Subpixel Urban Land Cover Estimation: Comparing Cubist, Random Forests, and Support Vector Regression

Jeffrey T. Walton

Abstract

Three machine learning subpixel estimation methods (Cubist, Random Forests, and support vector regression) were applied to estimate urban cover. Urban forest canopy cover and impervious surface cover were estimated from Landsat-7 ETM+ imagery using a higher resolution cover map resampled to 30 m as training and reference data. Three different band combinations (reflectance, tasseled cap, and both reflectance and tasseled cap plus thermal) were compared for their effectiveness with each of the methods. Thirty different training site number and size combinations were also tested. Support vector regression on the tasseled cap bands was found to be the best estimator for urban forest canopy cover, while Cubist performed best using the reflectance plus tasseled cap band combination when predicting impervious surface cover. More training data partitioned in many small training sites generally produces better estimation results.

Introduction

Describing the urban environment using remote sensing techniques has been an active area of research for many years. One popular method for characterizing urban land-cover is the V-I-S (vegetation-impervious surface-soil) model presented by Ridd (1995). The V-I-S model describes the complex urban landscape as a tripartite mixture of the fundamental components of an urban ecosystem: vegetation, impervious surfaces, and exposed soil (ignoring water surfaces). Ridd calls for "a standardized way to define these urban building blocks and to detect and map them in repetitive and consistent terms." This study compares three machine learning regression methods that can be used to map individual urban land constituents.

Mapping landscape components in urban areas using traditional hard classification techniques is impeded by the large proportion of mixed pixels. In moderate resolution imagery, such as Landsat ETM+, mixed pixels predominate because of the heterogeneous combination of landscape features that are smaller than the ground instantaneous field of view (GIFOV) of the sensor. Estimating subpixel proportions of land-cover components overcomes the difficulty of assigning a pixel to one thematic class and yields a better representation of the spatial distribution of the material of interest.

One successful method for estimating ground cover proportions is spectral mixture analysis (SMA) (Settle and Drake, 1993). SMA calculates the cover of each pixel as a proportion of several spectral endmembers assuming linear mixing of the reflectance from surfaces in the GIFOV. Wu and Murray (2003) estimated impervious surface distribution using SMA. Urban vegetation abundances have been estimated using SMA on Landsat imagery by employing a variation of the V-I-S model that uses bright surfaces, dark surfaces, and vegetation as the endmembers (Small, 2001; Small and Lu, 2006). Two limitations of linear mixing models were mentioned by Huang and Townshend (2003). First, linear spectral mixing of the land-cover reflectance is assumed. It has been shown that land-cover reflectance mix in a nonlinear fashion when multiple scattering effects from the background and canopy layers are considered (Borel and Gerstl, 1994). Ray and Murray (1996) also identified this effect when investigating the spectral characteristics of desert vegetation against the soil background, which in many ways is similar to identifying urban vegetation among the built structures in a highly developed portion of a city. The second limitation identified by Huang and Townshend is the endmembers used in SMA, such as vegetation, high albedo, and low albedo surfaces, do not correspond to specific physical land-cover components like tree canopy. An additional limitation is that land-cover constituents are often constrained so that they cannot overlap each other. In many circumstances it is desirable to permit co-occurring land covers, for example, tree canopies over impervious surfaces, by allowing the cover proportions to add to more than 100 percent. To overcome these limitations, empirical, non-parametric, machine learning techniques, such as decision tree regression (Huang and Townshend, 2003; Xu *et al*., 2005) or artificial neural networks (Liu and Wu, 2005; Carpenter *et al*., 1999), have been utilized for the problem of subpixel classification. This study will compare three machine learning subpixel estimation methods (Cubist, Random Forests, and support vector regression) applied to urban cover estimation.

Paul Smith's College, Forestry, Natural Resources, and Recreation Division, Routes 86 and 30, P.O. Box 265, Paul Smiths, NY 12970, and formerly with the Northern Research Station, USDA Forest Service, Syracuse, NY (jwalton@paulsmiths.edu).

Photogrammetric Engineering & Remote Sensing
Vol. 74, No. 10, October 2008, pp. 1213–1222.

099-1112/08/7410-1213/$3.00/0
© 2008 American Society for Photogrammetry and Remote Sensing

Subpixel Urban Cover Mapping

In other recent studies (such as Small and Lu, 2006), no distinction has been made with regard to what urban vegetation components are being mapped and also the impervious surface cover is divided into bright (high albedo) and dark portions. For the purposes of this study, two cover types will be specifically identified: urban forest canopy and urban impervious surfaces. Both cover types will be mapped at the subpixel level, meaning each resulting pixel will have a numerical value representing the cover type's fractional abundance in that pixel.

Urban Forest Canopy

Urban forest canopy, the area covered by tree and shrub canopies in an urbanized or developed area, is a fundamental measure of urban forest structure (Nowak, 1994). Urban forests include groups of trees in natural stands and individual trees that may be in a park, in a residential back yard, or along a street. Urban forest canopy has an important influence on ecological processes in urban environments (Zipperer *et al.*, 1997). Urban forest canopy has been quantified using aerial photo interpretation techniques (Rowntree, 1984; Nowak, 1996), automated classification of high-resolution digital imagery (Zhang, 2001; Myeong, 2003), and medium resolution satellite imagery (Wang, 1988; Iverson and Cook, 2000). Traditionally, the automated techniques have used "whole-pixel" classifications where each pixel is designated as either "forested" or "not forested" with some threshold being used to determine the cut-off between the two classes. Because of the heterogeneous nature and number of mixed pixels in urban areas, whole-pixel classifications tend to misrepresent the amount and spatial distribution of urban forest cover. Zhu (1994) developed a subpixel forest density map using a regression procedure from 1.1 km AVHRR multi-spectral imagery, which was later used to assess urban forest cover by Dwyer *et al.* (2000). The nonprofit group American Forests (URL: **http://www.americanforests.org/resources/urbanforests/**) has used the ERDAS Imagine® Subpixel Classifier (Applied Analysis, Inc.; URL: **http://www.discover-aai.com/software/products/IMAGINE_Subpixel_Classifier.htm**) to map and assess change of urban forest canopy. Recently, with the subpixel tree canopy layer from the National Land Cover Database (NLCD) 2001 (Homer *et al.*, 2004 and 2007) becoming available, urban forest cover assessments are being done utilizing this 30 m, nation-wide resource (unpublished data, USDA Forest Service, Syracuse, New York). The NLCD 2001 tree canopy estimate is generated using the Cubist software (Huang *et al.*, 2001b; Earth Satellite Corporation, 2002; Herold *et al.*, 2003).

Urban Impervious Surface Cover

The amount of impervious surfaces, such as building roofs, roads, parking lots, driveways, sidewalks, and paved patios, has become a primary indicator of urban watershed health (Schueler, 1994; Arnold and Gibbons, 1996; Barabec *et al.*, 2002). Mapped impervious surfaces are also a vital input for water quality modeling. Several techniques have been used to map urban impervious surfaces at the subpixel level. Artificial neural networks were used to estimate the subpixel proportion of impervious surfaces from Landsat TM imagery (Civco and Hurd, 1997; Flanagan and Civco, 2001). Impervious surfaces have also been classified from TM data by the ERDAS Imagine® Subpixel Classifier (Ji and Jensen, 1999; Flannigan and Civco, 2001). Wu and Murray (2003) estimated impervious surface distribution using spectral mixture analysis. More recently, the U.S. Geological Survey (USGS) has included a subpixel impervious surfaces layer in its NLCD 2001 product (Homer *et al.*, 2004 and 2007) that is generated using the Cubist software (Herold *et al.*, 2003; Yang *et al.*, 2003).

