PREDICTION OF LEAF AREA INDEX FOR SOUTHERN PINE PLANTATIONS FROM SATELLITE IMAGERY USING REGRESSION AND ARTIFICIAL NEURAL NETWORKS

Douglas A. Shoemaker
The Center for Applied Geographical Information Science
University of North Carolina-Charlotte
McEniry 320, 9201 University City Blvd.
Charlotte, NC 28223
E-mail: d.shoemaker@uncc.edu

Wendell P. Cropper, Jr.
School of Forest Resources and Conservation
University of Florida
214 Newins-Ziegler Hall
P.O. BOX 110410
Gainesville, FL 32611
E-mail: wcropper@ufl.edu

Abstract

Leaf Area Index (LAI) is an important predictor of southern pine forest productivity. In this study several models for estimating LAI values using remote sensing were developed and tested. Of these, a best model was selected based on performance and potential for operational application. The generalized southern pine LAI predictive model (GSP-LAI) was developed using artificial neural network (ANN) multivariate regression that incorporated important local information including phenological and climatic data. In validation tests the model explained > 75% of variance ($R^2 = 0.77$) with an RMSE < 0.50. The GSP-LAI model was applied to Landsat ETM+ image of the Bradford forest, north of Waldo, FL. Within the extent are substantial slash (*Pinus elliottii*) and loblolly (*Pinus taeda*) pine plantations. Based on image and stand data LAI values for 10,797 ha were estimated to range between 0 and 3.93 $m^2\ m^{-2}$ with a mean of 1.53 $m^2\ m^{-2}$.

Keywords. Landsat, LAI prediction, slash pine, loblolly pine, remote sensing

Introduction

Pine plantations of the southeastern United States constitute a large fraction of the world's industrial forests. In Florida alone, annual timber revenue exceeds $16 billion and is the dominant agricultural sector (Hodges et al., 2005). Although managing these forests for maximum yield is a primary economic goal, the rate at which these forests remove and sequester atmospheric carbon is of increasing interest (Hoen and Solberg, 1994; Stainback and Alavalapati, 2005; Litynski et al., 2006). Remote sensing coupled with modeling carbon dynamics could be

In Proceedings of the 6th Southern Forestry and Natural Resources GIS Conference (2008), P. Bettinger, K. Merry, S. Fei, J. Drake, N. Nibbelink, and J. Hepinstall, eds. Warnell School of Forestry and Natural Resources, University of Georgia, Athens, GA.

used for monitoring and verification of forest carbon storage rates (Prisley and Mortimer, 2004; Turner et al., 2004; Binford et al., 2006).

Leaf area index (LAI) is a key parameter for estimation of pine plantation productivity or net ecosystem exchange of carbon (NEE). In this study we focus on the estimation of LAI, a primary biophysical parameter used for forest productivity (NEE) modeling and carbon sequestration studies, and used by forest managers seeking to quantify canopy responses to silvicultural treatments (Cropper and Gholz, 1993; Gower et al., 1999; Reich et al., 1999, Cropper, 2000). LAI is the ratio of leaf surface area supported by a plant to its corresponding horizontal projection on the ground, and it is difficult and expensive to assess *in situ*, resulting in sparse sample sets that are necessarily localized at the stand scale, and thus difficult to extrapolate to larger extents (Fassnacht et al., 1997).

Determination of LAI from remotely sensed data has the advantages of being spatially explicit, scaleable from stand to regional or larger extents, and applicable to remote or inaccessible areas (Running et al., 1986). An ideal empirical model linking ground-referenced LAI to remote sensing data would make reliable predictions at various extents and image dates, and be general enough to incorporate important local information, such as climatological and phenological data.

As Gobron et al. (1997) point out, the range of variation that exists in global plant communities precludes the likelihood of a single universal relationship between LAI and remote sensing products. However, regional prediction of LAI in a relatively uniform system such as the extensive and economically important industrial pine plantations across the southeastern U.S., should be possible with reasonable accuracy.

There have been previous attempts to remotely estimate LAI for this forest type. Industrial plantations in the south typically consist of dense plantings of loblolly pine (*Pinus taeda*) and slash pine (*Pinus elliottii*) (Prestemon and Abt, 2002). Gholz et al. (1991) and Curran et al. (1992) studied a north-central Florida mature slash pine plantation where they correlated LAI values with remote sensing data collected by Landsat TM. Flores (2003) related ground-based indirect LAI values to hyperspectral remote sensing data for loblolly pine in North Carolina.

These studies used ordinary least squares (OLS) regression analysis to establish an empirical relationship between vegetative indices (VI) and ground-referenced LAI. The best understood VIs are the normalized difference vegetative index (NDVI) and the simple ratio (SR) (Birth and Mcvey, 1968), both of which make use of recorded values for the red and near infrared wavelengths. In the case of Gholz et al. (1991), three predictive equations were produced using NDVI. Flores used SR as the predictive variable for loblolly pine plantation LAI. We evaluated these models using a new dataset assembled for this study and found none exhibited significant predictive ability. While linear regression remains a popular approach, variations in surface and atmospheric conditions, as well as the structural considerations of satellite remote sensing, have prevented establishment of a universal relationship between LAI and VIs (Gobron et al., 1997; Fang and Liang, 2003).

Perhaps this failure is due to under- or mis-specification of the models. The biochemical and structural components of a forest canopy are complex, varying in both time and space (Raffy et al., 2003). Cohen et al. (2003) suggest the incorporation of other recorded spectra and the use of data from multiple dates as predictive variables, in order to improve regression analysis results. Multivariate regression techniques allow for the incorporation of more types of data, including important locational information, such as climate or categorical stand data. When OLS regression is used, variable selection techniques permit the exploration of a wide range of data for significance.

Despite these advantages, many of the assumptions necessary for OLS regression are violated by using remote sensing data that characteristically exhibits non-normality, and thus tends to suffer multicollinearity and autocorrelation. For this reason a nonparametric technique, regression with artificial neural networks (ANN), was investigated as an alternative to OLS regression.

Artificial neural networks are loosely modeled on brain function: a series of nodes representing inputs, outputs, and internal variables are connected by synapses of varying strength and connectivity (Jensen et al., 1999). The network architecture is typically oriented as a perceptron, which learns by passing information from inputs to outputs (forward propagation) and from output to inputs (back propagation) to optimize the accuracy of prediction by adjusting weights (Mehrotra et al., 2000). The ability to accommodate complexity can be made by altering the construction of the network to include multiple layers of internal nodes. These networks are robust in that many of the assumptions needed for OLS regression are relaxed, including the requirements of normality and independence.

Artificial neural networks have been coupled with remote sensing for a variety of applications in forestry. ANN models have been used to predict forest biomass (Foody et al., 2003), basal area and stem density and other stand structural characteristics (Corne et al., 2004; Ingram et al., 2005), and habitat suitability (Hilbert and Ostendorf, 2001) using a variety of remote sensing data. Nonparametric techniques such as ANN have demonstrated great potential for estimation of LAI. ANN models have been developed for LAI estimation from data such as Landsat Enhanced Thematic Mapper Plus (Fang and Liang, 2003), the Advance Spaceborne Thermal Emission and Reflection Radiometer (Hardin and Jensen, 2005), the Moderate Resolution Imaging Spectroradiometer (Fang and Liang, 2005), and hyperspectral airborne platforms (Schlerf and Atzberger, 2006).

In this study our objective was to develop a single general empirical model capable of producing reliable LAI predictions at various extents and image dates. We hypothesized that such a solution would require multivariate statistics to recognize and incorporate important local information such as climatological and phenological data. Three regression techniques, linear OLS, multiple OLS, and ANN were applied to a large dataset acquired by Landsat sensors over a 10-year period, and the resultant models were evaluated for performance using a validation process.

