Geocarto International
2009, 1–20, i*First* article

Taylor & Francis
Taylor & Francis Group

Hyperspectral remote sensing analysis of short rotation woody crops grown with controlled nutrient and irrigation treatments

Jungho Im[a]*, John R. Jensen[b], Mark Coleman[c] and Eric Nelson[d]

[a]*Department of Environmental Resources and Forest Engineering, State University of New York, College of Environmental Science and Forestry, Syracuse, New York, USA;* [b]*Department of Geography, University of South Carolina, Columbia, South Carolina, USA;* [c]*USDA Forest Service, Southern Research Station, Aiken, South Carolina, USA;* [d]*Savannah River National Laboratory, Savannah River Site, US Department of Energy, Aiken, South Carolina, USA*

(*Received 3 October 2008; final version received 15 October 2008*)

Hyperspectral remote sensing research was conducted to document the biophysical and biochemical characteristics of controlled forest plots subjected to various nutrient and irrigation treatments. The experimental plots were located on the Savannah River Site near Aiken, SC. AISA hyperspectral imagery were analysed using three approaches, including: (1) normalized difference vegetation index based simple linear regression (NSLR), (2) partial least squares regression (PLSR) and (3) machine-learning regression trees (MLRT) to predict the biophysical and biochemical characteristics of the crops (leaf area index, stem biomass and five leaf nutrients concentrations). The calibration and cross-validation results were compared between the three techniques. The PLSR approach generally resulted in good predictive performance. The MLRT approach appeared to be a useful method to predict characteristics in a complex environment (i.e. many tree species and numerous fertilization and/or irrigation treatments) due to its powerful adaptability.

Keywords: remote sensing; hyperspectral analysis; partial least squares regression; machine-learning regression trees; NDVI; leaf nutrients; leaf area index; biomass

Introduction

Forest plantation management practices impact forest productivity. Important forest management practices include pest control, irrigation and fertilization (Stanton *et al*. 2002, Coyle and Coleman 2005). Nutrient deficiency and water stress are major factors that limit forest productivity. Leaf area, nutrient concentration and carbon assimilation rate are affected by fertilization (i.e. soil nutrient availability), which in turn influences leaf nutrient levels (Samuelson *et al*. 2001, Xiao *et al*. 2003). Water availability also impacts nutrient uptake and photosynthesis (Landsberg 1986, Blake *et al*. 1996).

The health and productivity of forests can be predicted by measuring forest biophysical and biochemical characteristics at the leaf or canopy level (e.g. Hansen and Schoerring 2003, Yamashita *et al*. 2004, Chmura *et al*. 2007, Tilling *et al*. 2007). Unfortunately, measuring forest biophysical and biochemical characteristics at the foliar or canopy level with traditional *in situ* field methods is labour-intensive and time-consuming and difficult to conduct over large geographic areas. Consequently, remote sensing has been widely

*Corresponding author. Email: imj@esf.edu

ISSN 1010-6049 print/ISSN 1752-0762 online

DOI: 10.1080/10106040802556207
http://www.informaworld.com

adopted for some forest management applications (e.g. Johnson *et al.* 1994, Zhao *et al.* 2005, Chirici *et al.* 2007, Perry and Davenport 2007). In particular, the advent of airborne hyperspectral remote sensing at high spatial resolutions has made it possible to collect detailed spectral signatures in the region from 400 to 2500 nm to measure a number of biophysical, biochemical or physiological characteristics such as leaf area index (LAI), biomass and chlorophyll (e.g. Hu *et al.* 2004, Beeri *et al.* 2007). Less research has been performed on remote sensing of nutrients and trace elements such as nitrogen, phosphorous, potassium, calcium and magnesium especially in canopies of different species (e.g. Gong *et al.* 2002, Mutanga *et al.* 2003, Ferwerda and Skidmore in press).

Several approaches have been investigated for estimating biophysical and biochemical forest characteristics using hyperspectral remote sensing data. For example, some scientists employed regression analysis to correlate biophysical and/or biochemical characteristics with either reflectance or vegetation indices (Johnson *et al.* 1994, Jensen 2005, Tilling *et al.* 2007, Ye *et al.* 2007). Regression techniques tested include multiple regression, principal component regression and partial least squares regression (PLSR). Other scientists have focused on identifying the spectral reflectance red-edge position because it is closely associated with chlorophyll content and its seasonal variations (Curran *et al.* 1995, Cho and Skidmore 2006).

To learn more about the effects of nutrient and water availability on forest productivity, Coleman *et al.* (2004) created experimental plots of short-rotation woody crops on the U.S. Department of Energy's Savannah River Site near Aiken, SC (Figure 1). Since 2000, they have maintained plots containing four tree species subjected to a range of nutrient and irrigation treatments. They periodically collected *in situ* measurements on these plots including information on: LAI, stem biomass and leaf nutrient concentration for each treatment. Hyperspectral remote sensor data were obtained over this study area in 2006. The objectives of this study were to (1) identify the relationship between *in situ* and hyperspectral remote sensing measurement in terms of tree species and the different levels of water and nutrient availability, and (2) estimate biophysical and biochemical characteristics (i.e. LAI, biomass, leaf nutrients concentrations) from the hyperspectral imagery using three different digital image processing techniques. The leaf nutrient variables included nitrogen (N), phosphorous (P), potassium (K), calcium (Ca) and magnesium (Mg). The three digital image processing techniques used were (1) single regression using normalized difference vegetation index (NDVI), (2) partial least squares (PLS) regression and (3) machine learning regression trees (MLRT).

Study area and data

Study site

The study area was located in the northwest sector of the U.S. Department of Energy Savannah River Site, a National Environmental Research Park, near Aiken, SC (Figure 1). The climate is humid continental with an average annual temperature of 17.9°C. The dominant soil is Blanton sand with loamy subsoil at a depth of 120–200 cm (Rogers 1990). Previous vegetation of the site was plantation pine with an oak (*Quercus spp.*) understory.

Experimental design

Treatments at the short-rotation woody crop study site are depicted in Figure 2(a). The study area consists of five blocks with each block containing 14–28 plots. Five tree genotypes (two eastern cottonwood clones (*Populus deltoides Bartr. ST66* and *S7C15*),

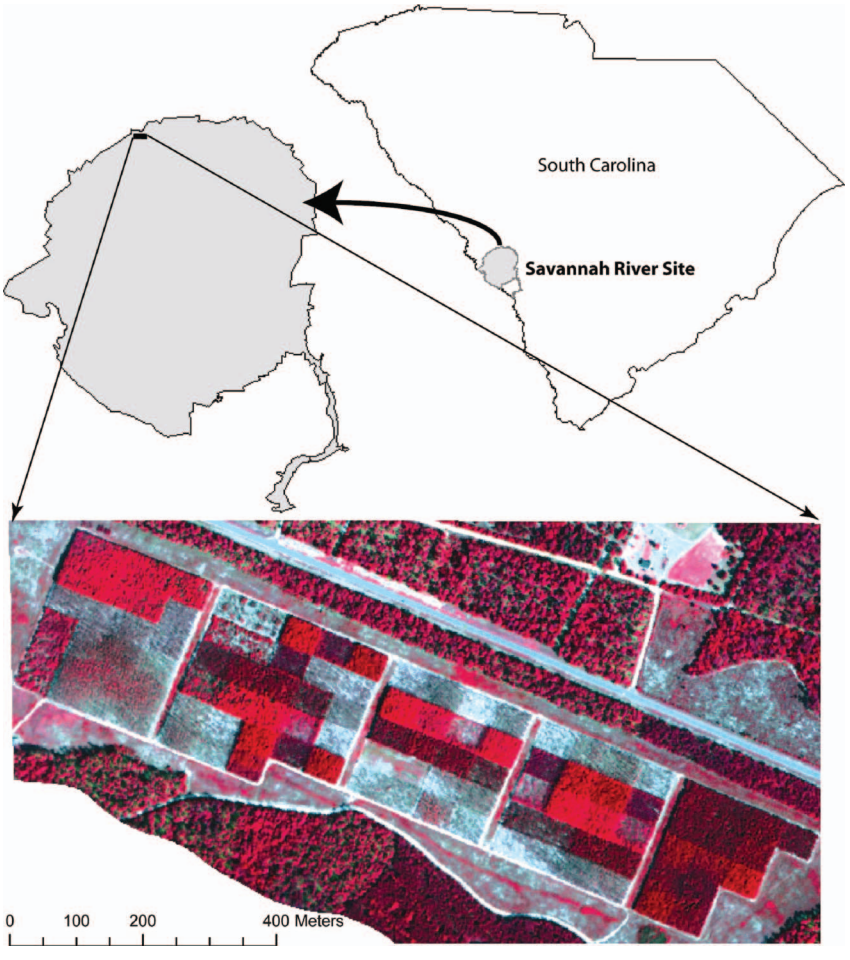
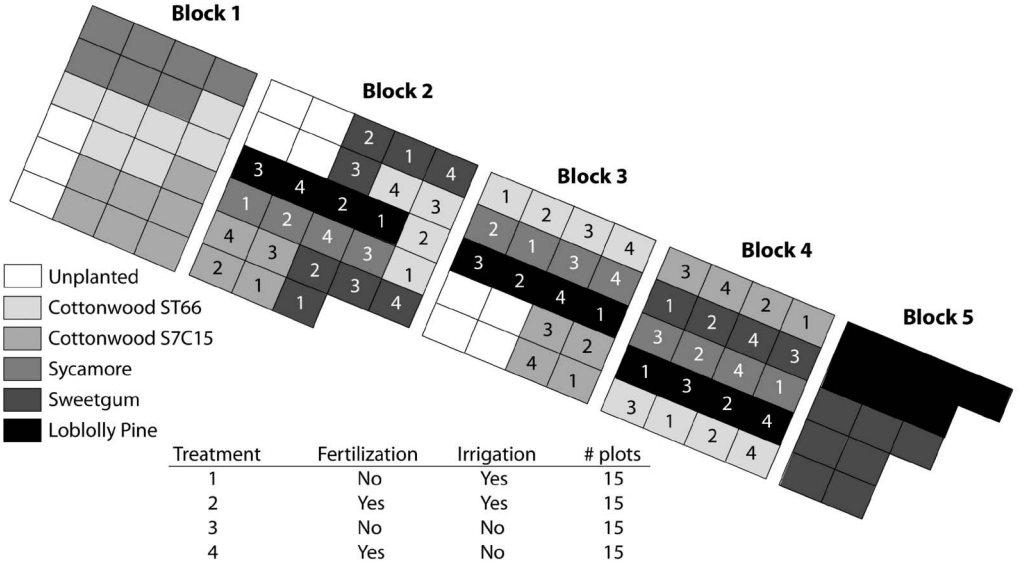