Subpixel Estimation Methods

Generally, the term "classification" refers to assigning a pixel to a type or thematic category. Alternatively, "regression" is the problem of finding a function that approximates the mapping from an input space into a real number based on a set of training values. For subpixel land-cover, we are interested in estimating the percent impervious or tree canopy cover as a real number between 0 and 100; therefore, regression techniques are the methods of focus. Three machine learning regression techniques (Cubist, Random Forests, and support vector regression) were compared for their effectiveness in estimating subpixel impervious surface and urban forest canopy cover.

Cubist

Cubist (Rulequest, 2006) is a rule-based regression technique developed by Quinlan (1997). Of the three regression techniques compared in this study, Cubist has the least algorithmic documentation because it is a commercial product and proprietary. However, a survey of Quinlan's earlier work and its deliberate progression (Murthy, 1998) can give insight into the probable techniques employed in Cubist. Decision tree classifiers such as ID3 (Quinlan, 1986), which chooses the branching attribute to maximize information gain in its divide-and-conquer tree generation strategy, and later C4.5 (Quinlan, 1993a), with (among other enhancements) its ability to handle continuous-valued attributes (Quinlan, 1996), are the foundations that Cubist is built upon. A method to reformulate decision trees into a set of production rules was devised by Quinlan (1987a and 1987b) and is in Cubist. With the M5 model (Quinlan, 1992), categorical decision trees were expanded to handle continuous classes by placing a multivariate linear model at each leaf. These *model trees* predict real values and are similar to regression trees, but a regression tree only contains a single value at each leaf (Witten and Frank, 2005). Model trees have been shown to be more accurate than ordinary regression (Quinlan, 1993b). An additional technique to improve prediction is to use similar training cases, or instances, to determine the value at the new location. Quinlan (1993c) proposed a method to combine instance- and model-based techniques to obtain a regression solution. Composite models combining instances and model trees have been shown to be more accurate than model trees alone (Quinlan, 1993b). It is assumed that Cubist is a composite model that combines a model tree, reformulated as rules, with the instance-based method (Rulequest, 2006). The resulting rules have linear multivariate models in their "then" statements. An option in Cubist allows the user to include instances if desired.

The USGS EROS Data Center and their cooperators have chosen to use Cubist to estimate the subpixel proportion of impervious surfaces and tree canopy for the NLCD 2001. The methodology and performance of using Cubist to estimate tree canopy cover from Landsat ETM+ imagery has been presented by Huang *et al.* (2001b), Earth Satellite Corporation (2002), and Herold *et al.* (2003). Similarly, Yang *et al.* (2003) and Herold *et al.* (2003) presented the procedure that is used by the NLCD to map impervious surfaces. The same method has been applied to map impervious surfaces in Norway using Landsat ETM+ imagery (Schuler and Kastdalen, 2005). MODIS imagery, in conjunction with training data developed from QuickBird images, was used to estimate the percent tree cover over the entire continent of Africa (Rokhmatuloh *et al.*, 2005). Currently, the Landscape Fire and Resource Management Planning Tools Project run by a consortium of U.S. Government land management agencies and other cooperators is using Cubist to estimate forest structure parameters for input into a fire simulator (LANDFIRE, 2006). Cubist has also been used in estimating house values with the aid of Landsat ETM+ derived environmental characteristics (Yu and Wu, 2006).

Random Forests

Random Forests (RF) is a classification and regression technique introduced by Breiman (2001) that uses many classification or regression trees in an ensemble. Ensemble learning methods aggregate the results from individual trees and generally produce better results (Breiman, 1996). RF, an extension of bagging (or bootstrap aggregation), uses random samples (with replacement) of the training data to generate many regression trees, which are grown without pruning, and the results of all averaged (Liaw and Wiener, 2002). In bagging, trees are grown by selecting the best split at each node using all the predictor variables. RF modifies the splitting by choosing the best split from a randomly selected subset of the predictors (Breiman, 2003; Liaw and Wiener, 2002). Selecting the predictors randomly creates more diversity among the trees and reduces their correlation (Prasad *et al.*, 2006). But because of its use of random choices, RF yields slightly different results each time it is run. The aggregation of output from many trees tends to smooth the variance between trees and gives the overall model more generalization capacity. In fact, as more trees are added, RF does not overfit, always converges, and has bounded generalization error (Breiman, 2001).

Three useful properties of RF are internal error estimates, the ability to estimate variable importance, and the capacity to handle weak explanatory variables. One of the attractive features of RF is that it can estimate error without having a set-aside testing dataset by using out-of-bag error estimates (Furlanello *et al.*, 2003; Lawrence *et al.*, 2006). The out-of-bag error estimates are created from the data that are not in the bootstrap sample used for each tree's development (Breiman, 2001). Because it is impossible with a random forest of hundreds of trees to understand the role of individual variables, the importance of a variable can be estimated by tracking how the prediction error changes as randomly permuted out-of-bag examples are applied after each tree is constructed (Breiman, 2001). Variable importance can be used to gain an understanding of the relative value of predictor variables to the solution and to potentially reduce the number of input variables. Often there is no one input or small group of inputs that strongly differentiates between classes or indicates its functional trend; this is commonly the case for remote sensing image analysis problems. RF has been shown to work well with such weak classifiers (Breiman, 2001; Lawrence *et al.*, 2006).

Random Forests software is available from the Breiman and Cutler web site as FORTRAN source (Breiman and Cutler, 2004), as an R package (Liaw and Wiener, 2002), or as a commercial product from Salford Systems (**http://www.salford-systems.com**).

RF has been used for several spatial mapping applications. Furlanello *et al.* (2003) used environmental variables to estimate the probability of tick presence. Similarly, Prasad *et al.* (2006), comparing RF with other regression techniques, predicted and mapped tree species range based on biophysical parameters. Bunn *et al.* (2005) geographically predicted plant growth (gross photosynthetic activity) from AVHRR derived data and surface weather observations. The RF classifier has also been employed in remote sensing applications; for example, Pal (2005) classified agricultural crops using Landsat ETM+ imagery. The ability of RF to handle high dimensional and weak input data has made it attractive for hyperspectral remote sensing classification. Lawrence *et al.* (2006) mapped invasive species and Ham *et al.* (2005) mapped vegetation and land-cover from hyperspectral imagery.

Support Vector Regression

Support vector machines (SVMs) are a family of classification and regression techniques based on statistical learning theory (Vapnik, 2000; Cristianini and Shawe-Taylor, 2000; Burges, 1998). Rather than dealing with the statistical properties of classes like traditional classifiers (such as mean and variance), SVMs focus classification decisions on the boundary between classes. Using a kernel function, SVMs map the input space (independent variables) to a higher dimensional space where complex nonlinear decision boundaries between classes become linear. In this mapped, high-dimensional space an optimal linear separator is found that maximizes the margin between classes (Russell and Norvig, 2003). By maximizing the margin, the solution is generalized and overfitting is reduced. There are three important properties of SVMs (Vapnik, 2000; Cristianini and Shawe-Taylor, 2000; Burges, 1998; Russell and Norvig, 2003): (a) Because the function of the margin is a convex quadratic form, a single, margin-maximizing solution can always be found; (b) Since the data appear in the margin function as dot products, any number of dot product kernels that better reproduce the complex (possibly nonlinear) decision boundary can be substituted. Popular kernel functions include linear, polynomial, radial basis, and sigmoid; and (c) Only data points closest to the separator play into the classification decision. These are the support vectors and may only be a small fraction of the training data limited to the critical area where two classes meet or overlap; this is the idea of sparseness. Due to these inherent properties, overfitting is reduced and a manageable level of complexity is maintained as the dimensionality of the data space increases (Burges, 1998). Very large input data spaces and large training datasets (100s of thousands of training points) can be handled (Burges, 1998).

