Available online at www.sciencedirect.com

SCIENCE DIRECT®

Remote Sensing of Environment 91 (2004) 508–520

Remote Sensing of Environment

www.elsevier.com/locate/rse

Hyperspectral versus multispectral data for estimating leaf area index in four different biomes

Kyu-Sung Lee[a], Warren B. Cohen[b,*], Robert E. Kennedy[c], Thomas K. Maiersperger[c], Stith T. Gower[d]

[a] School of Civil and Environmental Engineering, Inha University, Inchon 402-751, South Korea
[b] Forestry Sciences Laboratory, Pacific Northwest Research Station, USDA Forest Service, 3200 SW Jefferson Way, Corvallis, OR 97331, USA
[c] Department of Forest Science, Oregon State University, Corvallis, OR 97331, USA
[d] Department of Forest Ecology and Management, University of Wisconsin, Madison, WI 53706, USA

Received 2 September 2003; received in revised form 23 April 2004; accepted 29 April 2004

Abstract

Motivated by the increasing importance of hyperspectral remote sensing data, this study sought to determine whether current-generation narrow-band hyperspectral remote sensing data could better track vegetation leaf area index (LAI) than traditional broad-band multispectral data. The study takes advantage of a unique dataset, wherein field measurements of LAI were acquired at the same general time and grain size as both Landsat ETM+ and AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) imagery in four different biomes. Biome types sampled included row-crop agriculture, tallgrass prairie, mixed hardwood-conifer forest, and boreal conifer forest. The effects of bandwidth, band placement, and number of bands were isolated from radiometric quality by comparing regression models derived from individual AVIRIS channels with those derived from simulated ETM+ and MODIS channels using the AVIRIS data. Models with selected subsets of individual AVIRIS channels performed better to predict LAI than those based on the broadband datasets, although the potential to overfit models using the large number of available AVIRIS bands is a concern. Models based on actual ETM+ data were generally stronger than those based on simulated ETM+ data, suggesting that, for predicting LAI, ETM+ data suffer no penalty for having lower radiometric quality. NDVI was generally not sensitive to LAI at the four sites. Band placement of broad-band sensors (e.g., simulated ETM+ and MODIS) did not affect relationships with LAI, suggesting that there is no inherent advantage to MODIS spectral properties over those of ETM+ for estimating LAI. Spectral channels in the red-edge and shortwave-infrared regions were generally more important than those in the near-infrared for predicting LAI.
Published by Elsevier B.V.

Keywords: Hyperspectral; Multispectral; LAI

1. Introduction

Leaf area index (LAI) is an important property of vegetated systems. LAI, defined here as one-half the total surface area of leaves per unit ground area, can be used to infer processes (e.g., photosynthesis, transpiration, and evapotranspiration) and estimate net primary production (NPP) of terrestrial ecosystems (Bonan, 1993; Pierce & Running, 1988). As such, LAI is increasingly desired as a spatial data layer (i.e., map), to be used as input for modeling biogeochemical processes (Reich et al., 1999). Measuring LAI on the ground is difficult and requires a great amount of labor and cost (Gower et al., 1999). To produce an LAI map of a large area, a model relating field data with remote sensing data is typically developed, the model is inverted, and the remote sensing data are then used to extrapolate that relationship to the landscape (Cohen et al., 2003a).

Many studies have sought to establish relationships between LAI and remote sensing data (Badhwar et al., 1986; Peterson et al., 1987; Turner et al., 1999). Most of these studies have relied on empirical relationships between the ground-measured LAI and observed spectral responses (Curran et al., 1992; Peddle et al., 1999), although several have used canopy reflectance models (Jacquemond et al., 1995; Kuusk, 1998; Smith, 1993). With few exceptions,

* Corresponding author. Tel.: +1-541-750-7322; fax: +541-758-7760.
E-mail address: warren.cohen@oregonstate.edu (W.B. Cohen).

0034-4257/$ - see front matter. Published by Elsevier B.V.
doi:10.1016/j.rse.2004.04.010

such studies used broad-band multispectral data, like Landsat TM or ETM+ rather than narrow-band, hyperspectral sensors, such as the Airborne Visible Infrared Imaging Spectrometer (AVIRIS).

As a hyperspectral sensor, AVIRIS (Green et al., 1998) may have two theoretical advantages over a broad-band sensor like Landsat ETM+. First, if LAI affected spectral reflectance primarily in narrow spectral regions, then the narrow bands of AVIRIS would track these effects better than the broad bands of Landsat ETM+. Second, AVIRIS has higher radiometric quality than ETM+. AVIRIS has 12-bit data-depth, compared to the 8-bit depth of ETM+. The high signal-to-noise ratio of AVIRIS, particularly in the shortwave-infrared (SWIR) region, allows greater discrimination of subtle spectral responses. If such subtleties are in any way related to LAI, AVIRIS would have an advantage over Landsat ETM+.

It is difficult to infer from existing studies whether AVIRIS (or other hyperspectral sensors) offers an improved sensitivity to LAI over multispectral sensors. Some studies suggest no improvement of hyperspectral remote sensing over broad-band remote sensing for capturing vegetation properties. However, these studies have examined vegetation properties other than LAI (Lefsky et al., 2001), used early-generation, low signal-to-noise ratio AVIRIS data (Spanner et al., 1994), or relied exclusively on reflectance models rather than actual imagery and field data (Broge & Leblanc, 2001; Jacquemond et al., 1995). Studies that suggest an improvement of AVIRIS over broad-band sensors have not focused on LAI, inferring instead other vegetation properties such as vegetation fraction (Asner & Heidebrecht, 2002; Roberts et al., 1993), canopy chemistry (Wessman et al., 1988; Johnson et al., 1994), and plant species (Martin et al., 1998). Nearly all such studies have focused on a single biome type, which makes any conclusion difficult to generalize to other ecological systems. Also, given their dimensionality, use of hyperspectral data can lead to overfitting of statistical models (Thenkabail et al., 2000), and thus an overoptimistic view of their power.

Here, we present a study designed to resolve several of the uncertainties described above. The study capitalizes on an unusual body of remote sensing and field data collected at sites across four diverse biomes to examine the relationship between LAI and spectral reflectance. The study design allows controlled comparison of the different components of the potential LAI-spectral reflectance relationship, and thus allows inferences to be made about whether, where, and why hyperspectral imagery may improve maps of LAI.