Figure 1. The Study area on the Savannah River Site and color composite image (RGB = hyperspectral bands 760.8, 666.4, and 572.5 nm, respectively) of the experimental plots.

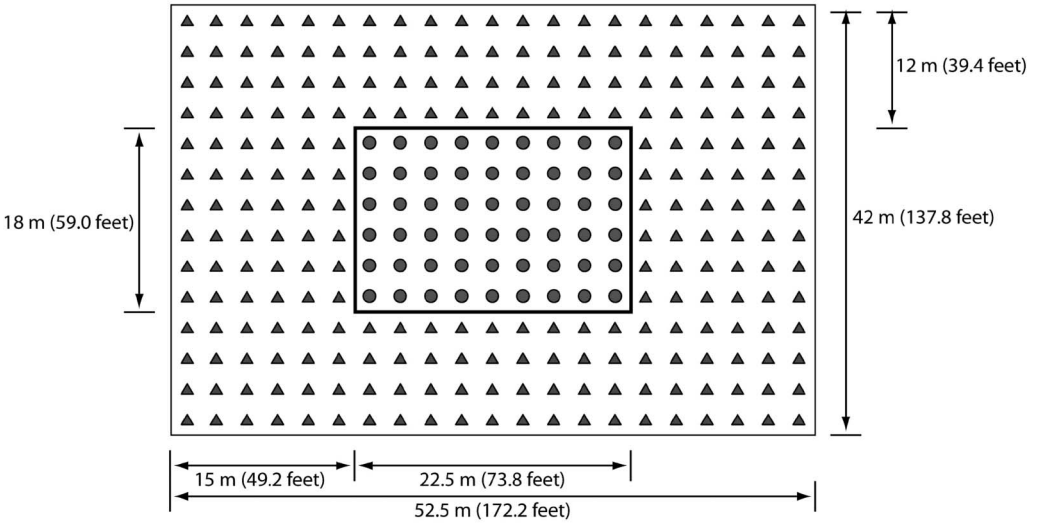
sycamore (*Platatus occidentalis L.*), sweetgum (*Liquidambar styraciflua L.*) and loblolly pine (*Pinus taeda L.*)) were established and subjected to a range of nutrient and water treatments. Each treatment plot contained 14 by 21 trees (i.e., 294 trees) with tree spacing of 2.5 by 3 m (Figure 2(b)). A central 0.04 ha measurement area including 54 trees in each plot is used for periodic *in situ* data collections. Treatments were composed of (1) control (no irrigation and fertilization), (2) irrigation, (3) fertilization and (4) irrigation + fertilization. Different fertilization levels were also applied to experimental plots in blocks 1 and 5, but they were not included in this study due to the limited *in situ* measurements (i.e., the shaded plots not numbered in the Figure 2(a)).

Irrigated and non-irrigated water treatments were investigated. Irrigated plots received water via a drip irrigation system. Irrigated plots received up to 5 mm per day, 6 days a week during the growing season in order to match evaporative demand (Treatments 1 and 2 in Figure 2(a)). The non-irrigated treatments received only 5 mm per week (Treatments 3 and 4 in Figure 2(a)). Fertilized plots have received a balanced fertilizer blend of macro and micro nutrients annually since planting. In 2006, each fertilized plot received



Treatment	Fertilization	Irrigation	# plots
1	No	Yes	15
2	Yes	Yes	15
3	No	No	15
4	Yes	No	15

a. Treatments at the short-rotation woody crops study site. Treatments 1 to 4 were replicated in the central blocks for all tree species. Sixty *in situ* measurements in the treatments of Blocks 2-4 were used in the study.



b. Individual treatment plot having 14 by 21 tree rows (294 trees). *In situ* measurements were collected on the 54 central trees (i.e. circle).

Figure 2. Overview of the experimental plots with different treatments.

120 kg N ha^{-1} (Treatments 2 and 4 in Figure 2(a)). Annual fertilization consisted of 26 equal weekly applications between 1 April and 1 October.

In summary, a total of 60 plots in blocks 2–4 were used in the study, which included 20 treatment plots (5 genotypes grown with 2 fertilizer and 2 irrigation treatments) with each replicated three times (Figure 2(a)). Detailed information about the experimental design is found in Coleman *et al.* (2004).

Hyperspectral imagery and data extraction

An Airborne Imaging Spectrometer for Applications (AISA) Eagle sensor system (CHAMP, CALMIT, University of Nebraska-Lincoln) was used to obtain hyperspectral imagery over the study area on 15 September, 2006. The imagery consisted of 63 channels (from 400 to 980 nm) with a spectral resolution of ~9-nm, a radiometric resolution of 12-bits, and a fine spatial resolution of 1 m × 1 m. The imagery was collected at an altitude of 1630 m above ground level (AGL) during cloud-free conditions at ~11:20 am (EDT) local time.

The hyperspectral imagery were pre-processed to ground percent reflectance using a Fast Line-of-sight Atmospheric Analysis of Spectral Hypercube (FLAASH) algorithm. The imagery was then rectified to a Universal Transverse Mercator (UTM) coordinate system. The geometric rectification was performed using GPS-derived coordinates located in the study site (i.e., each corner of each block) resulting in an estimated Root Mean Square Error (RMSE) of 0.48 pixels.

The radiometrically and geometrically corrected hyperspectral imagery was imported into ESRI ArcGIS 9.x. Since the *in situ* field measurements were collected on the central 54 trees of each treatment, hyperspectral measurements were extracted from these corresponding locations. Polygons surrounding the central 54 trees were created based on the GPS-derived coordinate values of each treatment plot. Zonal functions were used to compute the mean reflectance values of the hyperspectral pixels within the polygons. Each polygon contained ~405 pixels. The mean spectral reflectance values were used in the subsequent analyses.

In situ *measurements*

In situ biophysical and biochemical measurements included LAI, stem biomass and five leaf nutrient concentrations (i.e., N, P, K, Ca, and Mg). LAI measurements (m^2/m^2) were collected from 15–18 September, 2006 using a hand-held ceptometer. Stem biomass (Mg/ha) was calculated using diameter-at-breast-height (DBH) measurements obtained in December, 2006 as independent variables in allometric equations developed from biomass harvesting. Although the DBH measurements were obtained 2 months after the remote sensing data acquisition, the stem biomass only changed <0.04% during this time based on a subset of stem diameter measurements taken on 11 October, 2006. Composite hardwood leaf samples were collected for nutrient analysis on 5–7 July, 2006. Pine was sampled in January of 2006. We acknowledge that the acquisition date discrepancy for leaf nutrients could influence the data analysis.