For the problem of subpixel estimation, support vector regression (SVR) is used to define a real-valued output function given the independent input variables. In addition to the properties mentioned above for SVM classification, SVR applies the concept of an ε-insensitive loss function that ignores point errors within a distance of ε from the true value by weighting them with zero. The ε-insensitive loss function is convex quadratic and always has a minimum (Cristianini and Shawe-Taylor, 2000; Smola and Schölkopf, 2004). The resulting regression is "smooth and has a sparse representation" (Vapnik, 2000). Sparseness, the idea of representing the solution through a small subset of training points, indicates that the support vectors contain all the required information to define the function and results in "extremely efficient algorithms" (Cristianini and Shawe-Taylor, 2000). In an example comparison with other advanced regression techniques, Vapnik (2000) demonstrated a superior performance of a polynomial support vector regression machine over bagging and boosting regression trees. For certain choices of control parameters and kernel functions, SVR reduces to other known types of regression such as standard least squares or kernel ridge regression (Cristianini and Shawe-Taylor, 2000). In addition, SVR has been shown to be closely related to linear spectral mixture models, and when a linear kernel is used, both produce equivalent models (Brown *et al.*, 1999).

In remote sensing, SVMs have been applied to both thematic classification (Huang *et al.*, 2002a; Foody and Mathur, 2004; Pal, 2005) and estimation of continuous, real-valued properties (Brown *et al.*, 1999; Srivastava *et al.*, 2005; Camps-Valls *et al.*, 2006).

Methods

The general procedure used here for applying one of these estimation methods is: (a) A subset of the input imagery and reference data are used to train the estimator and develop a prediction model; (b) The model is then applied to the remaining input data to predict the target value across the whole image; and (c) Several accuracy metrics are used to assess the quality of the resulting image.

Datasets

The city of Syracuse, New York, is the study area for this urban land-cover analysis. With regard to urban forest cover, Syracuse is a fairly typical eastern U.S. city consisting of mostly deciduous trees and has approximately 26 percent forest cover. Impervious surface distribution is also typical with highly developed central business district and commercial corridors surrounded by less densely impervious residential areas. Input predictive data in the form of Landsat imagery and a complete coverage of reference data were available for the study area.

Input imagery data were obtained from the MRLC 2001 imagery archive. MRLC (Multi-Resolution Land Characteristics Consortium, URL: **http://www.mrlc.gov**) is a group of U.S. Federal agencies that joined together to purchase Landsat imagery of the entire United States and coordinate development of nation-wide land-cover products like the NLCD. Three dates of imagery are generally available for each Landsat scene corresponding to early growing season, peak growing season, and late growing season. Yang *et al*. (2001) describes the scene selection process, and Huang *et al*. (2001a) details the at-satellite reflectance corrections applied to the MRLC 2001 imagery. The dates of Landsat-7 ETM+ scenes used for this analysis were: early growing season on 28 April 2001 (path: 15, row: 30), peak growing season on 03 July 1999 (path: 16, row: 30), and late growing season on 07 September 2000 (path: 16, row: 30). For each scene, ten bands were available: six reflectance (Landsat bands 1 through 5, and 7), one high-gain thermal (Landsat band 9; resampled to 30 m), and three tasseled cap (brightness, greenness, wetness (Huang *et al*., 2002b)). Input data were limited to imagery bands readily available from MRLC. The spatial resolution of the Landsat images was 30 m. The total dataset consisted of 123,454 pixels covering approximately 111 km^2.

The reference data were in the form of a raster land-cover map of the entire study area developed from 0.61 m (2-foot) resolution, color infrared, aerial digital images acquired on 13 July 1999 (Myeong *et al*., 2003). The overall accuracy of the reference data is 82 percent (Myeong *et al*., 2003). This high-resolution cover map consisting of five classes (tree, grass, impervious surface, bare soil, and water) was summarized based on the Landsat pixels to generate two 30 m resolution reference layers (impervious surface and urban forest canopy) by assigning to each 30 m pixel the percent of the desired reference class in that pixel.

Two Experiments

Landsat input data and reference data were partitioned into training datasets that were used to develop the regression models. A training site random sampling scheme was designed to represent options that would be available for a typical land-cover estimation project. A range of training site sizes and number of training sites were created to test each estimator's robustness to training set size. Fewer and smaller training sites would be advantageous since less effort would be required to create the reference classification. Training site size, number of sites, and total number of training pixels for each classification run are summarized in Table 1. Data values for each training site size/number combination were selected by placing the non-overlapping sites randomly within the study area. Three different scenarios of input bands were evaluated: "all" (3 scenes with 10 bands each), "refl" (3 scenes with 6 reflectance bands each), and "tc" (3 scenes with 3 tasseled cap bands each). The same training subsets were used for each estimator and band scenario.

A second experiment was designed to directly compare the estimators using a single training site size of 334 m and 10 training sites arranged in 10 replications of non-overlapping, random selections of training pixels. The two-factor, 3×3 design consisted of all combinations of method (Cubist, Random Forests, and support vector regression) and input bands (all, refl, and tc). Training sets between each combination of factors were different, but could contain overlapping groups of pixels. The experiment was repeated for estimations of both urban forest cover and impervious surface cover. The results of this experiment were analyzed graphically to illustrate differences between the estimation methods and band combinations.

Software

Each of the three machine-learning regression estimators had various input parameters necessary for operation.

Cubist version 1.12 was used in this study. Cubist has the ability to generate composite models (including nearest neighbor instances) and committee models (groups of models that attempt to correct errors in previous committee member models). The default operation is for Cubist to create only rule-based models, and because preliminary results indicated that composite models did not improve accuracy, rule-only models were used for this study. Options also exist to limit the maximum number of rules (default: unlimited), the minimum number of cases a rule should handle (default: 1 percent), and an extrapolation percentage indicating the extent of prediction outside the range of the input data (default: 10 percent). Based on previous experience with the software, all user-specified options were left at their default values except the extrapolation percentage was set to zero for all Cubist runs.

TABLE 1. THE SIZE OF TRAINING SITES BASED ON THE PERCENT SAMPLE OF THE 123,454 PIXELS. SIZE IS THE LENGTH OF ONE SIDE OF A SQUARE TRAINING SITE. THE TOTAL NUMBER OF PIXELS IN THE TRAINING SET (SHOWN IN PARENTHESIS) VARIES SLIGHTLY WITHIN A COLUMN DUE TO THE RANDOM LOCATION OF THE TRAINING SITES WITH RESPECT TO THE FIXED 30 M PIXEL SPACING

Number of	Percent Sample of Total Image				
Training Sites	1%	2%	4%	8%	16%
5	472 m (1202)	667 m (2532)	943 m (4899)	1334 m (9991)	1886 m (19719)
10	334 m (1232)	472 m (2403)	667 m (5042)	943 m (9766)	1334 m (19904)
20	236 m (1266)	334 m (2477)	472 m (4933)	667 m (9866)	943 m (19581)
40	167 m (1226)	236 m (2502)	334 m (4969)	472 m (9857)	667 m (19671)
80	118 m (1246)	167 m (2503)	236 m (4978)	334 m (9992)	472 m (19848)
160	84 m (1259)	118 m (2482)	167 m (4990)	236 m (9916)	334 m (19937)

The Random Forests implementation used in this study is the “randomForest” library (Liaw and Wiener, 2002) in the free software environment for statistical computing, R (R Development Core Team, 2006). It is based on the original FORTRAN code developed by Breiman and Cutler (2004). Only two user supplied parameters are required to operate RF: the number of trees in the forest, ntree, and the number of variables randomly sampled at each split, mtry. Regression performance has been said to be insensitive to the choice of ntree and mtry (Breiman, 2001; Lawrence *et al.*, 2006) and attempts at tuning RF indicated the same. Values of ntree = 500 and mtry = 7 were used.