Methods

Study Sites

Two plantations of southern pine were used to develop LAI predictive models for this study: the Intensive Management Practices Assessment Center (IMPAC) operated by the University of

Florida's Forest Biology Research Cooperative (FBRC), and the Donaldson tract, part of the Bradford Forest managed by Rayonier, Inc. and currently part of a Florida Ameriflux eddy covariance monitoring program (Powell et al., 2005). Both plantations are regenerated with two southern pine species (loblolly and slash), which have similar physiology and seasonal foliage dynamics (Gholz et al., 1991).

The IMPAC site is located 10 km north of Gainesville, Florida USA (29°40' N, 82°20' W, Figure 1) The site is flat, with elevations varying less than 2 m, and experiences a mean annual

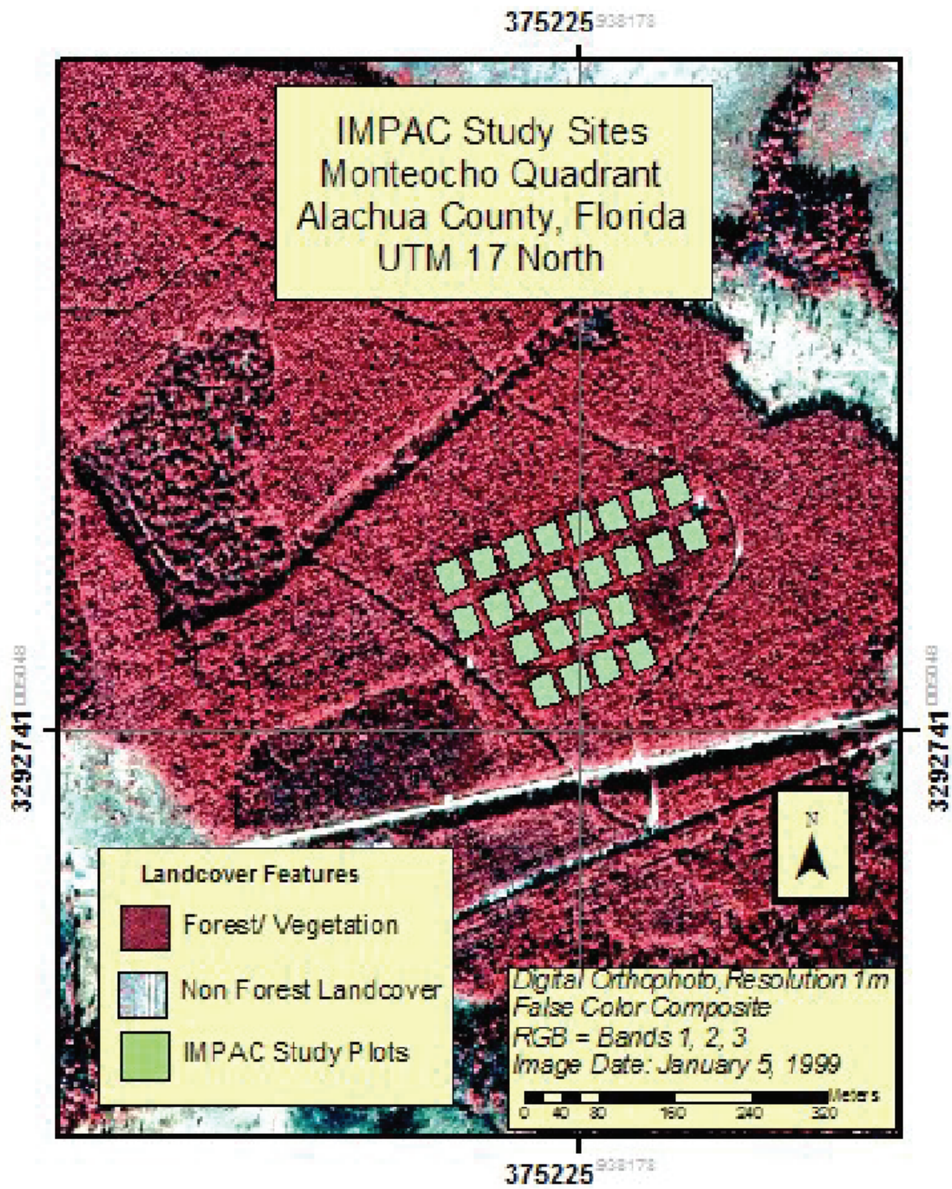

Figure 1. Map of the Intensive Management Practices Assessment Center, Alachua County, Florida, USA.

temperature of 21.7° C and 1,320 mm annual rainfall. Soils are characterized as sandy, siliceous hyperthermic Ultic Alaquods (Swindle et al., 1988). The stand was established in 1983 at a stocking rate of 1,495 seedlings per hectare; a dense planting typical of industrial pine plantations. The site was surveyed using a differential global positioning system (DGPS) in February 2004.

The IMPAC site consists of 24 study plots, each 850 m^2, treated with factorial combinations of species (loblolly and slash pine), fertilization (annual or none) and control of understory vegetation (sustained or none) in three replicates. Fertilization of respective plots occurred annually for ages 1-11, was ceased for ages 12-15, and resumed at age 16.

The Donaldson tract is located 12 km east of the IMPAC site (29°48' N, 82°12' W). The stand was established in 1989 and stocked at a rate of 1,789 slash pine seedlings per hectare. The site is flat and well-drained. Within the stand are four 2,500 m^2 plots from which leaf litterfall was collected starting at age 10 (1999). Plots were surveyed with GPS May 2002. Estimates of LAI based on needlefall (Gholz et al., 1991) from 10 randomly located traps were combined into a single value for all four plots.

<u>Remote Sensing Data</u>

For this study, we acquired 18 cloudless images recorded of the study area captured between 1991 and 2001 with Landsat 5 and 7 satellites (Table 1). This series of images is concurrent with significant cycles of dry and wet periods for the region (Figure 2) and contains examples of each of four canopy development phenological categories (Figure 3).

Images were captured with the both the Thematic Mapper (TM) sensor aboard Landsat 5 and the Enhanced Thematic Mapper Plus (ETM+) aboard Landsat 7. These sensors are functionally identical for the bandwidths used in the study: visible spectra blue (0.45-0.52 µm, or band 1), green (0.52-0.60 µm, or band 2), red (0.60-0.63 µm, or band 3), and three infrared spectra: near (0.69-0.76 µm, or band 4), mid (1.55-1.75 µm, or band 5), and reflected thermal (2.08-2.35 µm, or band 7). The spatial resolution of the Landsat data is generally 30 m. Band 6, which detects emitted thermal radiance between 10.5 and 12.5 µm, has a spatial resolution of 120 m for TM and 60 m for ETM+.

Brightness values (BV) were recovered from the data based on the center point of each study plot. All images were individually rectified using a second order polynomial equation with between 30 and 40 ground control points; while the images maintained the accepted rectification accuracy of 0.5 pixels, the accuracy of the overlay with study plots varied from image to image.

<u>Seasonal LAI Dynamics and Leaf Litterfall Data</u>

Loblolly and slash pines are evergreen trees that maintain two age classes of leaves throughout much of the year, needles produced in both the previous and current growing seasons (Gholz et al., 1991; Curran et al., 1992; Teskey et al., 1994). In north Florida, these classes overlap between July and September, establishing a period of peak leaf area categorized as maximum LAI. As such, the phenological year is typically categorized into four periods: minimum LAI,

Table 1. Catalog of images used in study (All images are Path 17N, row 39. Datum NAD83/GRS 80).