2. Data and methods

2.1. Study sites and LAI measurements

This study was conducted in context of the BigFoot project (http://www.fsl.orst.edu/larse/bigfoot), which was designed to provide local validation of global estimates of biophysical variables, including land cover, LAI, and NPP using MODIS data (Cohen & Justice, 1999; Cohen et al., 2003b; Turner et al., 2003). Four separate BigFoot study sites were used: agricultural cropland (AGRO), tallgrass prairie (KONZ), evergreen needleleaf boreal forest (NOBS), and temperate deciduous broadleaf-dominated forest (HARV), all in North America (Table 1). These sites, and related data collection, are described in more detail by Campbell et al. (1999), Burrows et al. (2002), and Cohen et al. (2003b). At each study site, a 5×5 km area included around 100 plots where LAI was measured. Each plot was 625 m^2, with plot center locations recorded with a real-time differential global positioning system (GPS) unit having an accuracy of < 0.5 m in both the x and y directions. The sample design permitted direct assessment of the spatial scale of actual LAI from the field data. Spatial scale of LAI at each site was in excess of 100 m.

At all plots, LAI was measured at each of five subplots and these were averaged to provide a single LAI value for each plot. Methods of LAI measurement involved both direct and indirect methods (Campbell et al., 1999; Cohen et al., 2003b; Gower et al., 1999), optimized for the type of vegetation and site (Table 1). At AGRO, LAI was measured from destructively sampled corn and soybean plants. At KONZ and HARV, LAI was estimated indirectly using an Li-Cor LAI 2000 canopy analyzer with appropriate clumping coefficients for each site (Campbell et al., 1999; Gower et al., 1999). At NOBS, overstory LAI was based on site-specific allometric equations developed by Gower et al. (1997) for over 250 trees ranging in size from < 1 to 25 cm diameter at breast height. The minimal understory LAI at NOBS was estimated from percent cover of understory plants. There was no understory at the HARV site. At all sites, the within-plot variation among the five subplot measurements was minimal and the average coefficient of variation was about 6%.

To build a meaningful relationship between phenologically varying vegetation and remotely sensed data, ground measurements must be contemporaneous with the date of image acquisition. The BigFoot project provided unusually robust matches between field data measurement and image

Table 1
Study sites, LAI measurement methods, and site characteristics

Study site	Location	Major cover type(s)	LAI measurement method	LAI (mean)
Agriculture cropland (AGRO)	Champaign, IL, USA	soybean, corn	destructive sampling	2.50
Konza Prairie (KONZ)	Manhattan, KS, USA	tallgrass prairie, shrub	optical	1.93
Northern old black spruce (NOBS)	Thompson, Manitoba, Canada	evergreen needleleaf boreal forest	allometric equations	4.15
Harvard Forest (HARV)	Petersham, MA, USA	mixed hardwood and conifer forest	optical	5.08

Table 2
Data acquisition dates for AVIRIS, ETM+, and field LAI measurements

Study site	AVIRIS	ETM+	Field LAI measurements
AGRO	June 22, 2000	June 29, 2000	June 16, 2000
KONZ	June 22, 2000	June 7, 2000	June 6, 2000
NOBS	July 10, 2000	July 10, 1999	June 21, 1999
HARV	May 16, 2000		June 18, 2000
		July 28, 2001	July 27, 2000

acquisition, both for Landsat ETM+ imagery and AVIRIS imagery (Table 2). At three study sites, Landsat ETM+ data was acquired at nearly the same date (AGRO and KONZ) or day of year (NOBS) as the AVIRIS data acquisition and the LAI measurements. At HARV, however, date discrepancies were larger, particularly with the AVIRIS acquisition. There, LAI and AVIRIS data were matched for one date, and LAI and ETM+ matched for another date.

2.2. Geometric and radiometric corrections

AVIRIS data were acquired during the months of May and June of 2000. Imagery was collected by the NASA ER-2 aircraft at 20-km altitude with approximate pixel size of 17 m. The data contain 224 spectral bands, each approximately 10 nm wide, ranging from 370 to 2500 nm. Landsat ETM+ images were purchased for the dates indicated in Table 2. AVIRIS and ETM+ images were georeferenced, radiometrically calibrated, and converted to surface reflectance. Initially, the ETM+ data were georeferenced to UTM coordinates (WGS84) using either precision-corrected high-resolution imagery (e.g., IKONOS) or USGS digital orthophoto quadrangles (Cohen et al., 2003b). Geometric registration of the ETM+ data to the high resolution data resulted in a root mean square error of position of less than 7 m. The AVIRIS data were co-registered to the ETM+ data (RMSE < 15 m) using 30–40 ground controls points.

When comparing remotely sensed data across sites, atmospheric correction of the data can be important. In this study, comparisons across sites never mix data between sensors or sites in the same statistical relationship, so strict correction is less necessary. Atmospheric correction should bring digital number values from disparate sensors into rough agreement in terms of range and dynamics, however, and thus we chose to atmospherically correct all images. The carefully designed experimental AVIRIS sensor has attendant, well-developed atmospheric correction algorithms. Atmospheric water-vapor is a key factor in the atmospheric correction of remote sensing data. The atmospheric correction of hyperspectral data has a clear advantage over multispectral data since the magnitude of water-vapor effects in every pixel can be directly assessed from a few spectral channels of the data themselves. In this study, we used the ACORN program (AIG, 2002), which was based on the MODTRAN-4 radiative transfer code. ACORN used two water absorption channels (940 and 1140 nm) in AVIRIS data to estimate the amount of water vapor at the time of data acquisition.

Such post-hoc atmospheric correction is more difficult with the broad bands of Landsat data. Although several studies have dealt with atmospheric correction of Landsat data (Liang, 2001), there is no method as elegant and site-specific as that described for the AVIRIS data above. For this study, we chose to apply the ACORN program to the ETM+ data, using a standard atmospheric model and a fixed value of water vapor. While this method was unlikely to produce a perfect match with the more-detailed AVIRIS data, it did place the ETM+ data into approximately the same range of variation as the AVIRIS data.