The SVR implementation used in this study is included in the e1071 library (Dimitriadou *et al.*, 2006) of the R statistical computing environment and is based on LIBSVM (Chang and Lin, 2006). For this study, the radial basis function was used for the kernel. Two general SVM parameters ε and cost, along with the radial basis function parameter gamma, must be specified; ε specifies the size of the insensitive region inside of which errors are ignored, and cost is the penalty multiplier applied for constraint violations. The R library e1071 provides a useful wrapper function for tuning the SVM model parameters. After running the tuning function several times with a representative input dataset, these values were selected: $\varepsilon = 0.01$, cost = 10, and gamma = 0.01.

Comparison Metrics

Each regression method generates its own accuracy measures, but for consistency in comparison of the techniques, two standard accuracy metrics were calculated. Error was calculated by subtracting the pixel’s reference value from the estimated value. Thus, a negative error indicates that the model underestimated the value. From this error, a pair of difference measures were computed: mean absolute error (MAE) and root mean square error (RMSE). MAE is the average absolute difference of the estimated value from the reference. RMSE is the square root of the mean squared error. MAE and RMSE are highly, positively correlated and roughly report the same quality of the model. However, large errors will tend to have a larger effect in RMSE because the error term is squared. Therefore, considering both metrics together may yield some insight into the relative distribution of the errors.

Execution time was also recorded for each estimator to judge the relative time required for different training set sizes and to compare estimators. For this purpose, all techniques were run on a dual processor, 3.1 GHZ Intel Pentium IV computer with Microsoft Windows® 2000 and execution times were recorded in seconds. Because of memory limitations of R under Windows® that limited the size of training sets for RF, the largest training sets could not be run for RF on this computer and were omitted.

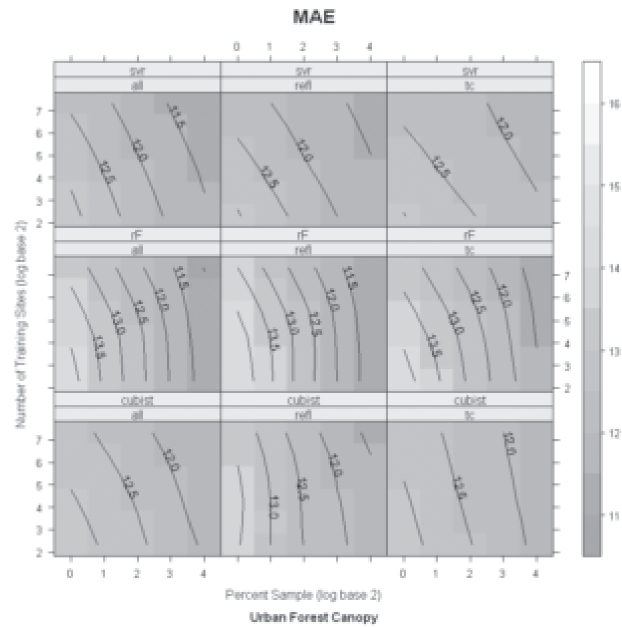

Figure 1. Isoline plot of MAE of urban forest canopy for each method (svr = support vector regression, rF = randomForests, and cubist = Cubist) and band combination (all = reflectance, thermal and tasseled cap; refl = reflectance only; and tc = tasseled cap only) by the number of training sites and percent of image used for training. The scales on each panel are log(base 2) so percent sample ranges from 1 percent to 16 percent and number of training sites ranges from 5 to 160. A color version of this figure is available at the ASPRS website: **www.asprs.org**.

Results

MAE and RMSE for both urban forest canopy and impervious surface cover estimations show a decrease in error (or increase in accuracy) as the number of training sites increases and the percent of the image sampled for training increases for all methods and band combinations (at the upper right in each panel of Figure 1 and Figure 2). As expected, MAE and RMSE values are highly correlated and do not show any obvious anomalies; therefore, only the MAE figures are presented. The total execution time to build the estimation model and apply it to the entire study area increases exponentially for both SVR and RF methods as the amount of training data increases (Figure 3); however, the increase in run time for Cubist was very small and appears to only increase in an arithmetic progression for the range of training input sizes used. Cubist run times were those reported by the software and were much faster than the other methods. In general, the RF execution times were the slowest especially for the larger training set sizes. Execution times for urban forest cover estimation were slightly less than those for impervious surface cover estimation, but followed the same pattern (not shown).

In the second experiment, the interaction plots (Figure 4 and Figure 5) show which method or band combination is different than another by plotting the means of the ten replications for each factor combination. Estimation method generally has more effect for urban forest canopy estimation, and the band combination used in the model is more effective for impervious surface cover estimation.

Discussion

Although the accuracy for each of the three methods are roughly the same and accuracy for all methods increases as more of the image is used for training, there are some subtleties that indicate differences between the methods. Examining the isolines for urban forest canopy MAE (Figure 1) show nearly vertical lines for RF and Cubist, and more angled lines (from upper left to lower right) for SVR. The nearly vertical isolines for RF and Cubist indicate that the spatial allocation of the training blocks (i.e., fewer large versus many small) is not important to decrease error, but rather just increasing the percent sample will tend to decrease MAE. For example, a 16 percent sample arranged as 5, 1886 m sites yields results roughly equivalent to 160, 334

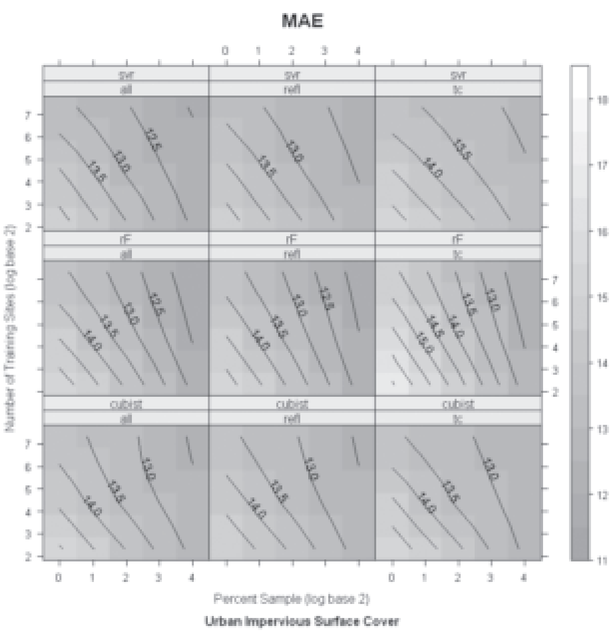

Figure 2. Isoline plot of MAE of impervious surface cover for each method (svr = support vector regression, rF = randomForests, and cubist = Cubist) and band combination (all = reflectance, thermal and tasseled cap; refl = reflectance only; and tc = tasseled cap only) by the number of training sites and percent of image used for training. The scales on each panel are log(base 2) so percent sample ranges from 1 percent to 16 percent and number of training sites ranges from 5 to 160. A color version of this figure is available at the ASPRS website: **www.asprs.org**.