Image date	Sensor	Phenological period	PHDI
1/17/1991	TM	Declining LAI	-1.75
3/22/1991	TM	Minimum LAI	-0.63
10/16/1991	TM	Declining LAI	2.63
1/20/1992	TM	Declining LAI	1.59
8/31/1992	TM	Maximum LAI	1.22
3/27/1993	TM	Minimum LAI	1.85
8/18/1993	TM	Maximum LAI	-2.76
1/25/1994	TM	Minimum LAI	2.78
9/6/1994	TM	Maximum LAI	1.30
6/7/1996	TM	Expanding LAI	0.83
9/30/1997	TM	Maximum LAI	-0.86
6/29/1998	TM	Expanding LAI	0.59
1/7/1999	TM	Minimum LAI	-1.90
9/4/1999	TM	Maximum LAI	-2.38
1/2/2000	ETM+	Declining LAI	-2.29
4/7/2000	ETM+	Minimum LAI	-2.71
8/13/2000	ETM+	Maximum LAI	-4.02
1/4/2001	ETM+	Declining LAI	-3.05

LAI = Leaf area index
PHDI = Palmer hydrological drought index

leaf expansion, maximum LAI, and declining LAI (Figure 3). This dynamic must be well understood to interpret LAI from remotely sensed data.

In situ estimates of LAI were calculated by leaf litterfall collection. Needlefall was collected monthly from six 0.7 m^2 traps distributed randomly within each of the 24 IMPAC study plots from year 8 (1991) through 2001. A similar method was used at the Donaldson site's four study plots, the results of which were aggregated into a single value for the tract. LAI from litterfall was estimated using foliage accretion models (Martin and Jokela, 2004). LAI results were presented as hemi-surface leaf area and converted to projected leaf area for integration with remote sensing data.

Integration of Ground Referenced LAI and Remote Sensing Data

LAI data based on monthly leaf litterfall collection from all 24 study plots was ground referenced to plot centroids based on GPS surveys. Landsat images were overlaid with plot

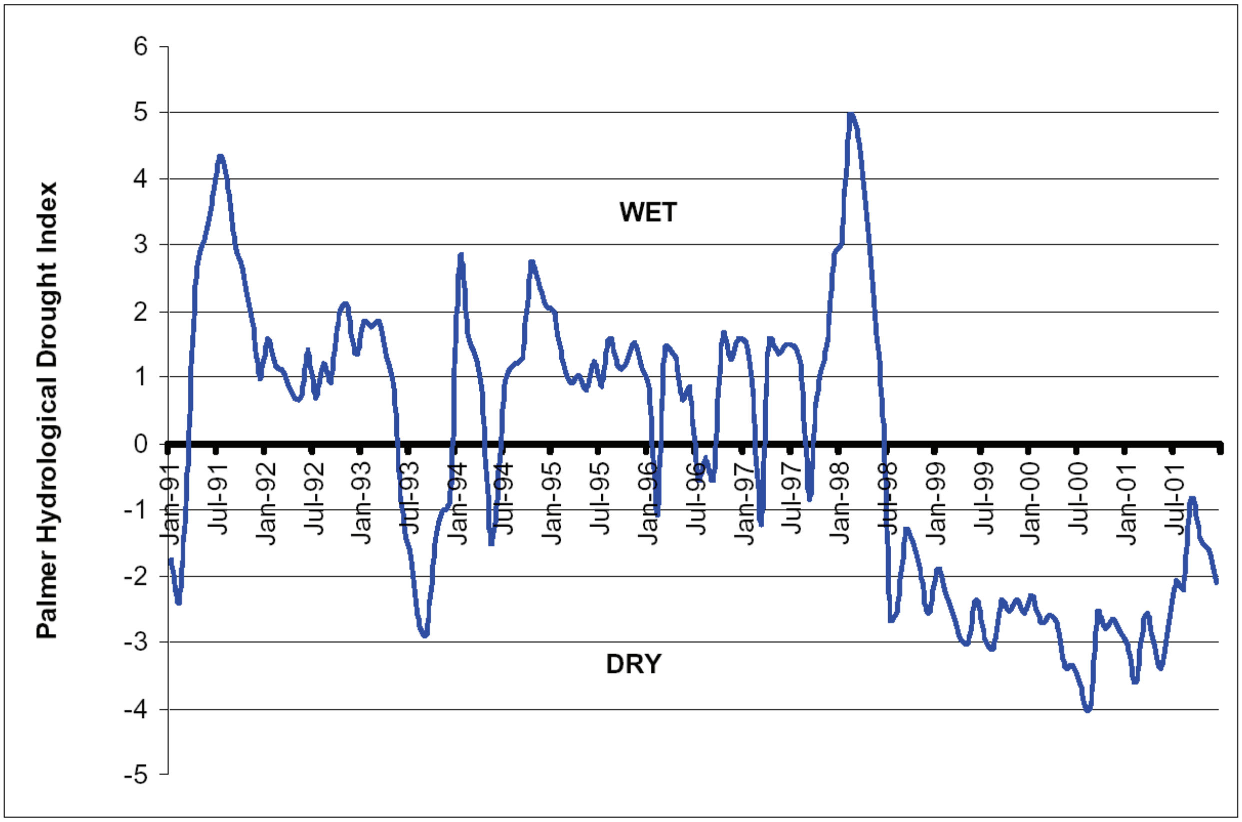

Figure 2. Palmer Hydrological Drought Index for North-Central Florida climate during study period 1991-2001 (National Climatic Data Center 2005).

locations within a geographic information system (GIS). Surface reflectance data and ground referenced LAI were related by a point method which joined LAI values to pixels based on the presence of a plot centroid. LAI data, aggregated monthly, were matched with image date based on proximity.

The integration resulted in a dataset based on the point method of 453 samples, which linked 28 locations with their respective surface reflectance values at specific times over a period of 11 years. All rows were randomized within the table and 51 cases were extracted and withheld for external validation.

The data were densified with vegetative indices including normalized difference vegetative index (Birth and Mcvey, 1968), simple ratio (Crist and Cicone, 1984) and tasseled cap analysis components (Crist and Cicone, 1984). Ancillary data incorporated into the set included climate indices and categorical plot data representing species type, plot treatment, and phenological period. The complete list of variables used in modeling is included in Appendix A.