Methodology

The study methodology is summarized in Figure 3. After radiometric and geometric pre-processing, the mean spectral reflectance values were extracted from the GPS-derived polygonal areas within each plot. How the spectral patterns differed based on nutrient and/or water availability (measured *in situ*) and tree species was investigated. After the spectral pattern analysis, a predictive analysis of the biophysical and biochemical characteristics was performed using each of the three digital image processing techniques, including: NDVI-based simple linear regression (NSLR), PLS regression (PLSR) and MLRT.

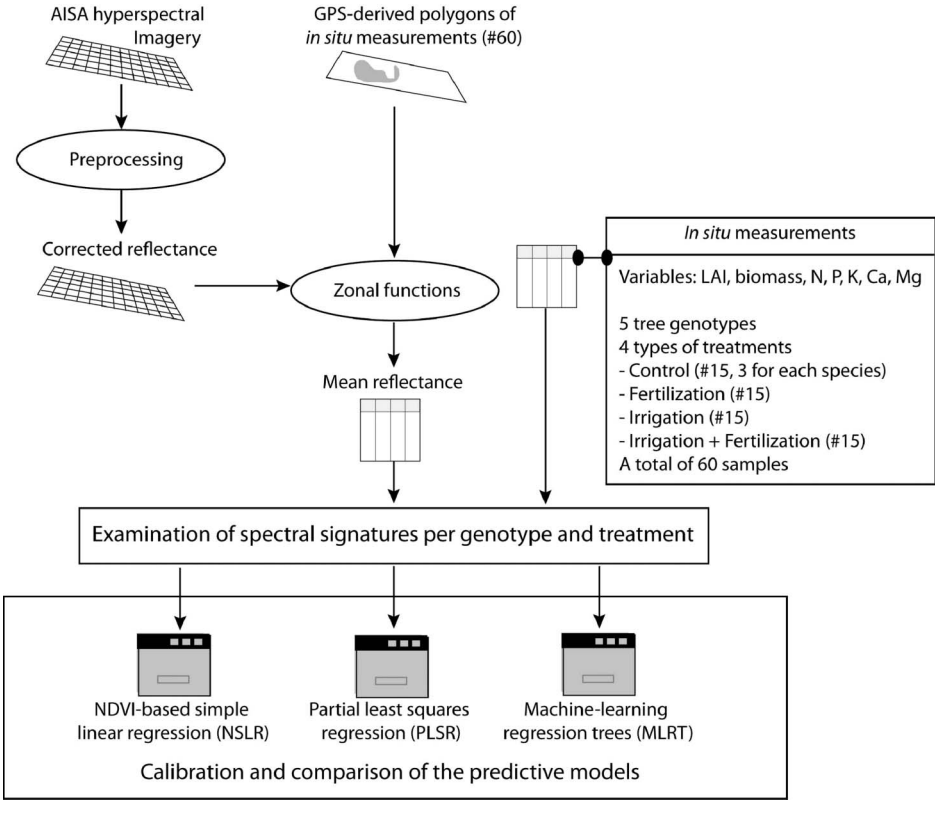


Figure 3. Process flow diagram of the research.

Normalized difference vegetation index

There are many types of vegetation indices applicable for hyperspectral analysis (Jensen 2005). This study focused on one representative vegetation index called the NDVI. The NDVI is a classical index based on red (R) and near-infrared (NIR) reflectance:

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{R}}{\mathrm{NIR} + \mathrm{R}} \tag{1}$$

Numerous studies have reported that NDVI is useful for estimating vegetation characteristics such as LAI, biomass, chlorophyll and nitrogen (e.g. Aparicio *et al*. 2000, Hansen and Schjoerring 2003, Jensen 2007). Scientists using traditional multi-spectral data usually have only one or two red and NIR bands available to use in the NDVI algorithm. Scientists using hyperspectral data may select among numerous red and NIR bands to use in the NDVI algorithm. Selecting an optimum band combination from hyperspectral data is critical to compute NDVI and requires a large size of *in situ* samples to calibrate and validate the NDVI. This study simply focused on a standard NDVI by using the NIR band cantered at 856.8 nm and the red band centred at 657.0 nm (Jensen 2005).

Partial least squares regression

PLSR uses concepts of principal component analysis and multiple regression at the same time. PLSR reduces a number of independent variables, which might display multi-collinearity, to a few manageable non-correlated variables using component projection. These non-correlated variables represent the relevant latent structural information of the reflectance measurements for predicting the dependent variable such as LAI (Hansen and Schjoerring 2003). One of the advantages of PLSR is the minimal demands on sample size and residual distributions. Another feature is the ability to model multiple dependent variables. Unfortunately, it is often difficult to interpret the PLSR loadings because they are from a cross product between dependent and independent variables (Garson 2006). Thus, PLS regression has great potential as a predictive model, but not necessarily as an interpretive model.

The number of non-correlated variables used in the subsequent analyses is another consideration. Visual inspection of the scores associated with model effects and validation residual variance plots is generally used to locate the optimum number of non-correlated variables through PLS regression. This study used 95% cumulative model effects as a criterion for the optimum number of non-correlated variables. The treatments contained five different tree genotypes. To incorporate genotypes as an independent variable, five additional dummy variables were created with values of 0 and 1 (e.g., for sweetgum: 0, 0, 1, 0, 0). Consequently, two sets of PLSR analyses were performed for seven dependent variables: (1) using 63 reflectance independent variables, and (2) using 63 reflectance variables + 5 dummy variables to distinguish tree genotypes.

Machine learning regression trees

Traditional regression trees use a binary recursive partitioning process (Breiman *et al.* 1984). Training samples are input to the regression trees to generate rule-based models for predicting a target variable using a recursive partitioning process. A split occurs if the model's combined residual error for two subsets is significantly lower than the residual error of the single best model in the process (Huang and Townshend 2003). Advantages of MLRT include the ability to handle non-linear relationships between independent and dependent variables, and the use of both continuous and discrete variables as input data. To predict continuous biophysical and biochemical characteristics, this study used Cubist by *RuleQuest Inc.*, which uses a modified regression tree system to create rule-based predictive models from the data. Each rule has an associated multivariate linear model. These linear models are not mutually exclusive, allowing overlap between models. Output values are averaged to arrive at a final prediction. The predictability of Cubist MLRT has been examined in several studies (e.g. Huang and Townshend 2003, Yang *et al.* 2003, Moisen *et al.* 2006). Since Cubist allows categorical (i.e., discrete) variables as input variables, three categorical variables were used including tree genotypes, irrigation and fertilization. MLRT using just the 63 reflectance variables was also performed for comparison.

Since only 60 samples were available for all the genotypes and treatments, validation of the models was conducted using a cross-validation technique. RMSE based on the predicted value (Q_i) and observed value ($Q_{\text{observed } i}$) was used as a measure of performance:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (Q_i - Q_{\text{observed}})^2}{N}} \tag{2}$$

The coefficient of determination (R^2), which provides some information about the goodness of fit of a model, was also used to measure calibration performance.

Results and discussion

Spectral signature patterns

All samples for each genotype and type of treatment from the AISA hyperspectral imagery were averaged to produce spectral reflectance curves. Spectral responses of each tree genotype associated with water and/or nutrient treatments are shown in Figure 4. The cottonwood present in this study exhibited a thin, relatively sparse canopy. Spectral patterns associated with cottonwood were often quite different from those associated with the other species (Figure 4(a and b)). This was because the bare ground visible to the hyperspectral sensor through the relatively sparse cottonwood canopy influenced the spectral reflectance. Because of the bare ground effect, the control treatments from less dense canopies resulted in higher reflectance curves compared with the irrigated and/or fertilized treatments.

Sycamore and sweetgum plots had more dense continuous canopies than cottonwood. The spectral reflectance curves associated with the nutrient and water treatments for these species were often dramatically different from the control reflectance characteristics. For example, irrigated sycamore exhibited substantially lower spectral reflectance in the visible part of the spectrum ($p = 0.026$ at 619 nm) and approximately the same reflectance as the control in the near-infrared part of the spectrum ($p = 0.843$ at 857 nm). Fertilized sycamore yielded unique spectral patterns in the near-infrared producing much higher reflectance than the control and irrigation treatments ($p = 0.016$ at 857 nm) (Figure 4(c)). Interestingly, the fertilization treatment of sycamore resulted in slightly higher reflectance than the fertilization + irrigation treatment in the near-infrared region. The sweetgum treatments produced spectral reflectance patterns similar to sycamore. Irrigation was, however, also an influencing factor, which yielded a distinctive spectral pattern, compared with the control treatment (Figure 4(d)). The loblolly pine treatments showed similar spectral patterns to the sweetgum treatments. However, the near-infrared reflectance of the loblolly pine treatments was lower than that of the sweetgum treatments, which had denser canopy with broad leaves.