To aid in interpreting later results, the comparative effects of atmospheric correction were investigated. Corrected AVIRIS data were convolved into ETM+-like bands (described below, under Derived Image Datasets) for comparison with corrected ETM+ imagery. Areas with minimal landscape change were isolated, and pixels compared one to one. A simple regression of AVIRIS-simulated ETM+ reflectance values on corrected actual ETM+ reflectance values was calculated, and the results summarized in Table 3. Fits between the two images were relatively good at AGRO, although the AVIRIS-simulated ETM+ data appeared to have a much narrower dynamic response in Band 4. At KONZ, strongly negative intercept values with slopes near 1 in the visible bands (bands 1–3) indicate that correction of atmospheric scattering in the ETM+ image was incomplete. The relationship in bands 4 through 7 was generally better than for the visible bands. The situation

Table 3
Regression results for atmospheric corrections (x = simulated ETM+; y = ETM+)

Site	Band	Intercept	Slope	R^2
AGRO	1	−3.21	1.11	0.62
	2	−3.90	1.08	0.67
	3	−6.75	1.32	0.70
	4	−13.97	1.81	0.59
	5	3.80	0.82	0.73
	7	1.30	0.82	0.79
KONZ	1	−5.16	1.00	0.71
	2	−5.13	0.94	0.70
	3	−10.78	1.21	0.66
	4	−2.19	1.32	0.45
	5	1.00	0.72	0.55
	7	−2.55	0.73	0.66
NOBS	1	−3.61	0.60	0.25
	2	−7.16	0.98	0.56
	3	−8.60	0.95	0.50
	4	−8.21	0.96	0.71
	5	−3.48	0.80	0.69
	7	−2.45	0.64	0.66
HARV	1	1.84	0.07	0.01
	2	3.65	0.05	0.00
	3	2.41	0.03	0.00
	4	3.31	0.63	0.24
	5	−2.35	0.81	0.20
	7	1.64	0.50	0.06

was similar at NOBS, with the degree of mismatch in the visible bands even stronger, and the discrepancies in the infrared bands slightly more pronounced. The situation at HARV was more complicated than at the other sties. In the visible bands, there was almost no relationship between images (slopes of AVIRIS on ETM+ less than 0.10), while the relationships in the infrared bands (4 through 7) were generally similar to the relationships seen at the other sites. The lack of dynamic range in the visible bands of the AVIRIS image may have been caused by its early-season acquisition date (May 16, during flushing of deciduous leaves) or by heavy atmospheric scattering in the visible bands.

Overall, the ETM+ correction appeared to be less effective than the AVIRIS correction. Again, because this study did not mix data from different sites or sensors into a single statistical relationship, this disparity was not a problem. The generally high statistical relationship between the ETM+ and AVIRIS images (as reflected by R^2 value in Table 3), however, suggests that the two images are tracking similar properties, and that statistical tests applied separately to the different sensors at a given site and between sites can yield meaningful comparisons. The one exception might be the HARV site, where special conditions may require special interpretation.

2.3. Derived image datasets

The goal of this study was to determine the relative value of broad-band, multispectral versus narrow-band hyperspectral imagery for measuring LAI. The general format of the study was to develop simple statistical relationships between field-measured LAI and some raw or derived spectral image values. By varying the character of the spectral images, it was possible to isolate the effects of bandwidth and band placement from the effects of radiometric resolution, spatial resolution, and atmospheric condition.

The base spectral images were simply the atmospherically corrected images of spectral reflectance for the original ETM+ and AVIRIS imagery. Bands 1 through 5 and band 7 were used for the ETM+ data. Of the 224 bands in the full AVIRIS dataset, 193 bands were used that did not contain strong water-absorption or instrument noise. In addition to these two base spectral reflectance images, several other datasets were derived from the AVIRIS data (Table 4).

Two simulated image sets were derived from the AVIRIS imagery. The first was a simulated ETM+ image, with the same six bands as the base ETM+ image. Simulation was achieved by weighting the appropriate narrow AVIRIS spectral bands within the ETM+ spectrum according to the spectral response function of the ETM+ instrument. This simulated image had the same radiometric quality as the AVIRIS data from which it was derived, but essentially the same band placement and bandwidth as the original, or base ETM+ data. The second derived image was one of simulated MODIS (Moderate Resolution Imaging Spectrometer) image. Only bands 1 through 7 of the MODIS instrument were simulated, again using a weighted average of AVIRIS bands. The MODIS bands differ from their ETM+ analogs in being narrower and in including an additional band at the far end of the near-IR wavelength region (1230–1250 nm).

For both of the base spectral images, a normalized difference vegetation index (NDVI) image was calculated. NDVI has seen ubiquitous use in predicting LAI. Following Teillet et al. (1997), we derived an AVIRIS NDVI dataset

Table 4
Image datasets derived from AVIRIS to be compared with LAI

Dataset	AVIRIS bands selected (center wavelength, nm)		Methods
Simulated ETM+	93 bands covering the spectrum of six reflected ETM+ bands		weighted averaging by the spectral responsivity of each ETM+ band
Simulated MODIS	29 bands covering the spectrum of the first 7 MODIS bands		weighted averaging by the spectral responsivity of each ETM+ band
NDVI	2 AVIRIS bands (655, 846)		(NIR-R)/(NIR+R)
Selected AVIRIS channels	AGRO (2)	703, 1334	stepwise multiple regression with significance level 0.15
	KONZ (18)	750, 760, 956, 1004, 1080, 1089, 1117, 1254, 1284, 1424, 1504, 1583, 1603, 1613, 1643, 1693, 1723, 1743	
	NOBS (21)	433, 452, 462, 904, 947, 1080, 1593, 1633, 1713, 1723, 1733, 1773, 1783, 1971, 1981, 2111, 2141, 2171, 2191, 2201, 2300	
	HARV (23)	413, 433, 442, 491, 655, 665, 694, 703, 751, 760, 1454, 1464, 1802, 1991, 2001, 2051, 2061, 2081, 2131, 2251, 2390, 2410, 2420	
7-channel selected AVIRIS	AGRO	655, 674, 1165, 1314, 1504,1971, 2400	stepwise multiple regression with maximum R^2 for seven independent variables
	KONZ	760, 956, 1284, 1504,1563, 1643, 1743	
	NOBS	452, 520, 1633, 1723,1773, 2201, 2300	
	HARV	578, 607, 646, 684, 2390, 2410, 2449	
PC of AVIRIS	193 bands (excluding water-absorption and noise bands from 224 bands of original AVIRIS)		the first 10 principal components of 193 AVIRIS bands

using channel 30 (655 nm) and channel 53 (846 nm). It is possible that some other combination of specific red and near-infrared bands would have provided slightly improved results (Gong et al., 2003), but identifying this specific band combination was not our objective. The NDVI image from the ETM+ data was derived using bands 4 and 3 in the standard manner (Turner et al., 1999).