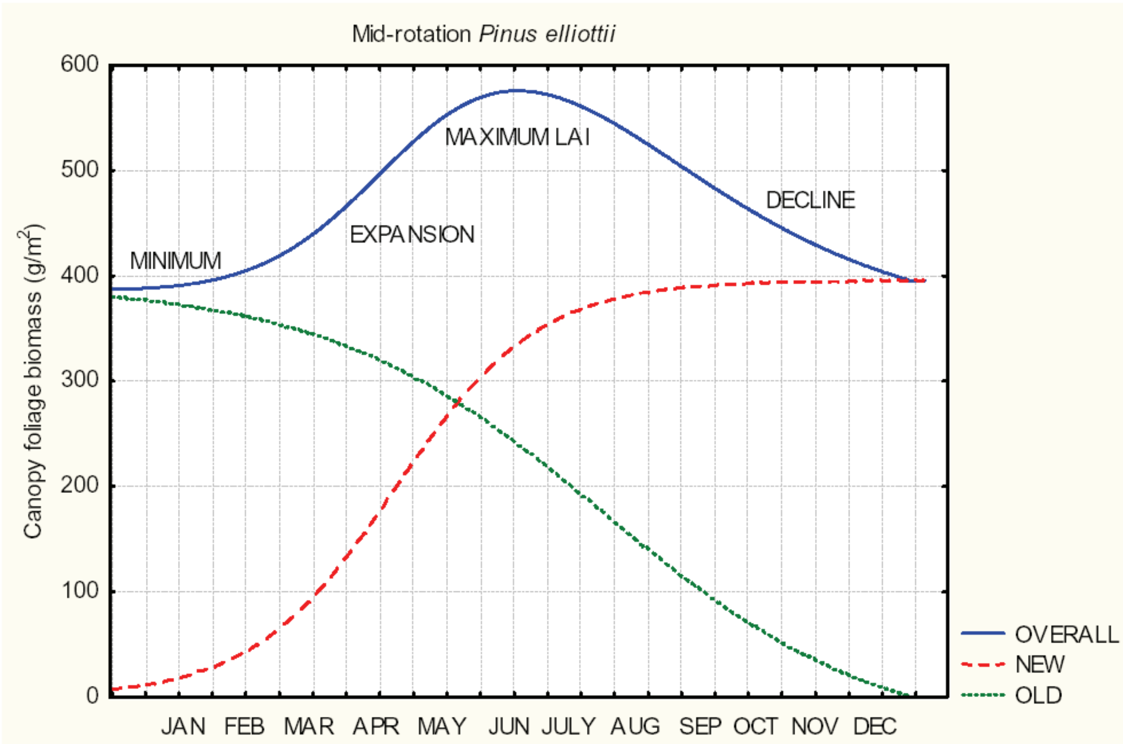

Figure 3. Annual cycle of variation in leaf phenology in southern pines illustrating the typical two populations of needles (after Cropper and Gholz, 1993).

<u>Climate Variables</u>

Local site water balance was represented by the Palmer Hydrological Drought Index (PHDI), a monthly index used to access long-term influences on plant productivity (Karl and Knight, 1985; National Climatic Data Center, 2005). The variety of indices developed by Palmer and others standardize climatic indicators to allow comparison of site water balance at different times and locations. The PHDI was used instead of the better-known rainfall-based Palmer Drought Severity Index (PDSI) as it accounts for local outflows and storage of water based on short-term trends.

Time scales at which climate influences leaf area are not completely understood. Therefore several variables were developed to explore specific lags: a) a simple annual lag, b) the PHDI summation values during the leaf expansion period, c) a summation with an annual lag, and finally a summation for PHDI during leaf expansion for current and previous growing years. This last variable is an attempt to capture the cumulative effect of climate when represented by two age classes of needles present during the maximum LAI period.

Statistical Analysis

Statistical analysis was performed on the integrated data set, including descriptive, principal component, and autocorrelation analysis, using NCSS statistical software (Hintze, 2001). The likelihood of spatial autocorrelation was explored using GEODA 0.9.5 geostatistical software (Anselin, 2003). Three types of regression processes were evaluated; two were based on ordinary least squares (OLS), and the third was based on an artificial neural network (ANN) approach.

OLS Regression

Linear regression represents the standard form of OLS regression where a single independent variable, often a vegetative index, was regressed against the dependent variable LAI. Linear regression has been the typical approach in previous studies including Gholz et al. (1991) using NDVI and Flores (2003) using SR. In the multiple form of OLS regression, many independent variables, including surface reflectance data, vegetative indices, climate data, and categorical data were regressed against the dependent variable LAI. Stepwise variable selection was used to identify variables significant at p-value < 0.05.

Artificial Neural Network

Construction and processing of ANNs was accomplished with the neural network module of Statistica statistical software (StatSoft, 2004). Architectures were limited to Multilayer Perceptron (MLP) with a maximum of four hidden layers as suggested by Jensen et al., (1999). Sample sets were bootstrapped based on available cases. One hundred architectures were evaluated per model, with the top 5 retained based on the lowest ratio of standard deviation between residuals and observation data. From these five, a best model was selected based on the relationship between predicted and observed values from the training and validation sets (R^2, RMSE).

The ANN used to calculate LAI values consisted of a network with 18 inputs and two fully connected hidden layers with 16 and 7 nodes respectively. The MLP was trained using back propagation, in which network weights are modified to reduce error through gradient descent (Mehrota et al., 2000). Each input value was transformed prior to network calculations. Continuous variables were transformed with two-parameter linear equations, and categorical inputs (0 or 1) were replaced with constant values between 0 and 1. The input vector *VI* was constructed by appending a value of 1.0 (representing the bias or threshold value) onto the list of transformed input values. The first hidden layer weights are then applied:

$$VO1 = VI\, W1 \qquad (1)$$

where *W1* is a 19 row, 16 column matrix of hidden layer one weights and *VO1* is the output vector from the first hidden layer. Each of the values (*i*) in *VO1* then has a sigmoid activation function applied to it:

$$\frac{e^{VO1_i} - e^{-VO1_i}}{e^{VO1_i} + e^{-VO1_i}} \tag{2}$$

Similarly, vector *VO1* is then multiplied by the second hidden layer weight matrix (*W2*, 17 rows and 7 columns). The output vector (*VO2*, with a threshold value of 1.0 appended to it) is filtered through the sigmoidal activation function (eq. 2). Finally, a scaled LAI value is calculated as follows:

$$\mathrm{LAI_s} = \sum_{i=1}^{n+1} \mathrm{VO2_i} \cdot \mathrm{W3_i} \tag{3}$$

where *W3* is a vector of weights and *n* is the number of nodes in the second hidden layer. LAI can be calculated from the scaled value as follows:

$$\mathrm{LAI} = \frac{\mathrm{LAI_s} + 0.09978868}{0.2386889} \tag{4}$$

Three classes of multiple regression models are evaluated in this work: (1) simple models whose constituent variables are generated solely from remote sensing data and corresponding vegetation indices only; (2) intermediate models that additionally incorporate image date (and therefore phenological information) and climate data; and (3) the most complex models that add stand level data such as species and treatment. Following the precedent set by Gholz et al. (1991), the simple and intermediate models sets were developed for single species and single phenological periods.

Results

LAI values from leaf litterfall collection vary from just under 0.5 to 4.5 with a mean of 2.38 m^2 m^{-2}. There is considerable overlap in LAI for slash and loblolly (Figure 4). Loblolly pine leaf production was slightly more sensitive to fertilization, with an increase of 1.0 in mean LAI as compared to 0.56 for slash (Figure 5).

One of the limitations of relating LAI to remote sensing data is spatial autocorrelation. For example Band 6, which detects emitted thermal radiation, was removed from analysis as it exhibited significant spatial autocorrelation (Moran's I = 0.53), likely due to its coarse resolution of 120 m (Landsat TM), and resulting extent, which overlays several plots at once. Spatial autocorrelation was not indicated for the reflectance values of the other 5 bands and LAI (Morans I = 0.03 and -0.02 respectively).

