morcrette, J., 1997. Second simulation of the satellite signal in the solar spectrum, 6S: an overview. *IEEE Transactions on Geoscience and Remote Sensing, 35*, 675-689.

Walton, J.T., 2008a. Difficulties with estimating city-wide urban forest cover change from national, remotely-sensed tree canopy maps. *Urban Ecosystems*, *11*, 81-90.

Walton, J.T., 2008b. Subpixel urban land cover estimation: Comparing cubist, random forests, and support vector regression. *Photogrammetric Engineering & Remote Sensing*, *74*, 1213-1222.

Walton, J.T., Nowak, D., & Greenfield, E.J., 2008. Assessing urban forest canopy cover using airborne or satellite imagery. *Arboriculture and Urban Forestry*, *34*, 334-340.