The fourth derived image was a subset of the 193 AVIRIS bands. Band selection has been one of primary concerns in analyzing hyperspectral data, with selections largely dependent on the type of information desired (Price, 1997; Warner et al., 1999). In this study, we selected a subset of AVIRIS channels using stepwise multiple regression on LAI. A relaxed *p*-value tolerance of 0.15 was used, with as many as 23 AVIRIS channels being selected at each of the four sites. This is about one-fifth the number of observations (~ 100) at each site, and thus at or below the recommended maximum ratio of selected bands to data points to minimize model overfitting, as discussed by Thenkabail et al. (2000). For each site, this fourth derived image was named the "selected AVIRIS channels" image.

Even though the number of spectral bands contained in the selected AVIRIS channels image did not exceed the maximum ratio of bands to data points recommended by Thenkabail et al. (2000), we desired to further minimize the possible effect of overfitting. Moreover, we were interested in removing the effect of using more hyperspectral bands than multispectral bands when comparing models derived from each. This would enable a more direct assessment of bandwidth and placement. Thus, we derived a fifth image dataset of only the first seven selected channels from the multiple stepwise regression. This corresponds to the maximum number of bands in other data sets (simulated MODIS, and simulated and actual ETM+). We called this the "7-channel selected AVIRIS" image.

The sixth derived dataset was calculated from a standard principal component (PC) analysis of the 193 original AVIRIS bands. For all study sites, the first 10 principal components explained about 99% of total variability of original AVIRIS dataset. This dataset captured the spectral dimensionality of the AVIRIS instrument, but may have obscured subtle and specific narrow-band relationships with LAI. The first 10 principal components of 193 AVIRIS bands were named the "PC of AVIRIS" image.

For all of these spectral images, spectral reflectance values were derived for the field plots in the same manner: each plot was identified in each spectral image using a vector map of plot locations, and the four pixels spanning each plot boundary were spatially averaged to obtain a mean spectral vector for each plot. Although the four pixels area of AVIRIS and ETM+ is larger than the actual plot size (625 m^2), the reflectance values of the four pixels exhibited minimal variance due to high spatial autocorrelation. The use of four pixel averages also minimized possible misrepresentations that might be caused from location error related to geometric correction.

2.4. Correlation analyses

The simplest statistical investigation was calculation of a band-by-band correlation between spectral image value and field-measured LAI. This was conducted on the two base spectral images (ETM+ and AVIRIS) as well as on the simulated ETM+ and the simulated MODIS images. Because each spectral image represented a different combination of spectral strengths and weaknesses, discrepancies in correlation at particular wavelengths can be used to begin inferring which spectral qualities are important in measuring LAI.

The next statistical investigation was a regression of AVIRIS- or ETM+-derived NDVI on the field-measured LAI values. Although similar to many studies in the literature, this investigation is unique in having such close temporal agreement between AVIRIS and ETM+ data and in representing four distinct biomes. We also regressed the multiple-band images on field-measured LAI. The multiple-band images involved were the ETM+ image, the simulated ETM+ image, the simulated MODIS image, the PC of AVIRIS image, the selected AVIRIS channels image, and the 7-channel selected AVIRIS image. We used canonical correlation analysis (CCA), a multivariate statistical method used in a similar context by Cohen et al. (2003a), to derive an integrated spectral index to plot against measured LAI values. CCA is a generalized form of multiple regression designed to align two sets of variables (multiple *X*'s and multiple *Y*'s) by maximizing the correlation between a composite of variables from one set with a composite of variables from the other set (Tabachnick & Fidell, 1989). When one side of the equation has only one variable (such as LAI in this study), CCA provides a set of coefficients to transform the other set of variables (in this case, spectral vectors) into a single transformed variable called the canonical variable (or axis) that is maximally correlated with the single variable on the other side of the equation. In such cases, the canonical correlation coefficient (r) is equivalent to the square root of the coefficient of determination (R^2) of multiple regression.

3. Results and discussion

3.1. Correlations between LAI and spectral reflectance

Correlation coefficients between LAI and spectral reflectance were highly variable by both wavelength and site (Fig. 1). In general, the non-forested and forested sites behaved differently. For the two non-forested sites (AGRO and KONZ), the highest positive correlations were in the near-IR regions for all image sources (ETM+, AVIRIS, simulated ETM+, and simulated MODIS). For all other spectral regions at these sites, correlation coefficients were negative, or near zero. AGRO showed high correlations across the spectral range, and strong agreement between