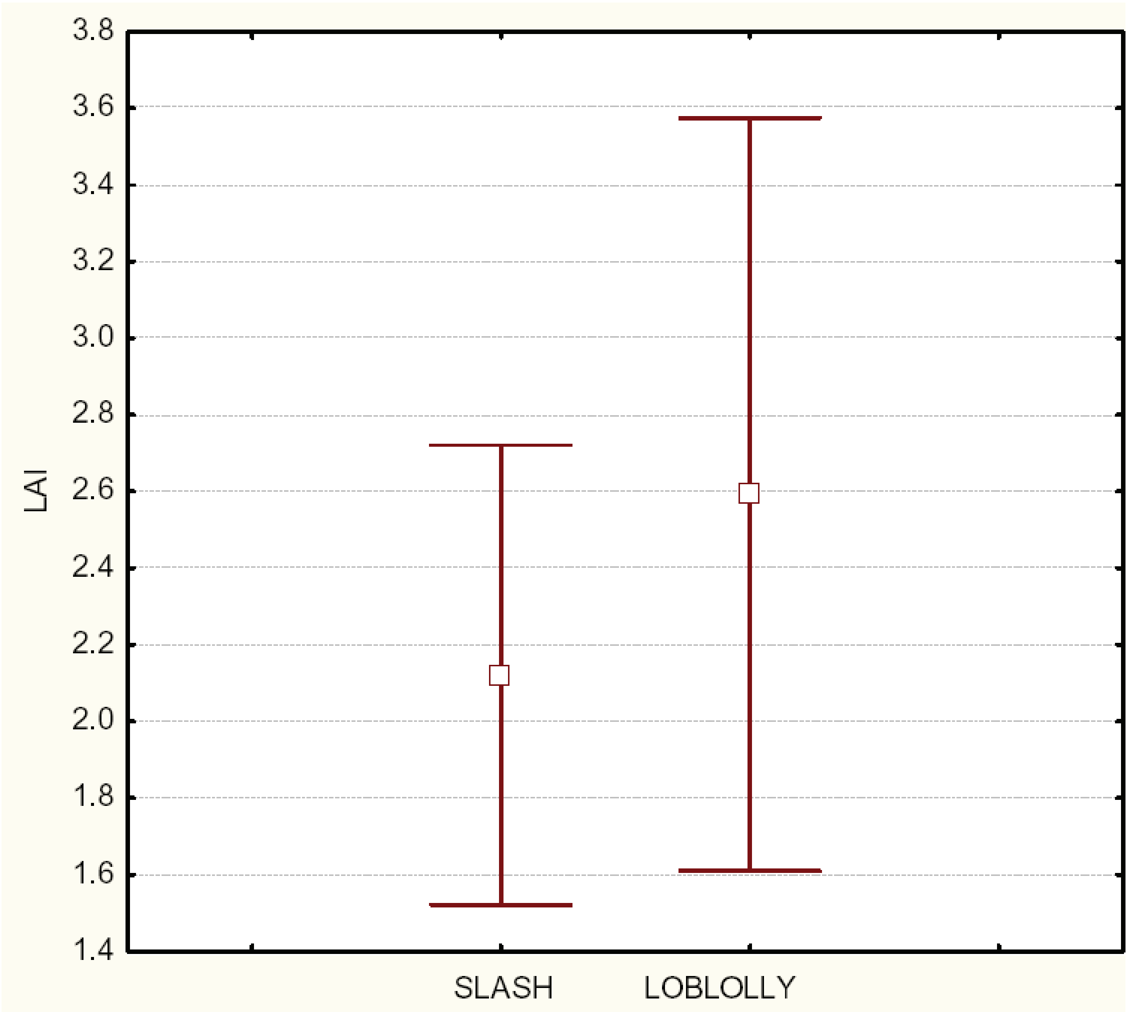

Figure 4. The mean and range of estimated (litterfall-based) LAI values for slash and loblolly pine for all sites, 1991-2001.

Principal component analysis of spectral variables used revealed eigenvalues near 0.0 for 5 of the 9 resultant components, indicating multicollinearity. When two or more of the independent variables of a multiple regression are correlated, the data is said to exhibit multicollinearity. Multicollinearity may result in wide confidence intervals on regression coefficients. In general, the simplest possible predictive model is desirable. Simpler models are easier to apply to new cases because of the reduced requirements for input data. Complex environmental systems with multiple interacting biological and physical components are, however, not likely to be adequately modeled by the simplest models. In this study we have examined a range of models from simple linear models through non-linear ANN multiple regression models. Our goal was to find a model that was a good predictor for separate validation data. The latter requirement was necessary as a guard against overtraining (Mehrotra et al., 2000).

Figure 5. Differences in effect of fertilizer treatment on slash and loblolly pine. Mean and range LAI values are indicated.

Linear Models

Of the 20 models tested 16 failed to reject the null hypothesis $\beta_1 = 0$. No model exceeded an R^2 = 0.12. These simple models were not adequate predictors of LAI for the training data. Even the published models with a history of useful predictors of southern pine LAI failed for this dataset.

Multiple Regression Models

All models tested were statistically significant for slope, representing improvements over linear models. R^2 values ranged from 0.31 to 0.70. In validation testing, increasingly complex models accounted for greater variation in LAI for training data, but the performance using test data was mixed (Table 2). ANOVA analysis of significant variables appears in Table 3. Significant variables include presence or absence of fertilization treatments and phenological periods.

Table 2. Summary of OLS multiple regression models fitted to dataset.

Class	Label	Species	*n*	R^2	RMSE
		(Training)			
Remote sensing data only	PASEND	slash	79	0.31	0.459
	PALEND	loblolly	74	0.33	0.707
Categorical and climate data	PBSTOT	slash	208	0.42	0.497
	PBLTOT	loblolly	194	0.43	0.794
General model	PCTOT	both	402	0.70	0.49
		(Testing)			
Remote sensing data only	PASEND	slash	13	0.51	0.37
	PALEND	loblolly	8	0.05	1.05
Categorical and climate data	PBSTOT	slash	27	0.02	1.40
	PBLTOT	loblolly	20	0.17	0.92
General model	PCTOT	both	47	0.63	1.97

PASEND LAI = -0.54 + 0.0057 (B1) - 0.00527 (B5) + 0.00808 (TCA-2)

PALEND LAI = -2.48 + 1.23 (SR) + 0.11 (TCA-3)

PBSTOT LAI = 2.35 - 0.79 (EXP) - 0.045 (LAG-PHDI) - 0.63 (MAX) - 0.40 (MIN) - 6.32 (NDVI) + 0.06 (PHDI) + 1.20 (SR) + 0.06 (TCA-3)

PBLTOT LAI = 2.04 - 1.03 (EXP) - 0.74 (MAX) - 0.68 (MIN) - 14.78 (NDVI) + 3.02 (SR) + 0.09 (TCA-3)

PCTOT LAI = 4.48 - 1.038 (EXP) - 0.902 (FERT) - 0.508 (HERB) - 0.835 (MAX) - 0.515 (SPP) + 0.0308 (TCA-3)

B1= Band 1; B5= Band 5; TCA-2, 3= Tassel cap analysis component 2, 3; SR= Simple ratio vegetative index; MIN, EXP, MAX= phenological period: minimum LAI, expanding LAI, maximum LAI; PHDI= Palmer hydrological drought index; LAG-PHDI= PHDI one year previous; NDVI= Normalized difference vegetative index; FERT= Fertilization; HERB= Herbicide application; SPP= Species of tree. Details about the variables are contained in Appendix A.

Table 3. ANOVA analysis of highly significant variables in OLS multiple regression. (Other, less significant variables are not shown.)

Model	Variable	R^2	F-ratio	P-value	Power (5%)
PCTOT	FERT	0.22	286.14	< 0.0001	1.0000
	MAX	0.16	209.67	< 0.0001	1.0000
	EXP	0.11	136.95	< 0.0001	1.0000
PBSTOT	MAX	0.15	51.48	< 0.0001	1.0000
	TCA-3	0.09	31.17	< 0.0001	0.9998
PBLTOT	TCA-3	0.16	51.31	< 0.0001	1.0000
	MAX	0.08	26.71	< 0.0001	0.9993
PASEND	TCA-2	0.22	23.39	< 0.0001	0.9976
PALEND	SR	0.22	23.09	< 0.0001	0.9973
	TCA-3	0.21	21.76	< 0.0001	0.9959
	B1	0.14	15.60	0.0002	0.9737

ANN Models

The ANN predictions significantly improved on OLS multiple regression. R^2 values ranged from 0.40 to 0.85 in training validation, and from 0.02 to 0.77 using the test data (Table 4). The generalized southern pine LAI predictive model (GSP-LAI) was selected as the top-performing model (Figure 6). In validation tests the model explained > 75% of variance ($R^2 = 0.77$) with an RMSE < 0.50.