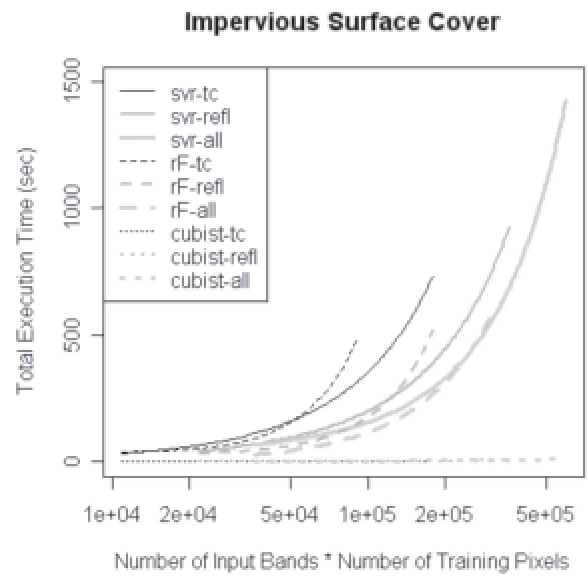

Figure 3. Total execution time to build the estimation model and apply it to the entire study area. Execution time was smoothed using a 2nd order polynomial fit. The longest RF runs were omitted due to processing limitations. A color version of this figure is available at the ASPRS website: **www.asprs.org**.

m training sites. Conversely, for SVR and all of the impervious surface estimations (Figure 2), having the data arranged in more sites yields better results. Because adjacent pixels tend to be correlated, a more representative sample may be created as more sites are used for the same percent sample. One explanation for the difference between urban forest canopy and impervious surface estimation may be due to how evenly distributed the cover type is throughout the study area. In this study, trees tend to be clustered in several large tracts and heavily canopied neighborhoods while impervious surfaces are more evenly distributed.

There were very substantial differences in execution time for the three models (Figure 3). Execution times for SVR and RF increased rapidly as more training data was added. The execution time for Cubist increased more slowly; the size of the datasets used in this experiment were easily handled by Cubist. One factor that may account for some of the noted differences is the model's implementation. Cubist is an optimized, stand alone program; whereas, the other methods are implemented inside the R environment where there may be some overhead penalty. Execution time is important when considering if the addition of more training data is worth the time investment to generate and process that data. Based on the asymptotic curve (Figure 6), the time required to reduce error by 1 percent increases dramatically as more training data are used. Additionally, error will only decrease to some limiting value. For projects with very large data sets, Cubist, by far, offers the best performance.

For urban forest cover estimation, the tasseled cap band combination (tc) is better than the reflectance bands alone (refl) and all bands (all) (Figure 4) for all three estimators. The SVR method yields better results than Cubist, which in turn is better than RF for all band comparisons (Figure 4a). And in general, SVR is more accurate (based on both metrics) than the other two methods for every band combination. For urban forest canopy, the three tasseled cap bands (brightness, greenness, and wetness) contain all the necessary information to produce accurate maps and, in fact, including other bands will result in poorer accuracy.

In contrast, for impervious surface cover estimation (Figure 5), the tc bands produce the worst results with refl and all band combinations generating about equal accuracy. The tasseled cap bands appear to lack enough information for the three methods to estimate impervious surface cover better than when using the reflectance bands alone. When comparing methods using all input bands, Cubist consistently produced better results and SVR the worst (by comparing MAE and RMSE). RF produced mixed results between the two accuracy metrics, perhaps indicating that distribution of the errors in the RF output contains fewer large errors than the results from SVR. Using only the reflectance bands, all three methods produce about the same quality of results.

The RF estimator has been touted as having the capacity to handle weak explanatory input data (Breiman, 2001; Ham *et al.*, 2005; Lawrence *et al.*, 2006). For this study, when estimating urban forest cover, RF results were the best when using the minimalist tc band set and actually were degraded when additional, weakly descriptive variables were added like the entire reflectance band set (Figure 4). A concurrence of this can also be noted in the impervious surface estimations where slightly poorer results are shown when adding bands to the reflectance set to yield the all combination (Figure 5). In this comparison, RF did not seem to handle these situations any better than the other estimators.

It should be noted that the differences represent small errors, usually less than 1 percent cover for the three methods at a given band combination (Figure 4 and Figure 5). The error

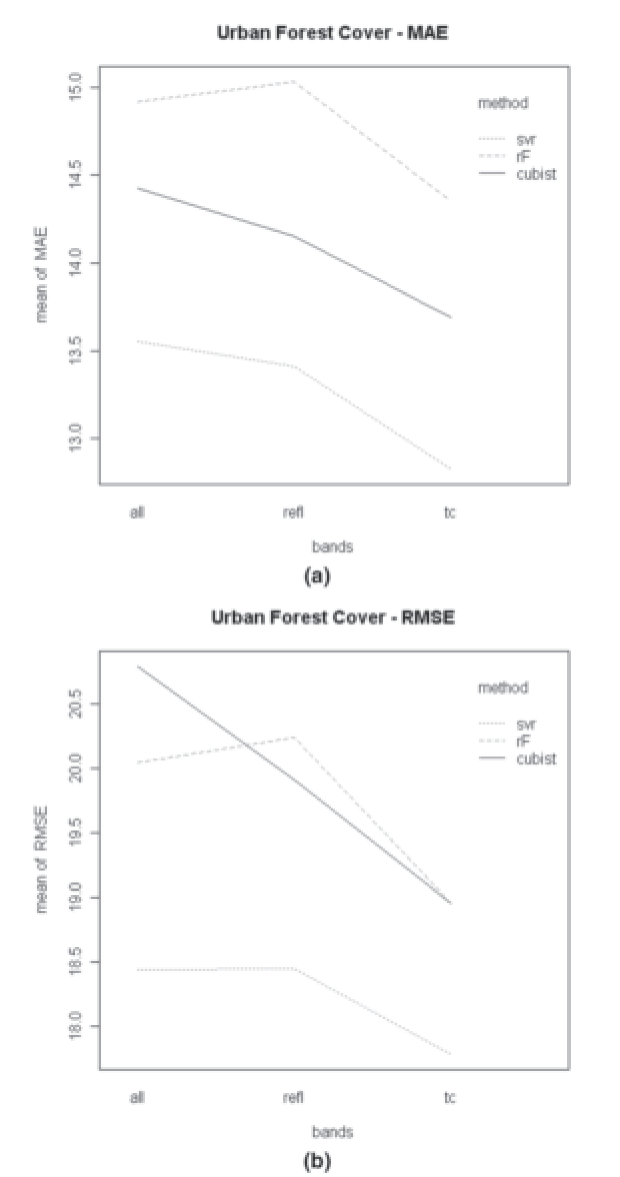

Figure 4. Urban forest cover accuracy interactions between estimation method and band combinations. Mean values of (a) MAE and (b) RMSE in percent cover for the 10 replications are plotted for each method/band combination. A color version of this figure is available at the ASPRS website: **www.asprs.org**.