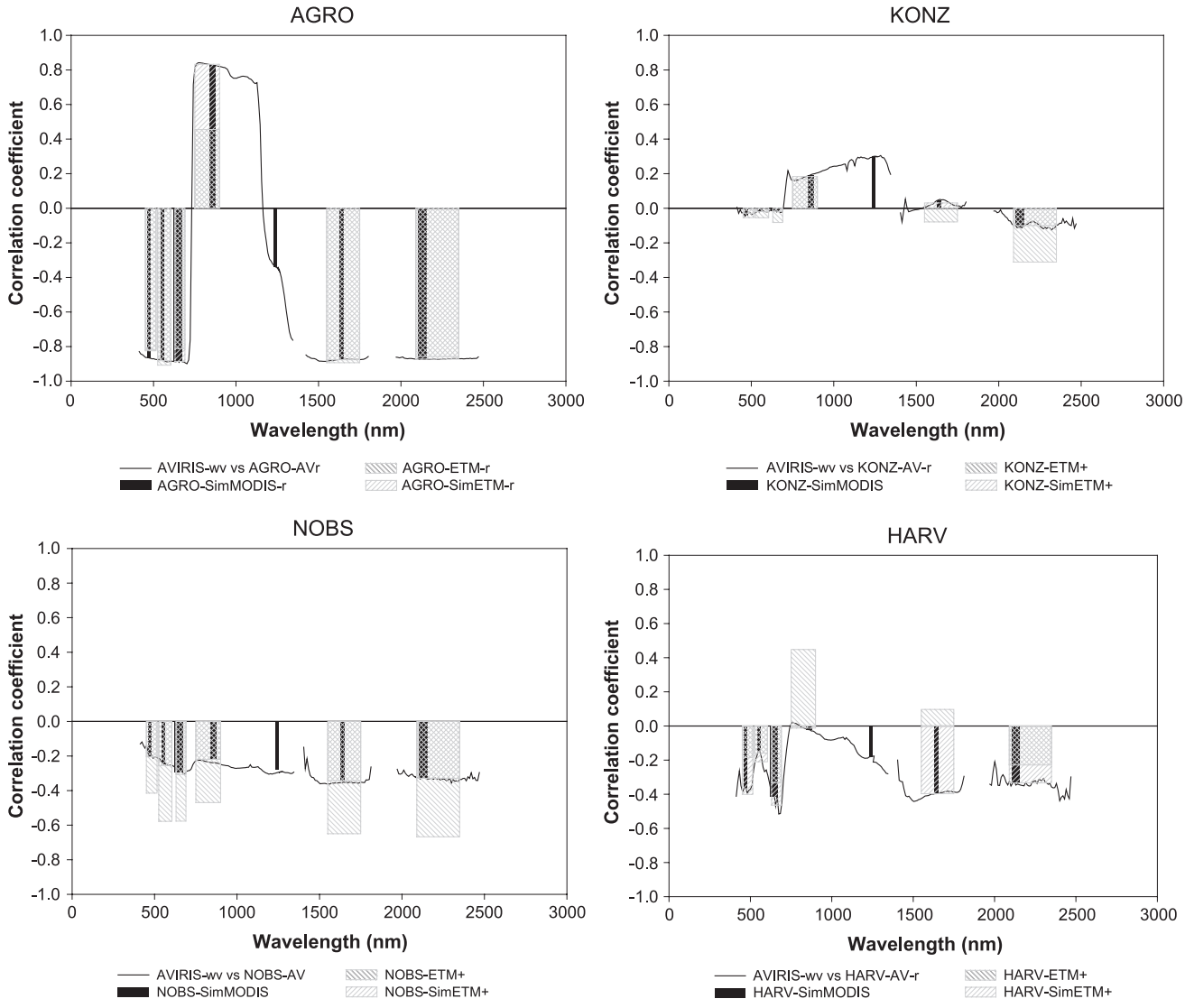

Fig. 1. Correlation coefficients for the relationship between field-measured LAI and spectral reflectance at the four study sites along the AVIRIS channels and the other three multispectral bands.

ETM+ and derived AVIRIS datasets. Agreement was weaker at KONZ, particularly in the visible region and the SWIR regions (around 1600 and 2200 nm). Notably, LAI at KONZ was more strongly correlated in several bands to the actual ETM+ data than to the simulated ETM+ data, the simulated MODIS data, or the original AVIRIS data.

At the two forested sites, correlation between spectral data and LAI was negative across nearly the entire spectral range. At NOBS, the SWIR regions had stronger negative correlations than either the near-IR region or the red region. LAI at NOBS was more strongly correlated to the actual ETM+ data than to any of the AVIRIS or AVIRIS-derived image datasets. The situation at HARV was less clear than at the other sites. Agreement between ETM+ data and AVIRIS and AVIRIS simulated data was generally strong in the visible region, but quite weak (even oppositional) across the infrared regions. If disagreement between sensors were related to discrepancies in atmospheric correction (summarized in Table 3), agreement patterns would be reversed: agreement low in the visible bands, and high in the infrared. It appears that date of image acquisition plays the dominant role in the differences between the sensors, with the ETM+ image captured at the height of biomass for the system (Table 2) and the AVIRIS image captured before full flushing of the deciduous canopy.

Several general inferences can be made from these correlations. The agreement between the original AVIRIS bands and the simulated ETM+ and MODIS bands suggests that narrow bands have no discernible advantage over broad bands, at least when viewed on a single band-by-band basis. Moreover, the comparison between simulated ETM+ and real ETM+ data suggests that the real ETM+ data suffers no penalty for its lower radiometric quality. Regardless of

sensor, the relationship between LAI and spectral values seems to exhibit few patterns of correlation that are universal across the sites: different regions of the spectrum are important in different biomes. It is interesting to note that the contrast between the near-IR and the red regions—the foundation of most of the common vegetation indices in use today—is the dominant feature primarily in the agricultural system, where the common vegetation indices were originally exploited with great success. These results suggest that complex natural systems may be better served by a vegetation index that includes regions of the spectrum other than the near-IR and red wavelengths. As seen in Fig. 1, the magnitude of correlation coefficients in SWIR region is higher than other wavelength, particularly at the forested sites.

3.2. Correlations between NDVI and LAI

NDVI (and its counterpart the simple ratio) has been the most widely used spectral vegetation index to estimate LAI (Cohen et al., 2003a). Thus, it is important in a study such as this to contrast other results against the NDVI. The four sites in this study represent a diverse group of vegetation types, and having results across these sites should enable some general observations to be made about the strength of NDVI–LAI relationships.

Results suggest that the NDVI was generally not sensitive to values of LAI measured at the four study sites (Fig. 2). Only at AGRO was there any meaningful trend, and that trend was evident primarily for the relatively low LAI soybeans only (Fig. 2). There was a slight trend in the relationship between NDVI and LAI at HARV, but only at the very lowest LAI values. This general low correlation between NDVI and LAI has been noted in numerous studies (Chen & Cihlar, 1996; Cohen et al., 2003b; Turner et al., 1999). NDVI has been (for nearly three decades) a popular index with which to estimate LAI across diverse systems, but these results suggest that other indices may be more appropriate. Fortunately, numerous recent studies have noted a strong contribution of SWIR bands to the strength of relationships between reflectance and LAI (Brown et al., 2000; Cohen & Goward, 2004; Nemani et al., 1993).

At three of the sites (AGRO, KONZ and NOBS), AVIRIS NDVI values were considerably higher than ETM + NDVI values. This can be caused by insufficient removal of atmospheric scattering effects in the visible bands of the ETM+ data. Scattering in the visible bands inflates the red component of the NDVI relationship, which dampens the contrast between near-IR and red, and diminishes NDVI values (Turner et al., 1999). This is consistent with our observations of post-atmospheric correction imagery, summarized in Table 3: atmospheric correction of ETM+ in the visible bands was generally incomplete, leading to inflated ETM+ reflectance in those bands.