Discussion

In this study, we created GSP-LAI, a model that effectively predicted LAI for southern pine plantations comprised of slash or loblolly pine. The model's development was guided by three major factors: 1) a focus on a relatively simple and well understood forest system for which there was ample data, 2) a desire to create an operational solution with wide applicability, and 3) the willingness to employ sophisticated regression techniques.

The intensively managed pine plantation is a simpler system than natural pine forests or mixed coniferous / deciduous forests in terms of species diversity and the reduction of canopy layers (Gholz et al., 1991). Although seemingly an ideal system for LAI prediction, previously published southern pine LAI predictors applied to new remote sensing data lead to results so inaccurate as to be unusable as inputs for forest productivity modeling. New simple linear regression models constructed using single vegetative indices and trained on the study's large database offered no improvement.

Table 4. Summary of ANN models fitted to dataset.

Class	Label	Inputs	Hidden Layers	Nodes per layer	Training species	n^a	R^2	RMSE
				(Training)				
Remote sensing data only	ASEND5	6	2	16, 12	slash	79	0.40	0.422
	ALEND9	7	1	4	loblolly	74	0.40	0.650
Include categorical and climate data	BSCLIM10	14	2	16, 6	slash	213	0.42	0.490
	BLCLIM5	15	2	16, 7	loblolly	190	0.49	0.784
General model	GSP-LAI	18	2	16, 7	both	402	0.85	0.347
				(Testing)				
Remote sensing data only	ASEND5	6	2	16, 12	slash	26	0.02	1.10
	ALEND9	7	1	4	loblolly	18	0.26	1.30
Include categorical and climate data	BSCLIM10	14	2	16, 6	slash	27	0.39	0.52
	BLCLIM5	15	2	16, 7	loblolly	24	0.12	0.94
General model	GSP-LAI	18	2	16, 7	both	51	0.77	0.40

[a] Number of cases available for bootstrap sampling.

It is unclear whether the previously published models were ever intended for use outside of the image from which they were created; they were developed with relatively few samples and with few sample dates. Climate history and leaf phenology would necessarily differ from remote sensing data that were used for model calibration. These shortcomings lead to a criterion that LAI estimation should not be limited to a single image, location, phenological period, or satellite sensor. The poor performance of linear regression techniques when applied to a robust dataset

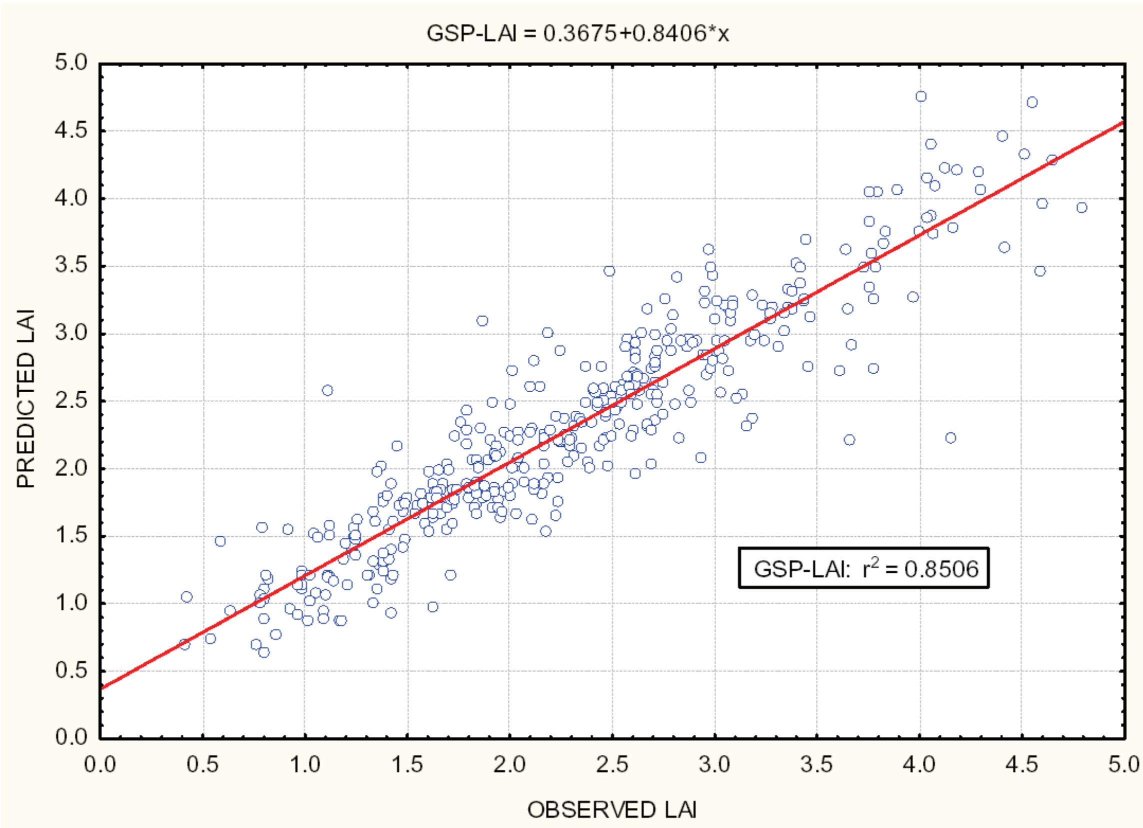

Figure 6. Measured and predicted LAI values from the GSP-LAI for training data.

lead us to the conclusion that even this simple system was too complex to be modeled by a single variable.

In addition to the remote sensing data there are many variables that might be useful predictors of the system, including measures of climate, histories of management treatments such as fertilization, the presence and contribution of understory, the phenological period, and others. To incorporate these variables, multivariate regression techniques were necessary.

The availability of ANN regression functions in modern statistical software allowed for the quick exploration of predictive networks to compare against OLS regression results. In both OLS and ANN regression results, the highest performing models were the most general, capable of incorporating both continuous and categorical variables into a single solution. The assignment of categorical variables is a useful and under-exploited technique, permitting the development of models with wide domains of application.

OLS Multiple Regression Models

OLS regression revealed some of the probable drivers of this system, namely phenological period and management treatment. Tassel cap component 3 was the only consistent remote sensing variable used between models (Table 3). This component, also known as “wetness,” is typically associated with evapotranspiration (ET), which is expected to increase with increased LAI. Tassel cap components are the product of coefficients for all 6 bands of reflected radiation that TM and ETM+ record and as such exploit more spectra than the commonly used NDVI and SR (Cohen et al., 2003).

ANN Models

The best general model was an ANN regression. This non-parametric technique was able to incorporate climate data as represented by PHDI and its lagged derivatives. Climate, while assumed to be important, is typically ignored in empirical LAI models. It is a difficult problem: eligible satellite images are all captured on sunny days, and the various temporal scales on which local climate influences vegetation are mostly unknown, and likely to be species- and site-specific. Typical data used in multitemporal analyses exhibit serial autocorrelation, necessitating transformations in order to become valid OLS inputs. The improved performance conveyed by the ANN regression suggests that 1) climatic variables are significant, and 2) OLS regression was unable to use the variables as employed. The GSP-LAI model is deterministic and easily implemented. The Python code for the model is available from the corresponding author.