Although a few bands resulted in statistical difference in the spectral reflectance of the two cottonwood genotypes derived from the hyperspectral imagery, it was not possible to identify that such difference was caused by the effects of irrigation and/or fertilization because of the thin canopy and bare ground effects. Discrimination between the thin canopy and bare ground was not possible from the data at the 1 m $\times$ 1 m resolution. On the other hand, the effects of irrigation and/or fertilization were easily identified in other treatments such as sycamore, sweetgum and loblolly pine based on the spectral signatures. Statistical ANOVA tests for the spectral signatures reported that fertilization was a more discriminant factor than irrigation especially in sycamore and loblolly pine.

NDVI-based regression (NSLR)

The relationship between NDVI and LAI for all species is shown in Figure 5(f) resulting in a R^2 of 0.738 and an RMSE of 0.8194 m^2/m^2 through cross-validation. We also investigated the relationship between NDVI and LAI per genotype (Figure 5(a through e)). Irrigation and fertilization effects were easily detected in the relationships. Fertilization appeared to be more critical to plant growth in terms of LAI than irrigation. This corresponds to the results of the statistical ANOVA tests for the spectral signatures.

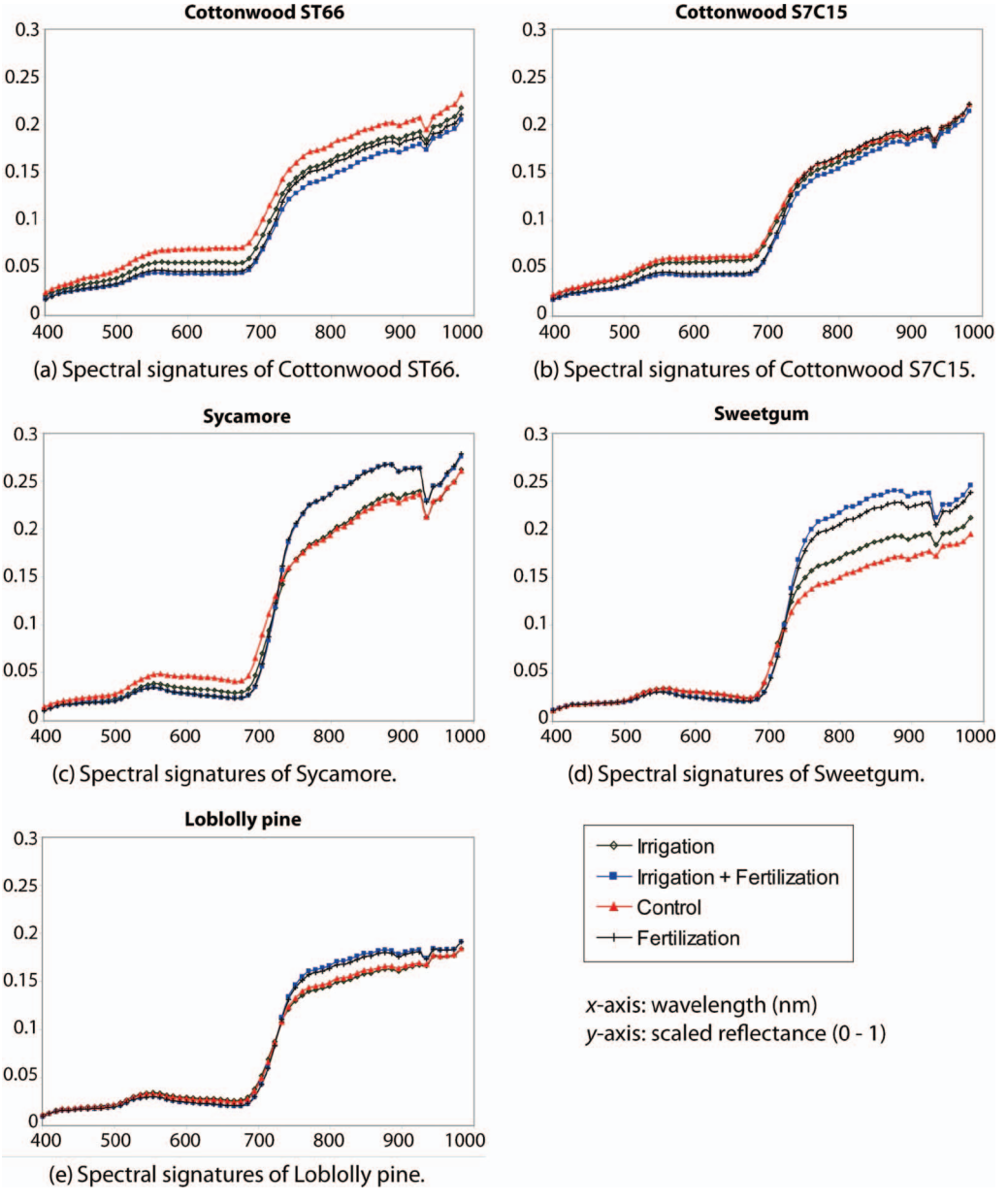


Figure 4. Spectral signatures of the five tree genotypes associated with irrigation and/or fertilization treatments.

The NDVI resulted in relatively lower R^2 values for the two cottonwood genotypes, and the irrigation and fertilization effects were not that explicit compared with the other species (Figure 5(a and b)). This was because the reflectance values used in the NDVI computation were not solely from the canopy. The NDVI showed very high correlation with LAI for sweetgum and loblolly pine, over $R^2 = 0.9$ (Figure 5(d and e)). Different band combinations for calculating NDVI might produce better results especially for cottonwood and sycamore.

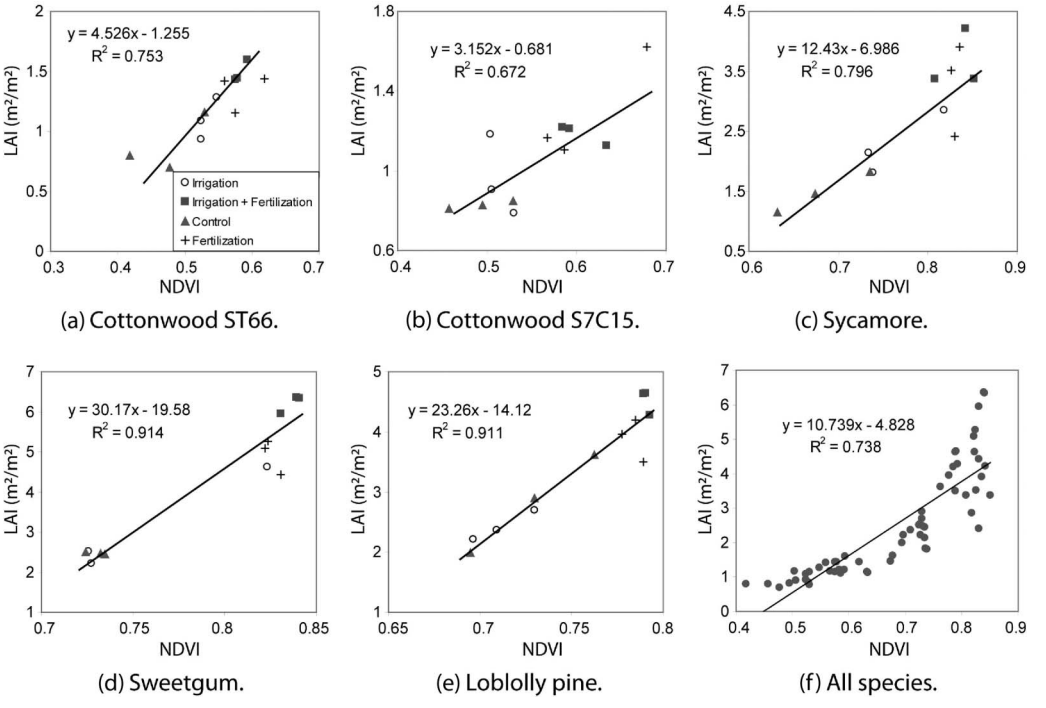


Figure 5. Regression plots between the NDVI and LAI for individual tree genotypes and for all species using the NSLR approach.

The standard NDVI resulted in an R^2 of 0.569 to estimate stem biomass (Figure 6(a)). During the calibration for each genotype, similar patterns to the LAI calibration were found: relatively lower R^2 for the two cottonwood genotypes and strong linear relationships for sweetgum and loblolly pine. Irrigation and fertilization effects were clearly found in the relationships except for the two cottonwood genotypes.