3.3. Canonical correlation analysis

CCA was used here as a convenient way of integrating several selected wavebands into a single index from each dataset. In Fig. 3, canonical axis scores from different image datasets are shown regressed against LAI values by site. Were these relationships to be inverted and used in mapping, it would desirable to withhold some proportion of plots to assess map error. However, the purpose of this study is to illustrate the relative information content of the different image datasets, and thus it is appropriate to use these best-case conditions to explore the tradeoffs in different image source datasets in tracking LAI.

Inferences about the relative importance of band placement, bandwidth, sensor radiometric quality, and model overfitting can be made by comparing different pairs of graphs in Fig. 3. The top two graphs at each site use the full complement of AVIRIS band placement and radiometric quality, but the top graph (selected AVIRIS bands, $n=2$–23) capitalizes on individual AVIRIS narrow bands to develop a site-specific relationship with LAI, while the second graph (10 principal components of AVIRIS) uses compressed AVIRIS data which obscures the relative strengths of individual bands. A better fit by the former over the latter might suggest that individual narrow bands have an advantage over compressed data. However, given that the PC of AVIRIS image contained over 99% of the variance of the full 193 channel dataset, it is difficult to assess the relative effects of overfitting between models developed on that dataset versus the models developed with the selected AVIRIS channels dataset. Comparing the graphs derived from the selected AVIRIS channels dataset with the 7-channel selected AVIRIS dataset (third set of graphs) reveals the effect of adding more bands to, and potentially overfitting of, models based on AVIRIS data. Comparing the fourth and fifth graphs (ETM+ data versus simulated ETM+ data, respectively) at each site sheds light on whether improved radiometric quality of AVIRIS data can provide an advantage over actual ETM+ data. By comparing simulated ETM+ and simulated MODIS imagery, the fifth and sixth graphs address the issue of broadband width and placement. Finally, comparisons among all graphs suggest which combination of these effects has the greatest potential for capturing patterns in LAI, within the context of the potential effects of model overfitting.

At all sites, comparison of the top two graphs suggests that the use of individual narrow bands allowed for a better fit of LAI than PC-based compressed bands. At AGRO, the two are comparable, but at all other sites the fit from the selected AVIRIS channel-derived CCA axis is better than that derived from the PC of AVIRIS-derived CCA. Comparing these results with the third row of graphs reveals the potential effects of model overfitting. With the exception of AGRO, the models based on seven AVIRIS channels are less powerful than those based on the selected AVIRIS channels image. These are still

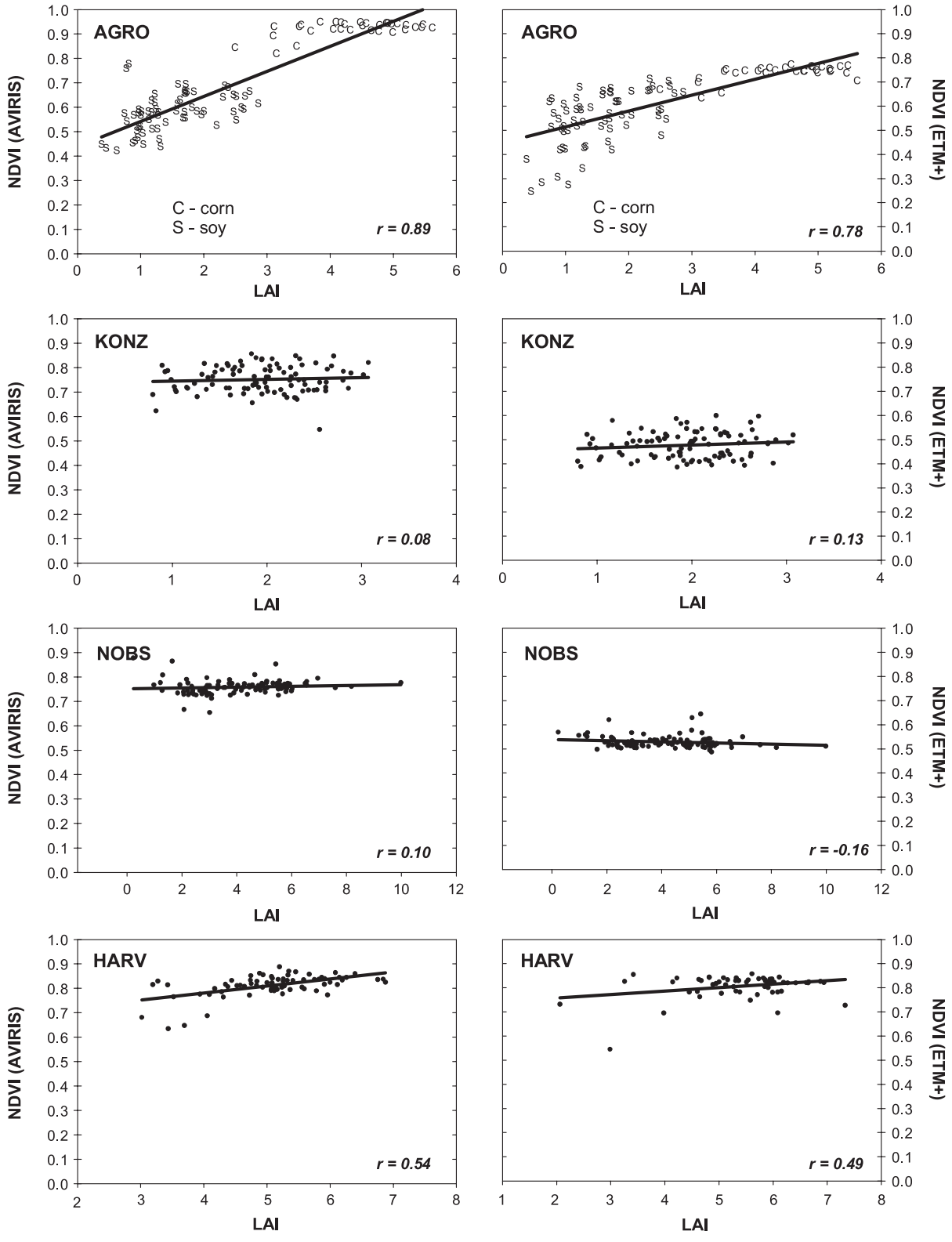

Fig. 2. Relationship between field-measured LAI and NDVI for ETM+ and AVIRIS at the four study sites.

stronger than the PC of AVIRIS models, suggesting that even if the PC of AVIRIS images (which contained 99% of the variation of the full 193 band AVIRIS datasets) are overfit, their compressed nature obscures their value relative to the even smaller subset of seven individual AVIRIS bands datasets.