Fertilization

In the OLS and ANN generalized models fertilization represents a significant variable (Tables 4 and 5). This result supports observations (Figure 5) and also Martin and Jokela’s (2004) analysis of IMPAC leaf litterfall data. Fertilization is a focal treatment in intensive management practices, and indications of canopy response in the form of LAI could direct the location and frequency of application. The availability of reliable LAI data could lead to a paradigm change in management practices were the goal becomes optimization of leaf growth based on site potential.

The improved performance of increasingly complex models provides insight into variables, which drive or improve the predictability of LAI. The climate variables are particularly interesting in that they are widely assumed to play a role in canopy appearance and yet are rarely incorporated in empirical analysis. Difficulties exist in how to characterize climate, i.e. in terms of rainfall or temperature, and on what temporal scales it operates. Climate data necessarily suffers serial autocorrelation, a violation of an assumption required for OLS regression.

The effectiveness of LAI predictions would be enhanced with a reduction of time between the acquisition of remote sensing data and its analysis. The use of ground-referenced LAI from litterfall necessitates an 18-month lag in processing from collection to value. Using optical methods to indirectly measure LAI *in situ* would likely reduce this lag, provided corrections as suggested by Gower et al., (1999) were incorporated to improve accuracy. With minor modification, the GSP-LAI model can be adapted to new remote sensor data that share ‘legacy’ characteristics with Landsat TM and ETM+.

Table 5. Sensitivity analysis of variables used in ANN multiple regressions.

	Model[a]				
Rank	ASEND	ALEND	BSCLIM	BLCLIM	GSP-LAI
1	TCA-1	B2	B1	END	END
2	B5	TCA-3	B4	MAX	FERT
3	B4	TCA-1	EXP-PHDI	B7	B2
4	B2	SR	TCA-3	LAG1-PHDI	B5
5	TCA-2	TCA-2	SUM-EXP-PHDI	B1	SPP
6	B1	B4	B5	NDVI	HERB
7		B1	MIN	MIN	PHDI
8			EXP	B3	TCA-1
9			LAG-PHDI	SR	MIN
10			TCA-2	PHDI	B3
11			SR	EXP-PHDI	EXP
12			B7	TCA-3	EXP-PHDI
13			PHDI	B2	SUM-EXP-PHDI
14			B2	TCA-2	LAG1-PHDI
15				B4	B7
16					LAG-PHDI
17					TCA-2
18					TCA-3

[a] Variable codes appear in Appendix A.

Due to mechanical malfunctions, the Landsat ETM+ sensor has become an unreliable source of remote sensing data. Data captured by the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) is particularly interesting for this application. Aboard the TERRA platform, ASTER flies the same orbit as Landsat and shares similar spectral, radiometric and temporal resolutions as ETM+, with bandwidths available. Integrating ASTER data into GSP-LAI would allow for continuous analysis into the reasonable future.

Conclusions

The development of empirical models relating ground-referenced parameters to remote sensing data may be greatly facilitated using multivariate regression techniques. The specification of ancillary variables is an effective way to include aspects of the unique ecology of a given system, in this study represented by seasonal leaf dynamics, variation in local climate, and management practices. The use of these local variables was essential for developing a model which met the objectives of multi-temporal and spatial applicability.

The evaluation of increasingly complex regression models was designed to expose simple solutions to the problem of LAI prediction if they existed. In this study none were found, and instead advanced non-linear techniques were required to incorporate important data with non-normal distributions and multicollinearity, such as serially correlated climate data.

Acknowledgements

Support for this research was provided by the United States Forest Service. We are grateful to Tim Martin and Michael Binford who provided access to valuable data, and for providing useful comments on the manuscript. This study would not have been possible without the cooperation of Alan Wilson and Brad Greenlee of Rayonier, Inc.

References

Anselin, L. 2003. GEODA 0.9 User's Guide. Urbana-Champaign, IL: University of Illinois.

Binford, M.W., H.L. Gholz, G. Starr and T.A. Martin. 2006. Regional carbon dynamics in the southeastern U.S.A. Coastal Plain: Balancing land-cover type, timber harvesting, fire, and environmental variation. Journal of Geophysical Research. 111: D24S92, doi:10.1029/2005JD006820.

Birth, G.S., and G.R. Mcvey. 1968. Measuring color of growing turf with a reflectance spectrophotometer. Agronomy Journal. 60: 640-649.

Cohen, W.B., T.K. Maiersperger, S.T. Gower, and D.P. Turner. 2003. An improved strategy for regression of biophysical variables and Landsat ETM+ data. Remote Sensing of Environment. 84: 561-571.

Corne, S.A., S.J. Carver, W.E. Kunin, J.J. Lennon, and W.W.S. van Hees. 2004. Predicting forest attributes in southeast Alaska using artificial neural networks. Forest Science. 50: 259-276.

Crist, E.P., and R.C. Cicone. 1984. A physically-based transformation of Thematic Mapper data - the TM tasseled cap. IEEE Transactions on Geoscience and Remote Sensing. 22: 256-263.

Cropper, W.P. 2000. SPM2: A simulation model for slash pine (*Pinus elliottii*) forests. Forest Ecology and Management. 126: 201-212.

Cropper, W.P., and H.L. Gholz. 1993. Simulation of the carbon dynamics of a Florida slash pine plantation. Ecological Modelling. 66: 231-249.

Curran, P.J., J.L Dungan, and H.L. Gholz. 1992. Seasonal LAI in slash pine estimated with Landsat TM. Remote Sensing of Environment. 39: 3-13.

Fang, H.L., and S.L. Liang. 2003. Retrieving leaf area index with a neural network method: Simulation and validation. IEEE Transactions on Geoscience and Remote Sensing. 41: 2052-2062.

Fang, H., and S. Liang. 2005. A hybrid inversion method for mapping leaf area index from MODIS data: Experiments and application to broadleaf and needleleaf canopies. Remote Sensing of Environment. 94: 405-424.

Fassnacht, K.S., S.T. Gower, M.D. MacKenzie, E.V. Nordheim, and T.M. Lillesand. 1997. Estimating the leaf area index of north central Wisconsin forests using the Landsat Thematic Mapper. Remote Sensing of Environment. 61: 229-245.

Flores, F.J. 2003. Using hyperspectral remote sensing to estimate leaf area index of loblolly pine plantations. Ph.D. dissertation. Raleigh, NC: North Carolina State University.

Foody, G.M., D.S. Boyd, and M.E.J. Cutler. 2003. Predictive relations of tropical forest Biomass from Landsat TM data and their transferability between regions. Remote Sensing of Environment. 85: 463-474.

Gholz, H.L., S.A. Vogel, W.P. Cropper, K. McKelvey, K.C. Ewel, R.O. Teskey, and P.J. Curran. 1991. Dynamics of canopy structure and light interception in *Pinus elliottii* stands, north Florida. Ecological Monographs. 61: 33-51.

Gobron, N., B. Pinty, and M.M. Verstraete. 1997. Theoretical limits to the estimation of the leaf area index on the basis of visible and near-infrared remote sensing data. IEEE Transactions on Geoscience and Remote Sensing. 35: 1438-1445.

Gower, S.T., C.J. Kucharik, and J.M. Norman. 1999. Direct and indirect estimation of leaf area index, f(APAR), and net primary production of terrestrial ecosystems. Remote Sensing of Environment. 70: 29-51.

Hardin, P.J., and R.R. Jensen. 2005. Neural network estimation of urban leaf area index. GIScience and Remote Sensing. 42: 251-274.