Calibrations for the five leaf nutrients concentrations are shown in Figure 6(b through f). All of calibrations resulted in weak linear relationships between leaf nutrients and the NDVI. When calibrated for individual genotypes, relatively strong linear relationships (e.g. $R^2 > 0.6$), however, were extracted. This is because the leaf nutrients were generally much more species-sensitive compared with the biophysical characteristics such as LAI and stem biomass. Thus, it failed to extract a strong linear relationship that could be applied to all the species. This is somewhat different from the results reported by Ferwerda and Skidmore (2007), which found stronger linear relationships across all species when compared with individual species. The major reason was that larger variation in the target variables for individual species existed in our study due to the diverse treatments at different levels of irrigation and fertilization.

The calibration and cross-validation results using the NSLR approach are presented in Table 1. Selecting the optimum two bands for NDVI computation is critical for this method. The optimum two bands might be different species by species. Although this study simply computed a standard NDVI using two representative NIR and red bands (centred at 856.8 and 657.0 nm), further exploration on an optimum band combination to compute NDVI will be necessary.

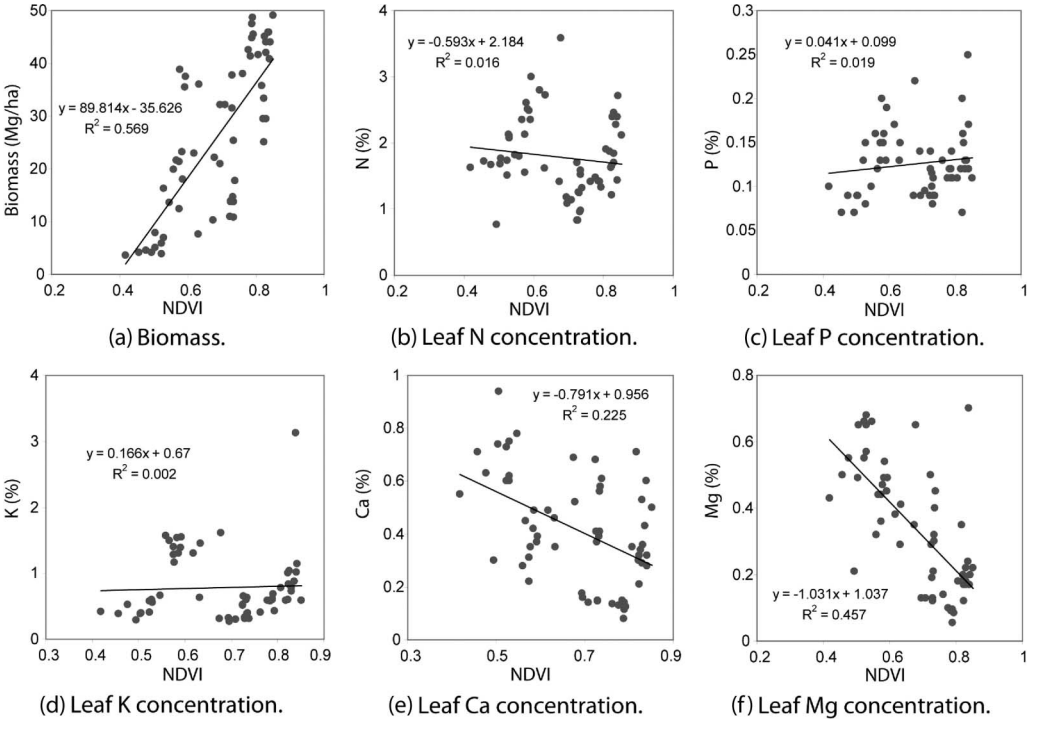


Figure 6. Regression plots between the NDVI and (a) biomass and (b)–(f) leaf nutrients concentrations for all tree species using the NSLR approach.

Table 1. Calibration and cross-validation results using the NSLR approach based on the use of bands centred on 856.8 and 657.0 nm.

Target variable	Calibration R^2	Cross-validation RMSE
LAI	0.738	0.8194 m^2/m^2
Biomass	0.569	9.8980 Mg/ha
N	0.016	0.5892%
P	0.019	0.0374%
K	0.002	0.5140%
Ca	0.225	0.1871%
Mg	0.457	0.1436%

Partial least squares regression

PLSR was performed using the SAS PLS procedure. Two different sets of input variables were used: 63 reflectance variables (PLSR1) and 63 reflectance + 5 dummy variables to distinguish the tree genotypes (PLSR2). Table 2 summarizes the calibration and cross-validation results using the PLSR approach. When applying a 95% of model effects as a criterion for selecting the optimum number of transformed components, two or three components were chosen based on the first input dataset (i.e. PLSR1). One or two additional components were added when calibrating with the second input dataset. The second input dataset that included the dummy variables yielded better calibration and

Table 2. Calibration and cross-validation results using the PLSR approach.

Target variable	PLSR1 (using 63 reflectance variables)		PLSR2 (using 63 reflectance + 5 dummy variables)	
	Calibration R^2	Cross-validation RMSE	Calibration R^2	Cross-validation RMSE
LAI	0.854 (3)*	0.6457 m^2/m^2	0.907 (3)	0.5118 m^2/m^2
Biomass	0.633 (2)	9.3579 Mg/ha	0.853 (3)	5.9233 Mg/ha
N	0.133 (2)	0.5589%	0.665 (4)	0.3732%
P	0.177 (3)	0.0353%	0.296 (3)	0.0329%
K	0.222 (3)	0.4693%	0.389 (3)	0.4219%
Ca	0.484 (2)	0.1575%	0.631 (3)	0.1341%
Mg	0.506 (2)	0.1415%	0.633 (3)	0.1224%

*Values in the parentheses represent the number of the components selected.

validation performance (i.e. higher R^2 and lower RMSE) than the first input dataset. This might be due to the different number of the components used. If more components were used in the PLSR analysis, then better performance was generally achieved. Thus, we applied additional components (i.e. the same number of components with the second case) to the PLSR analysis with the first input dataset. The R^2 values increased and the RMSE values decreased, but they were still lower than the performance measures from the second dataset. Consequently, the five dummy variables discriminating the tree genotypes improved prediction of the biophysical and biochemical characteristics of the five tree genotypes.

Figure 7 shows scatterplots between predicted and observed values for the seven target variables for the PLSR2 models with the R^2 values. The PLSR2 models yielded better estimations of LAI and stem biomass than the PLSR1. They did not exhibit strong relationships for the leaf nutrients, especially for the phosphorous (P) concentration.

The factor loadings represent how the PLS components were constructed from the centred and scaled independent variables. The factor loadings associated with the seven dependent variables under investigation for the PLSR2 models are depicted in Figure 8. Most of the components could be divided into three regions: (1) before the red edge region, (2) after the red edge region and (3) dummy variables. The factor loadings dramatically changed in the red edge regions for most of the components. For example, the first PLS component for LAI had low factor loadings between 400 and 720 nm and then dramatically increased to become flat after 741.7 nm. The second component showed a global minimum at 713.5 nm in the whole wavelength range (excluding the dummy variables) and then shifted to a global high loading at 760.8 nm.

Machine learning regression trees (MLRT)

Two input datasets were used with the MLRT approach: (1) 63 reflectance variables (MLRT1) and (2) 63 reflectance + 3 additional discrete variables, i.e. tree genotypes, fertilization and irrigation (MLRT2). Table 3 summarizes calibration (R^2) and cross-validation (RMSE) results for MLRT1 and MLRT2. Interestingly, additional input variables did not always improve calibration performance unlike the PLSR approach. For example, LAI was better predicted using the 63 input variables ($R^2 = 0.956$) compared to