There appeared to be no improvement in prediction of LAI from the higher radiometric quality of the AVIRIS data.

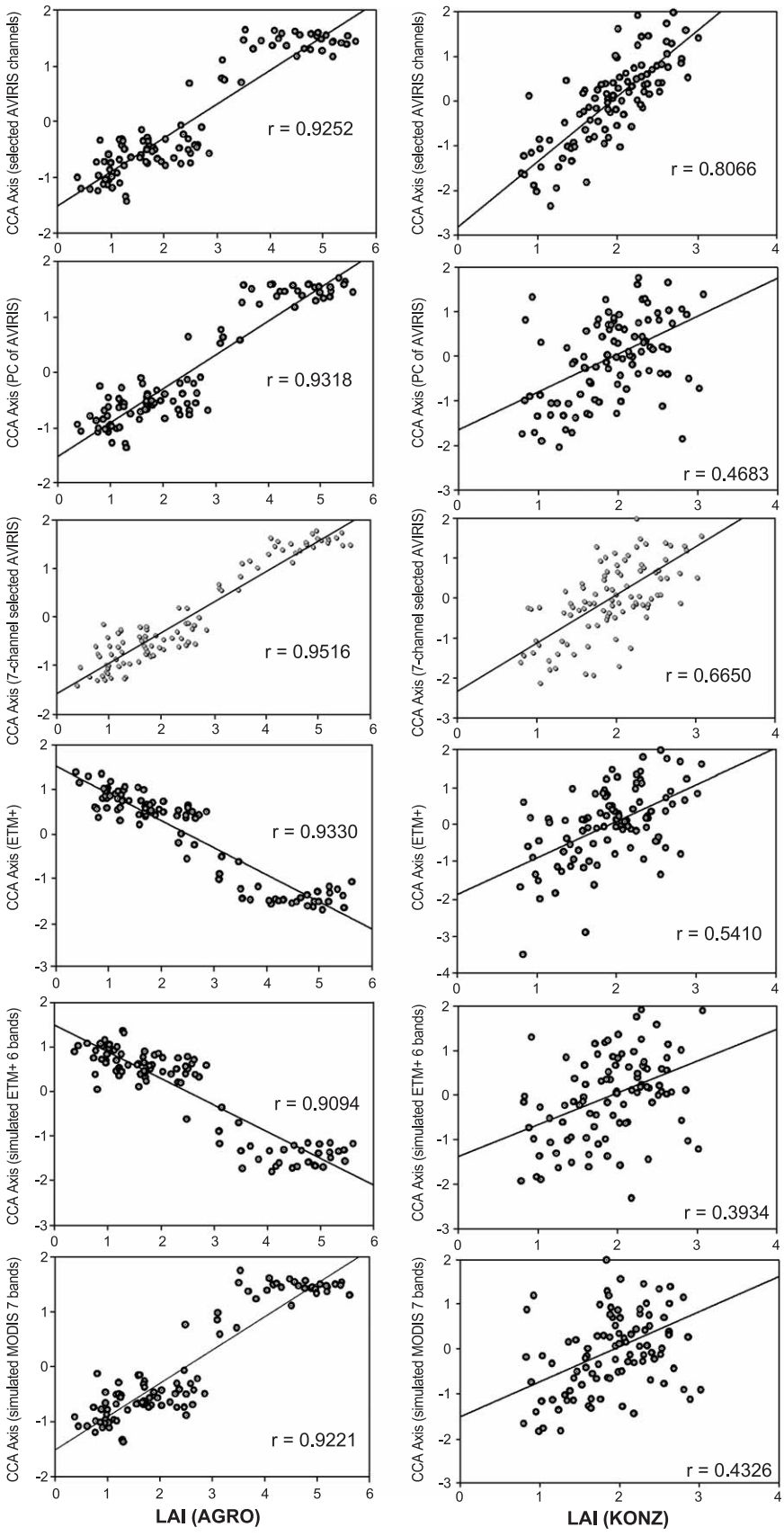

Fig. 3. The relationship between LAI and CCA-transformed spectral reflectance for six derived data sets at each study site.