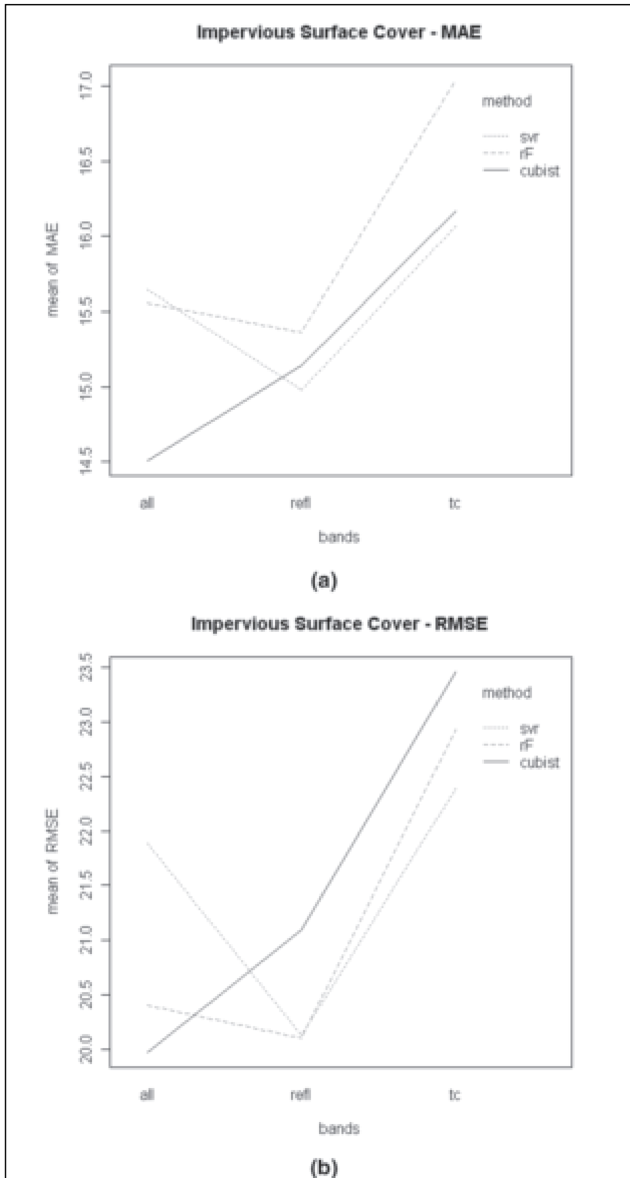

Figure 5. Impervious surface cover accuracy interactions between estimation method and band combinations. Mean values of (a) MAE and (b) RMSE in percent cover for the 10 replications are plotted for each method/band combination. A color version of this figure is available at the ASPRS website: **www.asprs.org**.

range between band combinations for a particular estimation method is also usually less than 1 percent for urban forest cover (Figure 4), but for impervious surface estimation, the range increases to approximately 1.5 percent for MAE and up to about 3 percent for RMSE. From a practical standpoint for predicting impervious surface and urban forest cover, results with differences of 1 percent or less are essentially the same. For larger errors, the results should be examined spatially in map form because certain types of input pixels may not be represented in the training set and yield errors that are obvious on a map. Practical importance levels will vary depending on the parameter being estimated.

All three models were fairly easy to use with Cubist the easiest; however, there is a relatively steep learning curve for R, which is not relevant for the specific estimators only their R implementation. Other features like the tuning functions and data manipulation/analysis capabilities in R make SVR and RF attractive. Additionally, SVR and RF are open source and have well-documented algorithms, which is important for those who want to know what is happening in the "black box." An added benefit of R is that it is available at no cost, although a routine similar to Cubist called M5Rules (Holmes *et al.*, 1999) is freely available through the Weka project (**http://www.cs.waikato.ac.nz/~ml/**; Witten and Frank, 2005).

The results of this study should be applicable to estimating urban forest and impervious surface cover for cities in many regions of the U.S., particularly the Northeast

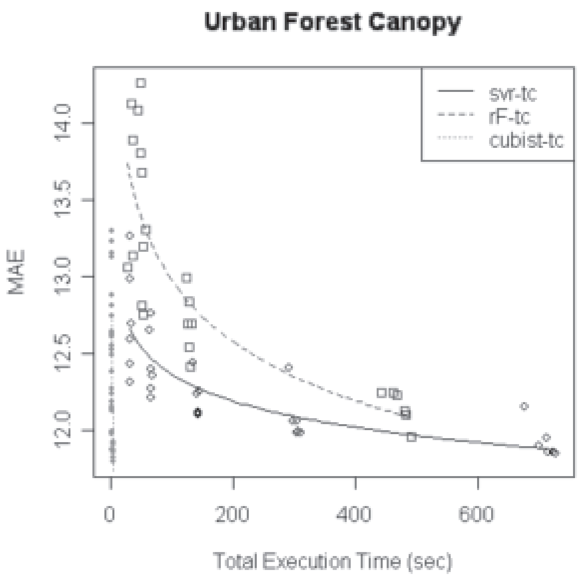

Figure 6. MAE decreases asymptotically as training data size (and therefore run time) increases. The lines are power functions fit through the raw data points. A color version of this figure is available at the ASPRS website: **www.asprs.org**.

with ecological and development characteristics similar to Syracuse. This method may have difficulty where urban vegetation is a very small fraction of the cover and dominated by impervious surfaces, where taller buildings cast large shadows over the landscape, and where impervious surfaces mix with high albedo bare soil or gravel-like surfaces, such as in the Southwestern U.S.

In this study, only readily available MRLC imagery was used as the input data. Inclusion of ancillary data in the prediction model, such as distance-to-roads, road density, and boundaries of known natural preserves, may greatly improve estimation accuracies.

Conclusions

Cubist, Random Forests, and support vector regression all seem to handle relatively well the nonlinear mixing of urban land covers and the prediction of materials of low abundance (such as vegetation in intensely developed downtown areas). These machine learning estimation techniques were used to map individual, real-world, urban land-cover constituents without the limitations of other methods. SVR using three dates of tasseled cap bands predicts urban forest canopy cover the best. Cubist using all bands (reflectance, thermal, and tasseled cap) would be the best choice for predicting urban impervious surface cover. In terms of execution time, Cubist was far superior to the other methods. Optimized implementations of RF and SVR may improve performance significantly.

Acknowledgments

Several people provided useful comments and reviews of this manuscript, including: Gregory Bacon, Andrew Lister, David Nowak, Rachel Riemann, and John Stanovick of the USDA Forest Service's Northern Research Station, and Stephen Stehman of SUNY-College of Environmental Science and Forestry. Funding for this project was provided, in part, by the USDA Forest Service's RPA Assessment Staff and State & Private Forestry's Urban and Community Forestry Program. The use of trade, firm, or corporation names in this article is for the information and convenience of the reader. Such use does not constitute an official endorsement or approval by the United States Department of Agriculture Forest Service of any product or service to the exclusion of others that may be suitable.

References

Arnold, C.L., and C.J. Gibbons, 1996. Impervious surface coverage: The emergence of a key environmental indicator, *American Planning Association Journal*, 62(2):243–258.

Barabec, E., S. Schulte, and P.L. Richards, 2002. Impervious surfaces and water quality: A review of current literature and its implications for watershed planning, *Journal of Planning Literature*, 16(4):499–514.

Borel, C.C., and S.A.W. Gerstl, 1994. Nonlinear spectral mixing models for vegetative soil surfaces, *Remote Sensing of Environment*, 47:403–416.

Breiman, L., 1996. Bagging Predictors, *Machine Learning*, 24:123–140.

Breiman, L., 2001. Random Forests, *Machine Learning*, 45:5–32.

Breiman, L., 2003. Manual – Setting Up, Using, and Understanding Random Forests, v 4.0, URL: *ftp://ftp.stat.berkeley.edu/pub/users/breiman/Using_random_forests_V3.0.pdf* (last date accessed: 12 June 2008).

Breiman, L., and A. Cutler, 2004. Random Forests, URL: *http://www.stat.berkeley.edu/users/breiman/RandomForests/* (last date accessed: 12 June 2008).

Bunn, A.G., S.J. Goetz, and G.J. Fiske, 2005. Observed and predicted responses of plant growth to climate across Canada, *Geophysical Research Letters*, 32:L16710, doi:10.1029/2005GL023646.

Brown, M., S.R. Gunn, and H.G. Lewis, 1999. Support vector machines for optimal classification and spectral unmixing, *Ecological Modelling*, 120:167–179.

Burges, C.J.C., 1998. A tutorial on support vector machines for pattern recognition, *Data Mining and Knowledge Discovery*, 2:121–167.

Camps-Valls, G., L. Bruzzone, J.L. Rojo-Alvarez, and F. Melgani, 2006. Robust support vector regression of biophysical variable estimation from remotely sensed images, *IEEE Geoscience and Remote Sensing Letters*, 3(3):339–343.

Carpenter, G.A., S. Gopal, S. Macomber, S. Martens, and C.E. Woodcock, 1999. A neural network method for mixture estimation for vegetation mapping, *Remote Sensing of Environment*, 70:138–152.