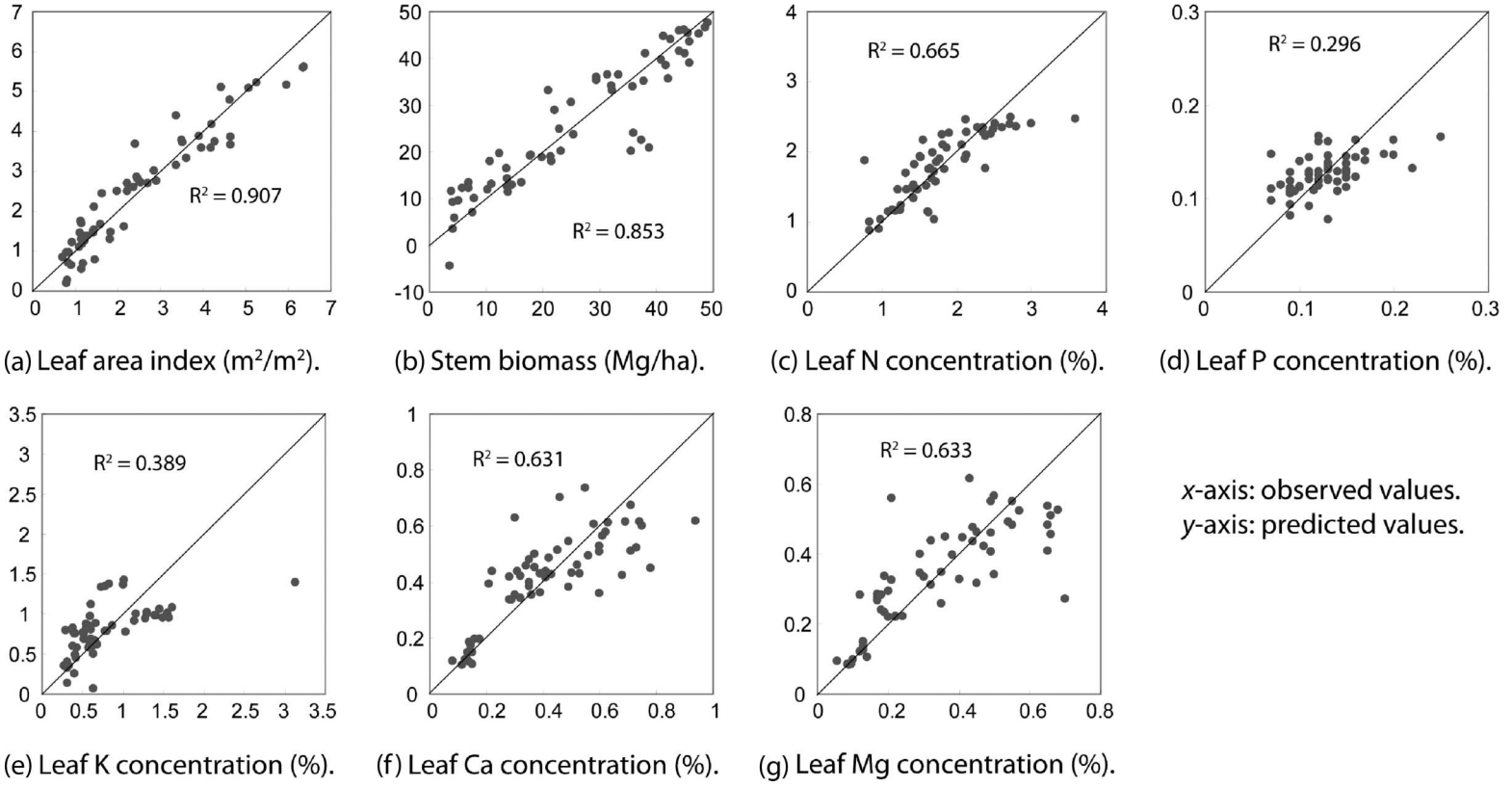


Figure 7. Scatterplots between predicted and observed values of the seven target variables for the PLSR models using the 68 predictors including the dummy variables.

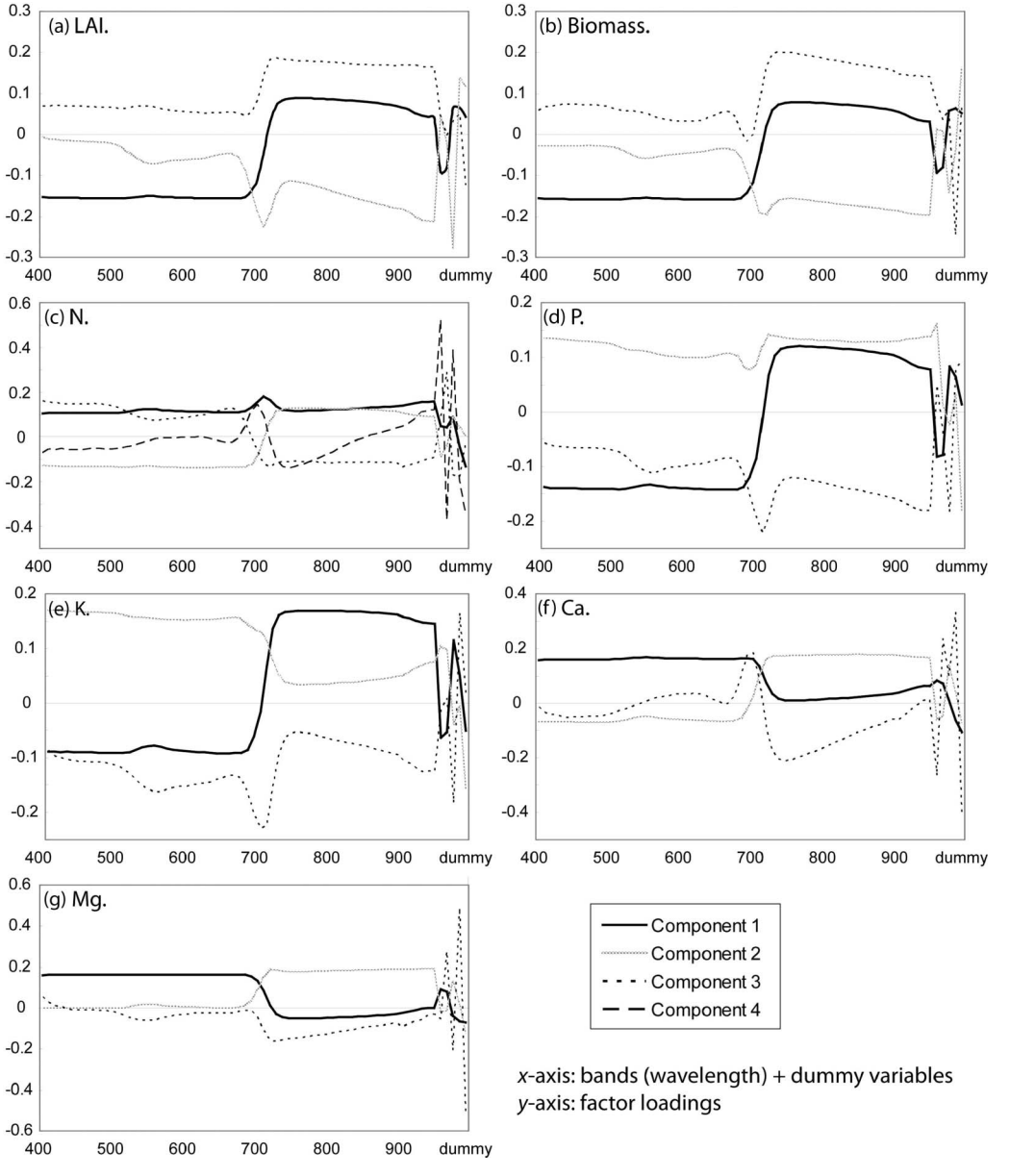


Figure 8. Factor loadings for the PLSR models using the 68 predictors including the dummy variables.

using the 66 input variables ($R^2 = 0.950$). On the other hand, the calibration of the leaf potassium (K) concentration was much improved by adding three additional discrete variables (R^2 from 0.557 to 0.936). This was possibly due to one sample of sweetgum, which had much higher K concentration compared with the other samples. When using only 63 reflectance variables, the sample was not well calibrated producing relatively lower R^2. The three additional variables helped predict the sample more accurately. This sample, however, influenced cross-validation by increasing the RMSE value. Scatterplots between

Table 3. Calibration and cross-validation results using the MLRT approach.

Target variable	MLRT1 (using 63 reflectance variables)		MLRT2 (using 63 reflectance + 3 discrete variables)	
	Calibration R^2	Cross-validation RMSE	Calibration R^2	Cross-validation RMSE
LAI	0.956 (2)*	0.4582 m^2/m^2	0.950 (2)	0.3937 m^2/m^2
Biomass	0.855 (2)	6.3923 Mg/ha	0.881 (2)	6.4617 Mg/ha
N	0.793 (3)	0.4836%	0.786 (3)	0.5092%
P	0.229 (1)	0.0335%	0.318 (2)	0.0326%
K	0.557 (2)	0.4270%	0.936 (4)	0.5588%
Ca	0.626 (1)	0.1528%	0.730 (2)	0.1490%
Mg	0.635 (2)	0.1430%	0.657 (2)	0.1366%

*Values in the parentheses represent the number of the rules generated from Cubist.

predicted and observed values of the seven target variables for the MLRT2 models with the R^2 values are shown in Figure 9. The MLRT2 models produced similar calibration performance to the PLSR2 models.

The addition of three discrete variables in the MLRT model generally increased R^2 values. But, the validation errors associated with the left-out samples increased in some cases (e.g., biomass, K, Ca). Overfitting of the models can explain this condition. Overfitting is one of the limitations of the MLRT approach wherein irrelevant details of the individual cases are modelled rather than learning the basic structure of the sample data (Waheed *et al*. 2006). MLRT is generally adapted well for non-linear problems as well as linearity. This adaptability of the model may lead to overfitting especially when a small number of samples are used. Since only 60 samples were available in this study, overfitting was detected in some of the MLRT models through the calibration and cross-validation processes. The overfitting problem may be solved by using more samples. A compromise would be to assign more of the samples to the validation dataset. A graphical comparison of calibration and cross-validation performance among the five models is presented in Figure 10.