Hilbert, D.W., and B. Ostendorf. 2001. The utility of artificial neural networks for modelling the distribution of vegetation in past, present and future climates. Ecological Modelling. 146: 311-327.

Hintze, J. 2001. Number Cruncher Statistical Systems (NCSS). Kaysville, UT: NCSS.

Hodges, A.W., W.D. Mulkey, J.R. Alavalapati, D.R. Carter, and C.F. Kiker. 2005. Economic Impacts of the Forest Industry in Florida, 2003. Gainesville, FL: University of Florida, Institute of Food and Agricultural Sciences.

Hoen, H.F., and B. Solberg. 1994. Potential and economic-efficiency of carbon sequestration in forest biomass through silvicultural management. Forest Science. 40: 429-451.

Ingram, J.C., T.P. Dawson, and R.J. Whittaker. 2005. Mapping tropical forest structure in southeastern Madagascar using remote sensing and artificial neural networks. Remote Sensing of Environment. 94: 491-507.

Jensen, J.R., F. Qiu, and M.H. Ji. 1999. Predictive modeling of coniferous forest age using statistical and artificial neural network approaches applied to remote sensor data. International Journal of Remote Sensing. 20: 2805-2822.

Karl, T.R., and R.W. Knight. 1985. Atlas of the Palmer Hydrological Drought Indices (1931-1983) for the contiguous United States. Asheville, NC: National Environmental Satellite Data and Information Service.

Litynski, J.T., S.M. Klara, H.G. McIlvried, and R.D. Srivastava. 2006. An overview of terrestrial sequestration of carbon dioxide: The United States Department of Energy's Fossil Energy RandD Program. Climatic Change. 74: 81-95.

Martin, T.A., and E.J. Jokela. 2004. Developmental patterns and nutrition impact radiation use efficiency components in southern pine stands. Ecological Applications. 14: 1839-1854.

Mehrotra, K., C.K. Mohan, and S. Ranka. 2000. Elements of Artificial Neural Networks. Cambridge, MA: MIT Press.

National Climatic Data Center. 2005. Palmer Hydrological Drought Index Florida- Division 2: 1895 - 2005 Monthly Averages. Available at: http://climvis.ncdc.noaa.gov/cgibin/ginterface Accessed 12 May 2005.

Powell, T.L., G. Starr, K.L. Clark, T.A. Martin, and H.L. Gholz. 2005. Ecosystem and understory water and energy exchange for a mature, naturally regenerated pine flatwoods forest in north Florida. Canadian Journal of Forestry Research. 35: 1568-1580.

Prestemon, J.P., and R.C. Abt. 2002. Timber Products: Supply and Demand. In Southern Forest Resource Assessment, 299-325, D.N. Wear and J.G. Greis, eds. Asheville, NC: U.S. forest Service, Southern Research Station.

Prisley, S.P., and M.J. Mortimer. 2004. A synthesis of literature on evaluation of models for policy applications, with implications for forest carbon accounting. Forest Ecology and Management. 198: 89-103.

Raffy, M., K. Soudani, and J. Trautmann. 2003. On the variability of the LAI of homogeneous covers with respect to the surface size and application. International Journal of Remote Sensing. 24: 2017-2035.

Reich, P.B., D.P. Turner, and P. Bolstad. 1999. An approach to spatially distributed modeling of net primary production (NPP) at the landscape scale and its application in validation of EOS NPP products. Remote Sensing of Environment. 70: 69-81.

Running, S.W., D.L. Peterson, M.A. Spanner, and K.B. Teuber. 1986. Remote-sensing of coniferous forest leaf-area. Ecology. 67: 273-276.

Schlerf, M., and C. Atzberger. 2006. Inversion of a forest reflectance model to estimate structural canopy variables from hyperspectral remote sensing data. Remote Sensing of Environment. 100: 281-294.

Stainback, G.A., and J.R.R. Alavalapati. 2005. Effects of carbon markets on the optimal management of slash pine (*Pinus elliottii*) plantations. Southern Journal of Applied Forestry. 29: 27-32.

StatSoft, Inc. 2004. STATISTICA. Tulsa, OK: StatSoft, Inc.

Swindle, B.F., D.G. Neary, N.B. Comerford, D.L. Rockwood and G.M. Blakeslee. 1988. Fertilization and competition control accelerate early southern pine growth on flatwoods. Southern Journal of Applied Forestry. 12: 116-121.

Teskey, R.O., H.L. Gholz, and W.P. Cropper. 1994. Influence of climate and fertilization on net photosynthesis of mature slash pine. Tree Physiology. 14: 1215-1227.

Turner, D.P., M. Guzy, M.A. Lefsky, W.D. Ritts, S. Van Tuyl, and B.E. Law. 2004. Monitoring forest carbon sequestration with remote sensing and carbon cycle modeling. Environmental Management. 33: 457-466.

Appendix A - Variables used in models.

Variable	Tag	Equation / bandwidth	Notes
Band 1	B1	0.45 – 0.52 µm Blue	Surface reflectance, 8-byte
Band 2	B2	0.52 – 0.60 µm Green	″ ″
Band 3	B3	0.60 – 0.63 µm Red	″ ″
Band 4	B4	0.69 – 0.76 µm Near infrared	″ ″
Band 5	B5	1.55 – 1.75 µm Mid infrared	″ ″
Band 6	B6	10.5 – 12.5 µm thermal	Variable not used (spatial autocorrelation)
Band 7	B7	2.08 – 2.35 µm Mid infrared	Surface reflectance, 8-byte
NDVI	NDVI	(B4 – B3)/(B4 + B3)	Vegetation index
Simple Ratio	SR	B4/B3	Vegetation index
Tasseled Cap1	TCA-1	0.2043(B1) + 0.4158(B2) + 0.5524(B3) + 0.5741(B4) + 0.3124(B5) + 0.2303(B7)	*n*-space vegetation index: "Brightness"
Tasseled Cap2	TCA-2	(-0.1603(B1)) + (-0.2819(B2)) + (-0.4934(B3)) + 0.7940(B4) + 0.0002(B5) + (-0.1446(B7))	*n*-space vegetation index: "Greenness"
Tasseled Cap3	TCA-3	0.0315(B1) + 0.2021(B2) + 0.3102(B3) + 0.1594(B4) + (-0.6806(B5)) + (-0.6109(B7))	*n*-space vegetation index: "Wetness"
Species	SPP	Loblolly = 1 Slash = 0	
Fertilizer	FERT	Fertilized = 1 Not Fertilized = 0	Based on previous season
Herbicide	HERB	Treated = 1 Untreated = 0	Maintained understory control
Minimum LAI	MIN	Within Minimum = 1 Other periods = 0	Minimum leaf area; ≈ March through
April expanding LAI	EXP	Within Expansion = 1 Other periods = 0	Increasing leaf area; ≈ May through June
Maximum LAI	MAX	Within Maximum = 1 Other periods = 0	Maximum leaf area; ≈ July through Sept.
Palmer Hydrological Drought Index	PHDI	Values generated by NOAA	Monthly: severity of dry and wet spells
One year lag PHDI	LAG-PHDI	Monthly PHDI – 1 year	Previous year's PHDI
Expansion period PHDI	EXP-PHDI	Avg. PHDI for March, April, May	PHDI during leaf expansion
Previous season expansion period	LAG1-PHDI	Lagged ave. PHDI for March, April, May	PHDI during leaf expansion; interacts with phonological period
Two consecutive years expansion period PHDI	SUM-EXP-PHDI	Sum lagged ave. PHDI for March, April, May	PHDI during leaf expansion; interacts with phonological period