Chang, C.-C., and C.-J. Lin, 2006. LIBSVM: A library for Support Vector Machines, URL: *http://www.csie.ntu.edu.tw/~cjlin/libsvm/* (last date accessed: 12 June 2008).

Cristianini, N., and J. Shawe-Taylor, 2000. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, Cambridge University Press, Cambridge, United Kingdom, 189 p.

Civco, D.L., and J.D. Hurd, 1997. Impervious surface mapping for the State of Connecticut, *Proceedings of the 1997 ASPRS/ACSM Annual Convention*, Seattle, Washington, (American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland), 12 p., URL: *http://nemo.uconn.edu/tools/impervious_surfaces/pdfs/Civco_Hurd_1997.pdf* (last date accessed: 12 June 2008).

Dimitriadou, E., K. Hornik, F. Leisch, D. Meyer, and A. Weingessel, 2006. e1071: Misc Functions of the Department of Statistics (e1071), TU Wien. R package version 1.5–13, URL: *http://cran.r-project.org/web/packages/e1071/index.html* (last date accessed: 12 June 2008).

Dwyer, J.F., D.J. Nowak, M.H. Noble, and S.M. Sisinni, 2000. Connecting people with ecosystems in the 21st century: An assessment of our nation's urban forests. General Technical

Report PNW-GTR-490, Portland, Oregon, U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 483 p. URL: *http://treesearch.fs.fed.us/pubs/12517* (last date accessed: 12 June 2008).

Earth Satellite Corporation, 2002. *Mapping Percent Canopy for the National Land Cover Database Mapping Zones 60 and 16: Demonstration Project 23737.001.38*, U.S. Geologic Survey Task Order 010112C0012, EROS Data Center, Sioux Falls, South Dakota, 14 p.

Flanagan, M., and D.L. Civco, 2001. Subpixel impervious surface mapping, *Proceedings of the 2001 ASPRS Annual Convention*, St. Louis, Missouri (American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland), 13 p. URL: *http://nemo.uconn.edu/tools/impervious_surfaces/pdfs/Flanagan_Civco_2001b.pdf* (last date accessed: 12 June 2008).

Foody, G.M., and A. Mathur, 2004. A relative evaluation of multiclass image classification by support vector machines, *IEEE Transactions on Geoscience and Remote Sensing*, 42(6):1335–1343.

Furlanello, C., M. Neteler, S. Merler, S. Menegon, S. Fontanari, A. Donini, A. Rizzoli, and C. Chemini, 2003. GIS and the random forests predictor: Integration in R for tick-borne disease risk assessment, *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*, Vienna, Austria, URL: *http://www.ci.tuwien.ac.at/Conferences/DSC-2003/Proceedings/FurlanelloEtAl.pdf* (last date accessed: 12 June 2008).

Ham, J., Y. Chen, M.M. Crawford, and J. Ghosh, 2005. Investigation of the random forest framework for classification of hyperspectral data, *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):492–501.

Herold, N.D., G. Koeln, and D. Cunnigham, 2003. Mapping impervious surfaces and forest canopy using classification and regression tree (CART) analysis, *Proceedings of the ASPRS 2003 Annual Conference*, Anchorage, Alaska (American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland), URL: *http://nemo.uconn.edu/tools/impervious_surfaces/pdfs/Herold_etal_2003.pdf* (last date accessed: 12 June 2008).

Holmes, G., M. Hall, and E. Frank, 1999. Generating rule sets from model trees, *Australian Joint Conference on Artificial Intelligence*, URL: *http://citeseer.ist.psu.edu/holmes99generating.html* (last date accessed: 12 June 2008).

Homer, C., C. Huang, L. Yang, B. Wylie, and M. Coan, 2004. Development of a 2001 National Land Cover Database for the United States, *Photogrammetric Engineering & Remote Sensing*, 70(7):829–840, URL: *http://www.mrlc.gov/references.php* (last date accessed: 12 June 2008).

Homer, C., J. Dewitz, J. Fry, M. Coan, N. Hossain, C. Larson, N. Herold, A. McKerrow, J.N. VanDriel and J. Wickham, 2007. Completion of the 2001 National Land Cover Database for the Coterminous United States, *Photogrammetric Engineering & Remote Sensing*, 73(4):337–341, URL: *http://www.mrlc.gov/references.php* (last accessed: 12 June 2008).

Huang, C., L. Yang, C. Homer, B. Wylie, J. Volgeman, and T. DeFelice, 2001a. At-satellite reflectance: A first order normalization of Landsat 7 ETM+ images, URL: *http://landcover.usgs.gov/pdf/huang2.pdf* (last date accessed: 12 June 2008).

Huang, C., L. Yang, B. Wylie, and C. Homer, 2001b. A strategy for estimating tree canopy density using Landsat 7 ETM+ and high-resolution images over large areas, *Third International Conference on Geospatial Information in Agriculture and Forestry*, Denver, Colorado, URL: *http://landcover.usgs.gov/pdf/canopy_density.pdf* (last date accessed: 12 June 2008).

Huang, C., L.S. Davis, and J.R.G. Townshend, 2002a. An assessment of support vector machines for land cover classification, *International Journal of Remote Sensing*, 23(4):725–749.

Huang, C., B. Wylie, L. Yang, C. Homer, and G. Zylstra, 2002b. Derivation of a tasseled cap transformation based on Landsat 7 at-satellite reflectance, *International Journal of Remote Sensing*, 23(8):1741–1748, URL: *http://landcover.usgs.gov/pdf/tasseled.pdf* (last date accessed: 12 June 2008).

Huang, C., and J.R.G. Townshend, 2003. A stepwise regression tree for nonlinear approximation: Applications to estimating subpixel land cover, *International Journal of Remote Sensing*, 24(1):75–90.

Iverson, L.R., and E.A. Cook, 2000. Urban forest cover of the Chicago region and its relation to household density and income, *Urban Ecosystems*, 4:105–124.

Ji, M., and J.R. Jensen, 1999. Effectiveness of subpixel analysis detecting and quantifying urban imperviousness from Landsat Thematic Mapper imagery, *Geocarto International*, 14(4):33–41.

LANDFIRE, 2006. LANDFIRE prototype products: Vegetation structure, U.S. Department of Agriculture, Forest Service, Fire Lab, URL: *http://www.landfire.gov/RapidRefreshProductDescription9.php* (last date accessed: 12 June 2008).

Lawrence, R.L., S.D. Wood, and R.L. Sheley, 2006. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForests), *Remote Sensing of Environment*, 100:356–362.

Liaw, A., and M. Wiener, 2002. Classification and regression by RandomForests, *R News*, 2(3):18–22. URL: *http://CRAN.R-project.org/doc/Rnews/* (last date accessed: 12 June 2008).

Liu, W., and E.Y. Wu, 2005. Comparison of non-linear mixture models: Sub-pixel classification, *Remote Sensing of Environment*, 94:145–154.

Murthy, S.K., 1998. Automatic construction of decision trees from data: A multi-disciplinary survey, *Data Mining and Knowledge Discovery*, 2:345–389.

Myeong, S., P.F. Hopkins, and D.J. Nowak, 2003. Urban cover mapping using digital, high-spatial resolution aerial imagery, *Urban Ecosystems*, 5:243–256.

Nowak, D.J., 1994. Understanding the structure of urban forests, *Journal of Forestry*, 92(10):42-46, URL: *http://www.fs.fed.us/ne/syracuse/Pubs/Downloads/94_DN_UnderstandingStr.pdf* (last date accessed: 12 June 2008).

Nowak, D.J., R.A. Rowntree, E.G. McPherson, S.M. Sisinni, E.R. Kerkmann, and J.C. Stevens, 1996. Measuring and analyzing urban tree cover, *Landscape and Urban Planning*, 36:49-57, URL: *http://www.fs.fed.us/ne/syracuse/Pubs/Downloads/96_DN_etal_Measuring.pdf* (last date accessed: 12 June 2008).