In summary, the MLRT approach using the 66 variables yielded the best performance for predicting LAI (highest R^2 and lowest RMSE). Stem biomass was best estimated using the PLSR approach (PLSR2). The MLRT approach calibrated biomass better, but RMSE through the cross-validation increased due to the overfitting of the model. As for leaf nutrients, the PLSR model including the dummy variables resulted in consistently good performance in both calibration and cross-validation. On the other hand, the NSLR approach produced relatively poor performance compared with the other approaches in estimating the forest characteristics (i.e. low R^2 and high RMSE). The MLRT approach exhibited overfitting problems when estimating some of the leaf nutrients concentrations.

This study investigated three techniques for estimating biophysical and biochemical characteristics of five tree genotypes with nutrient and/or irrigation treatments using hyperspectral imagery. The findings were as follows:

- Nutrient and irrigation availability influenced plant growth and status yielding different patterns in the biophysical and biochemical characteristics of the trees, which were also identified in the spectral responses extracted from the hyperspectral imagery. Fertilization was found to be more influential than irrigation when the spectral signatures were extracted by treatment.

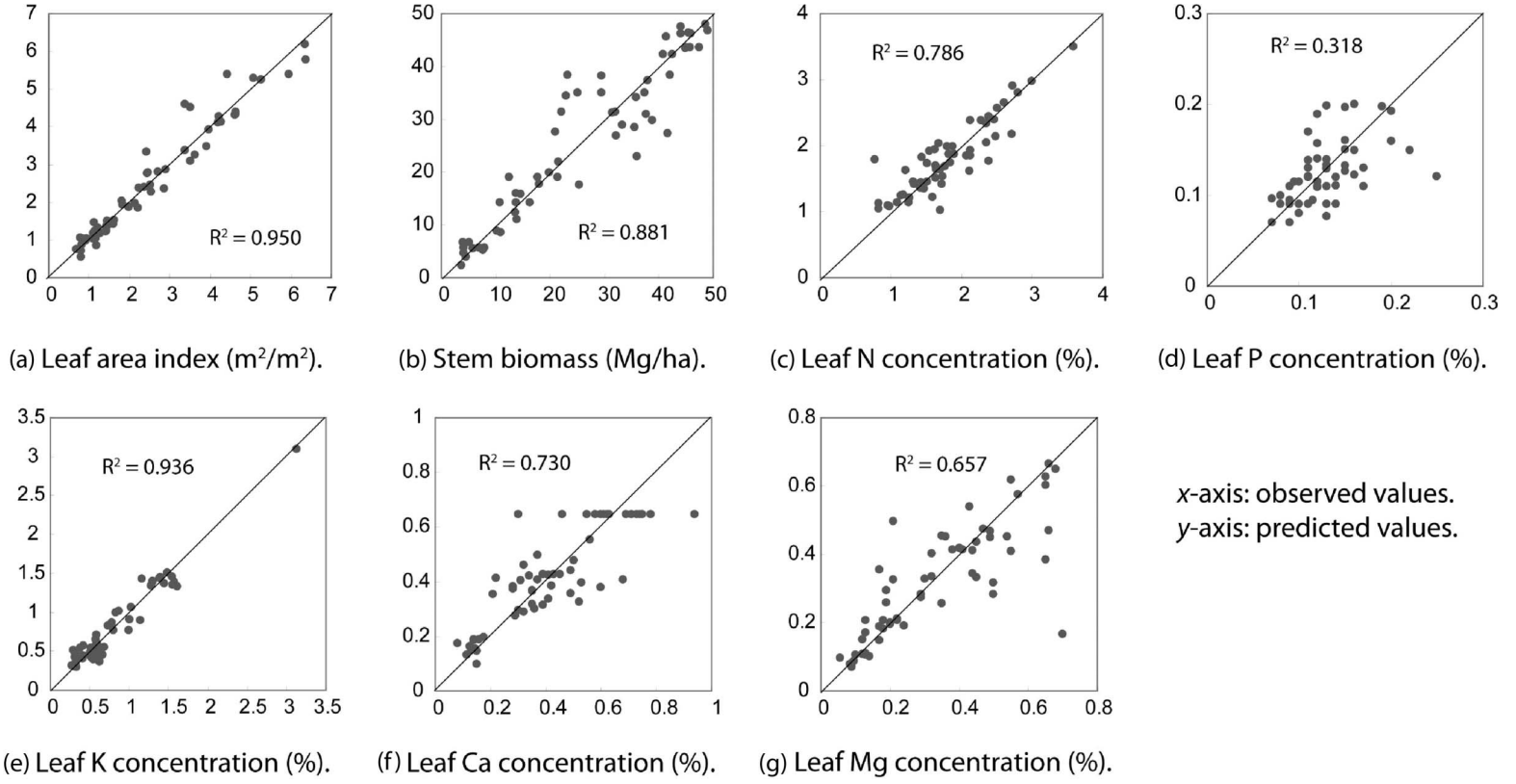
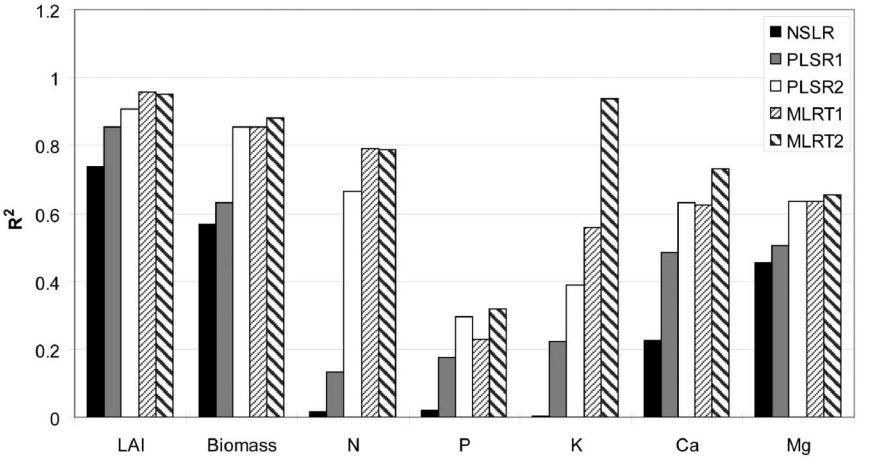
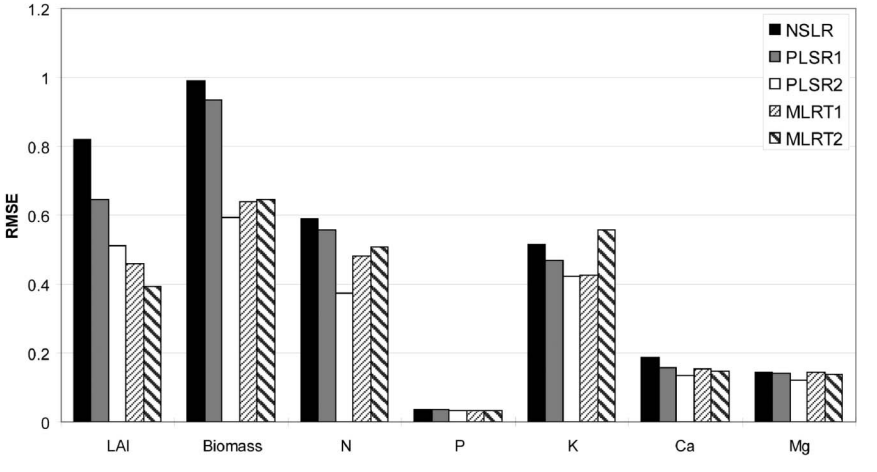


Figure 9. Scatterplots between predicted and observed values of the seven target variables for the MLRT models using the 66 predictors including three discrete variables.



(a) R^2 comparison between the techniques applied for each variable.



(b) *RMSE* comparison between the techniques applied for each variable:
LAI (m^2/m^2), Biomass ($x10^{-1}$ Mg/ha), N (%), P (%), K (%), Ca (%), and Mg (%).

Figure 10. Graphical comparison of calibration and cross-validation performance among the five models.

- The NSLR approach is simple and easy to implement. Selecting two bands to calculate NDVI is very important because each vegetation species generally has a unique spectral reflectance pattern. However, the NSLR approach based on the standard NDVI (i.e. using 856.8 and 657.0 nm) failed to develop strong relationships with the biochemical target variables covering all the species.
- The PLSR approach was generally suitable for predicting the biophysical and biochemical characteristics of the trees. It is noted that selecting the number of components to be used in the model was critical. Using dummy variables to distinguish between the tree genotypes increased the predictive performance of the PLSR model.