Pal, M., 2005. Random forest classifier for remote sensing classification, *International Journal of Remote Sensing*, 26(1):217–222.

Prasad, A.M., L.R. Iverson, and A. Liaw, 2006. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction, *Ecosystems*, 9:181–199.

Quinlan, J.R., 1986. Induction of decision trees, *Machine Learning*, 1:81–106.

Quinlan, J.R., 1987a. Generating production rules from decision trees, *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, Milan, Italy, Morgan Kaufmann, Los Altos, California, pp. 304–307.

Quinlan, J.R., 1987b. Simplifying decision trees, *International Journal of Man-Machine Studies*, 27:221–234, URL: *http://citeseer.ist.psu.edu/article/quinlan87simplifying.html* (last date accessed: 12 June 2008).

Quinlan, J.R., 1992. Learning with continuous classes, *Proceedings of the 5th Australian Joint Conference on Artificial Intelligence*, World Scientific, Singapore, pp. 343–348, URL: *http://citeseer.ist.psu.edu/quinlan92learning.html* (last date accessed: 12 June 2008).

Quinlan, J.R., 1993a. *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers, San Mateo, California, 302 p.

Quinlan, J.R., 1993b. A case study in machine learning, *Proceedings of the Sixteenth Australian Computer Science Conference (ACSC-16)*, URL: *http://citeseer.ist.psu.edu/quinlan93case.html* (last date accessed: 12 June 2008).

Quinlan, J.R., 1993c. Combining instance-based and model-based learning, *Proceedings of the Tenth International Conference on Machine Learning*, Morgan Kaufmann Publishers, San Francisco, pp. 236–243, URL: *http://citeseer.ist.psu.edu/quinlan93combining.html* (last date accessed: 12 June 2008).

Quinlan, J.R., 1996. Improved use of continuous attributes in C4.5, *Journal of Artificial Intelligence Research*, 4:77–90, URL:

http://citeseer.ist.psu.edu/quinlan96improved.html (last date accessed: 12 June 2008).

Quinlan, J.R., 1997. *New Tool for Continuous Classes (aka Regression)*, Machine Learning List, URL: *http://www.ics.uci.edu/~mlearn/MLlist/v9/13.html* (last date accessed: 12 June 2008).

R Development Core Team, 2006. *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3–900051–07–0, URL: *http://www.R-project.org* (last date accessed: 12 June 2008).

Ray, T.W., and B.C. Murray, 1996. Nonlinear spectral mixing in desert vegetation, *Remote Sensing of Environment*, 55:59–64.

Ridd, M.K., 1995. Exploring a V-I-S (vegetation-impervious surface-soil) model for urban ecosystem analysis through remote sensing: Comparative anatomy for cities, *International Journal of Remote Sensing*, 16(12):2165–2185.

Rokhmatuloh, D.N., H. Al Bilbisi, and R. Tateishi, 2005. Percent tree cover estimation using regression tree method: A case study of Africa with very-high resolution QuickBird images as training data, *IEEE International Geoscience and Remote Sensing Symposium 2005 (IGARSS '05)*, Vol. 3, pp. 2157–2160, doi: 10.1109/IGARSS.2005.1526445.

Rowntree, R.A., 1984. Forest canopy cover and land use in four eastern United States cities, *Urban Ecology*, 8:55–67.

Rulequest, 2006. Data Mining with Cubist, URL: *http://www.rulequest.com/cubist-info.html*, RuleQuest Research Pty Ltd., St. Ives, NSW, Australia (last date accessed: 12 June 2008).

Russell, S., and P. Norvig, 2003. *Artificial Intelligence: A Modern Approach*, Second edition, Prentice Hall, Pearson Education, Inc., Upper Saddle River, New Jersey, 1081 p.

Schueler, T.R., 1994. The importance of imperviousness, *Watershed Protection Techniques*, 1(3):100–111.

Schuler, D.V., and L. Kastdalen, 2005. Impervious surface mapping in Southern Norway, *Proceedings of the 31st International Symposium on Remote Sensing of Environment*, 20–24 May, Saint Petersburg, Russia (International Society for Photogrammetry and Remote Sensing), URL: *http://www.isprs.org/publications/related/ISRSE/html/papers/581.pdf* (last date accessed: 12 June 2008).

Settle, J.J., and N.A. Drake, 1993. Linear mixing and the estimation of ground cover proportions, *International Journal of Remote Sensing*, 14(6):1159–1177.

Small, C., 2001. Estimation of urban vegetation abundance by spectral mixture analysis, *International Journal of Remote Sensing*, 22(7):1305–1334.

Small, C., and J.W.T. Lu, 2006. Estimation and vicarious validation of urban vegetation abundance by spectral mixture analysis, *Remote Sensing of Environment*, 100:441–456.

Smola, A.J., and B. Schölkopf, 2004. A tutorial on support vector regression, *Statistics and Computing*, 14:199–222.

Srivastava, A.N., N.C. Oza, and J. Stroeve, 2005. Virtual sensors: Using data mining techniques to efficiently estimate remote sensing spectra, *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):590–600.

Vapnik, V.N., 2000. *The Nature of Statistical Learning Theory*, Second edition, Springer-Verlag, New York, 314 p.

Wang, S.-C., 1988. An analysis of urban tree communities using Landsat Thematic Mapper data, *Landscape and Urban Planning*, 15:11–22.

Witten, I.H., and E. Frank, 2005. *Data Mining: Practical Machine Learning Tools and Techniques*, Second edition, Morgan Kaufmann, San Francisco, 525 p.

Wu, C., and A.T. Murray, 2003. Estimating impervious surface distribution by spectral mixture analysis, *Remote Sensing of Environment*, 84:493–505.

Xu, M., P. Watanachaturaporn, P.K. Varshney, and M.K. Arora, 2005. Decision tree regression for soft classification of remote sensing data, *Remote Sensing of Environment*, 97:322–336.

Yang, L., C. Homer, K. Hegge, C. Huang, B. Wylie, and B. Reed, 2001. A Landsat 7 scene selection strategy for a National Land Cover Database, *Proceedings of the IEEE 2001 International Geoscience and Remote Sensing Symposium*, Sydney, Australia, unpagniated CD-ROM, URL: *http://landcover.usgs.gov/pdf/ l7_scene_select.pdf* (last date accessed: 12 June 2008).

Yang, L., C. Huang, C.G. Homer, B.K. Wylie, and M.J. Coan, 2003. An approach for mapping large-area impervious surfaces: synergistic use of Landsat-7 ETM+ and high spatial resolution imagery, *Canadian Journal of Remote Sensing*, 29(2):230–240, URL: *http://landcover.usgs.gov/pdf/imppaperfinalwithall.pdf* (last date accessed: 12 June 2008).

Yu, D., and C. Wu, 2006. Incorporating remote sensing information in modeling house values: A regression tree approach, *Photogrammetric Engineering & Remote Sensing*, 72(2): 129–138.

Zhang, Y., 2001. Texture-integrated classification of urban treed areas in high-resolution color-infrared imagery, *Photogrammetric Engineering & Remote Sensing*, 67(12):1359–1365.

Zhu, Z., 1994. *Forest Density Mapping in the Lower 48 States: A Regression Procedure*, Research Paper SO-280, U.S. Department of Agriculture, Forest Service, Southern Research Station, New Orleans, Louisiana, 11 p., URL: *http://www.srs.fs.usda.gov/pubs/rp/rp_so280.pdf* (last accessed: 12 June 2008).

Zipperer, W.C., S.M. Sisinni, R.V. Pouyat, and T.W. Foresman, 1997. Urban tree cover: An ecological perspective, *Urban Ecosystems*, 1:229–246.