- The MLRT models were somewhat limited in the study due to the overfitting problem. However, the adaptability of the MLRT approach, which was shown in the calibration, increased its potential for predicting the target variables. With more generalized and larger samples, the MLRT approach may result in better prediction of the target properties, especially in complex environments like the experimental plots in this study which contained many different tree genotypes and treatments.

Conclusions

Hyperspectral remote sensing technology has advanced vegetation monitoring by allowing the estimation of various vegetation biophysical and biochemical characteristics in an efficient manner. A range of quantitative methods are required to accurately monitor functional health of vegetation from hyperspectral remote sensing data. Research using these methods has heavily relied on data and information obtained only from the site under investigation. To generalize methods such as the approaches adopted in the study, more investigation of the approaches using different remote sensing data and/or plant species as well as larger sizes of *in situ* reference samples is necessary. It will provide a more robust basis for hyperspectral remote sensing monitoring of forest crops by reducing data- and/or site-specific problems.

This study has proven the capability of hyperspectral remote sensing to predict selected forest biophysical and biochemical characteristics in a very complex environment containing five tree genotypes subjected to different levels of irrigation and/or fertilization. The simple linear regression (i.e. NSLR) resulted in good performance in predicting the biophysical characteristics such as LAI and stem biomass for certain genotypes. But, it failed to predict the biochemical characteristics (i.e. leaf nutrients) covering all tree species and treatments. Advanced approaches (i.e. PLSR and MLRT) yielded much better estimation of the leaf nutrients concentrations. Particularly, the MLRT approach with a good size of samples appears to be a robust method for predicting such biophysical and biochemical properties from environmental complex forest environments due to its adaptability to sample data.

The valuable results of the experimental study deserve further research and improvement. A major limitation of the study is the small number of *in situ* samples, limiting comprehensive model evaluation. This limitation is the route for future research, which includes (1) additional investigation of the methods using different data and/or larger sizes of *in situ* samples, (2) examination of optimum band combinations for computing a narrow-band NDVI suitable for monitoring forest health and (3) exploration of other modelling techniques to estimate diverse biophysical and biochemical characteristics of forest crops such as physical models and neural networks.

Acknowledgements

Funding was provided by the US Department of Energy-Savannah River Operations office through the USDA-Forest Service Savannah River and the Forest Service Southern Research Station under Interagency Agreement DE-IA09-76SR00056; Savannah River National Laboratory under Contract Agreement DE-AC09-96SR18500; Department of Energy Oak Ridge National Lab Interagency Agreement 00-IA-11330135-221; Southern Research Station Research Work Units 4103, 4104, 4154, 4155, 4505, 4703, and The Timber Company, Weyerhaeuser, Champion International, and Union Camp. Douglas Aubrey provided database management support for field collected data.

References

Aparicio, N., *et al*., 2000. Spectral vegetation indices as non-destructive tools for determining durum wheat yield. *Agronomy Journal*, 92, 83–91.

Beeri, O., *et al*., 2007. Estimating forage quantity and quality using aerial hyperspectral imagery for northern mixed-grass prairie. *Remote Sensing of Environment*, 110, 216–225.

Blake, T.J., *et al*., 1996. Water relations. *In*: R.F. Stettler, H.D. Bradshaw, P.E. Heilman, and T.M. Hinckley, eds. *Biology of populus and its implications for management and conservation*. Ottawa: NRC Research Press, 401–422.

Breiman, L., *et al*., 1984. *Classification and regression trees*. New York: Chapman and Hall.

Chirici, G., Barbati, A., and Maselli, F., 2007. Modeling of Italian forest net primary productivity by the integration of remotely sensed and GIS data. *Forest Ecology and Management*, 246, 285–295.

Chmura, D.J., Rahman, M.S., and Tjoelker, M.G., 2007. Crown structure and biomass allocation patterns modulate aboveground productivity in young loblolly pine and slash pine. *Forest Ecology and Management*, 243, 219–230.

Coleman, M.D., *et al*., 2004. Production of short-rotation woody crops grown with a range of nutrient and water availability: establishment report and first-year responses. Technical report, SRS-72 Forest Service, US Department of Agriculture.

Coyle, D.R. and Coleman, M.D., 2005. Forest production responses to irrigation and fertilization are not explained by shifts in allocation. *Forest Ecology and Management*, 208, 137–152.

Cho, M.A. and Skidmore, A.K., 2006. A new technique for extracting the red edge position from hyperspectral data: the linear extrapolation method. *Remote Sensing of Environment*, 101, 181–193.

Curran, P.J., Windham, W.R., and Gholz, H.L., 1995. Exploring the relationship between reflectance red edge and chlorophyll content in slash pine leaves. *Tree Physiology*, 15, 203–206.

Ferwerda, J.G. and Skidmore, A.K., 2007. Can nutrient status of four woody plant species be predicted using field spectrometry? *ISPRS Journal of Photogrammetry & Remote Sensing*, 62, 406–424.

Garson, D., 2006. *Partial Least Squares Regression*. [Online]. Retrieved on 6 November 2007 from: http://www.statisticssolutions.com/Partial-Least-Squares-Regression.

Gong, P., Pu, R., and Heald, R.C., 2002. Analysis of *in situ* hyperspectral data for nutrient estimation of giant sequoia. *International Journal of Remote Sensing*, 23, 1827–1850.

Hansen, P.M. and Schjoerring, J.K., 2003. Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sensing of Environment*, 86, 542–553.

Hu, B., *et al*., 2004. Retrieval of crop chlorophyll content and leaf area index from decompressed hyperspectral data: the effects of data compression. *Remote Sensing of Environment*, 92, 139–152.

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

Jensen, J.R., 2005. *Introductory digital image processing*. Upper Saddle River: Prentice-Hall, 526 p.

Jensen, J.R., 2007. *Remote Sensing of the Environment*. Upper Saddle River: Prentice-Hall, 592 p.

Johnson, L.F., Hlavaka, C.A., and Peterson, D.L., 1994. Multivariate analysis of AVIRIS data for canopy biochemical estimation along the Oregon transect. *Remote Sensing of Environment*, 47, 216–230.

Landsberg, J.J., 1986. *Physiological ecology of forest production*. London: Academic Press.

Moisen, G.G., *et al*., 2006. Predicting tree species presence and basal area in Utah: a comparison of stochastic gradient boosting, generalized additive models, and tree-based methods. *Ecological Modeling*, 199, 176–187.

Mutanga, O., Skidmore, A.K., and Prins, H.H., 2003. Predicting *in situ* pasture quality in the Kruger National Park, South Africa, using continuum-removed absorption features. *Remote Sensing of Environment*, 89, 393–408.

Perry, E.M. and Davenport, J.R., 2007. Spectral and spatial differences in response of vegetation indices to nitrogen treatments on apple. *Computers and Electronics in Agriculture*, 59, 56–65.

Rogers, V.A., 1990. *Soil survey of Savannah River plant area, parts of Aiken, Barnwell, and Allendale counties, South Carolina*. Washington DC: USDA Soil Conservation Service.

Samuelson, L., *et al*., 2001. Production efficiency of loblolly pine and sweetgum in response to four years of intensive management. *Tree Physiology*, 21, 369–376.

Stanton, B.B., *et al*., 2002. Hybrid poplar in the Pacific Northwest: the effects of market-driven management. *Journal of Forestry*, 100, 25–33.

Tilling, A.K., *et al*., 2007. Remote sensing of nitrogen and water stress in wheat. *Field Crops Research*, 104, 77–85.

Waheed, T., *et al*., 2006. Measuring performance in precision agriculture: CART – A decision tree approach. *Agricultural Water Management*, 84, 173–185.

Xiao, Y., Jokela, E.J., and White, T.L., 2003. Growth and leaf nutrient responses of loblolly and slash pine families to intensive silvicultural management. *Forest Ecology and Management*, 183, 281–295.

Yamashita, T., *et al*., 2004. Comparison of two coniferous plantations in central Japan with respect to forest productivity, growth phenology and soil nitrogen dynamics. *Forest Ecology and Management*, 200, 215–226.

Yang, L., *et al*., 2003. Urban land cover change detection through sub-pixel imperviousness mapping using remotely sensed data. *Photogrammetric Engineering & Remote Sensing*, 69, 1003–1010.

Ye, X., *et al*., 2006. Estimation of citrus yield from airborne hyperspectral images using a neural network model. *Ecological Modeling*, 198, 426–432.

Zhao, C., *et al*., 2005. Predicting grain protein content of winter wheat using remote sensing data based on nitrogen status and water stress. *International Journal of Applied Earth Observation & Geoinformation*, 7, 1–9.