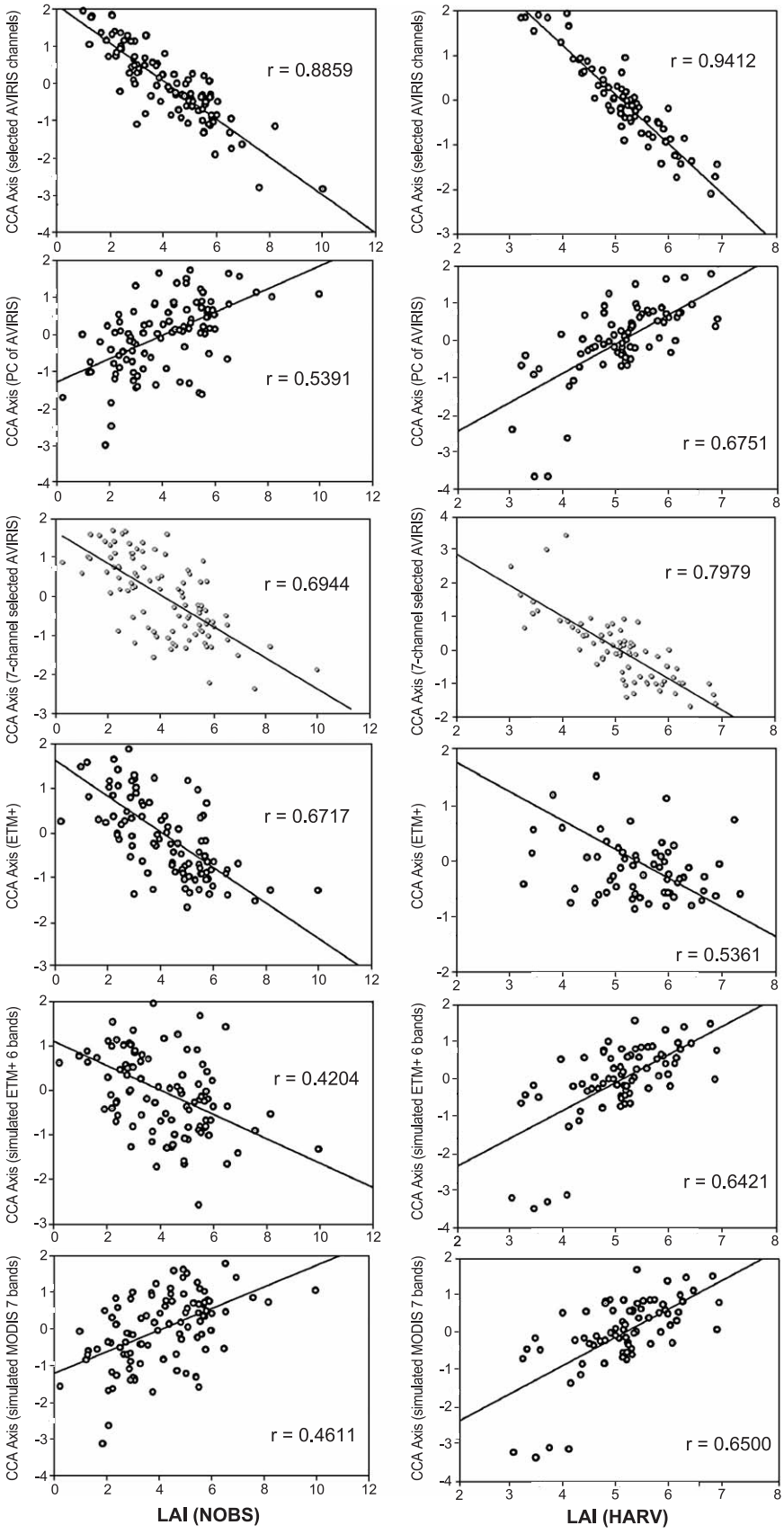

Fig. 3 (*continued*).

At AGRO, the actual ETM+ data and simulated ETM+ data were comparable. At NOBS and KONZ, the simulated data were moderately weaker than the ETM+ data. It is not possible to isolate indirect influences of this difference within our study, although two likely possibilities are differences in grain size or slight variations in image acquisition date. At HARV, the simulated ETM+ data better captured LAI patterns, which may be related to better atmospheric correction of the AVIRIS data, or simply due to the greater discrepancies in date of acquisition of image and field data at HARV. Bandwidth and band placement of simulated ETM+ and MODIS data appear to make little difference across sites.

Comparing across all image datasets, it appears that the use of individual, narrow bands of AVIRIS data yielded the best relationships with LAI. One exception was at AGRO, where all image datasets performed similarly. This result was driven primarily by the discrete contrast between the two agricultural vegetative types (corn and soybeans). That more bands of AVIRIS data added to the apparent strength of predictive LAI models is not surprising, given the relatively lower canonical correlation coefficients with the 7-channel selected AVIRIS dataset; however, the possibility that the models based on the selected AVIRIS channels images are overfit remains a major concern. Models based on the 7-channel selected AVIRIS dataset yielded higher correlations than those based on the broad-band datasets (i.e., 6-band actual and simulated ETM+, and 7-channel simulated MODIS datasets). This result does suggest that bandwidth and placement might be important factors in relative model strength.

3.4. Distribution of selected AVIRIS channels for the estimation of LAI

Two interesting observations arise from examining the bands chosen for the "selected AVIRIS" and the "7-channel selected AVIRIS" image datasets (Fig. 4). First, the area of vegetation's "red-edge" (around between about 710 and 750 nm) was identified as important at three of the four sites (AGRO, KONZ, and HARV). The correlation with LAI around this region has been reported in previous studies (Asner, 1998; Danson, 1996). Second, although the near-IR region (up to ~ 1100 nm) has been the cornerstone of the ubiquitous vegetation indices (NDVI, simple-ratio), it emerged as an important region of the spectrum for both datasets only at the KONZ site. However, for all datasets, bands from the SWIR region (beyond ~ 1300 nm) were important. A number of studies have identified this region of the reflectance spectrum as potentially important for tracking vegetative properties. Considering that the SWIR bands were important at all four biomes in this study, vegetation indices that do not include this spectral region may be less than satisfactory for estimating LAI. Other recent studies have demonstrated the same general result for a variety of vegetation attributes (Asner, 1998; Cohen et al., 2003b; Eklundh et al., 2001).

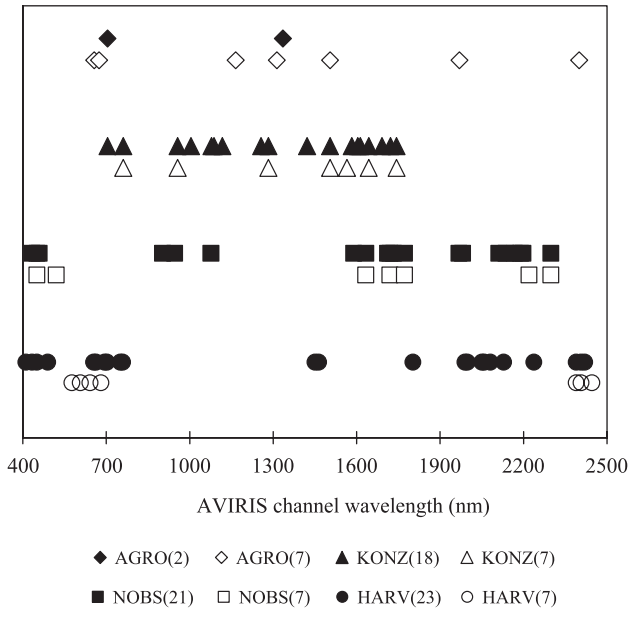

Fig. 4. Distribution of two datasets of AVIRIS spectral channels selected from the stepwise multiple regression to predict LAI at four study sites.

4. Conclusions

It has been more than 15 years since hyperspectral remote sensing began its emergence as an important remote sensing device (Goetz et al., 1985). With many airborne imaging spectrometer systems in use today and the rise of spaceborne hyperspectral sensors, better understanding of this type of image data is increasingly needed. This study was designed to explore whether and why narrow-band, hyperspectral data may improve remote sensing of LAI over broad-band, multispectral data. Its strength was in utilizing a unique field and remote-sensing dataset developed for the BigFoot project, allowing comparison of sensor types using consistent techniques across four diverse vegetative systems.

The major results of this work can be concluded as follows:

- *Number of bands appears to be the most important advantage of using hyperspectral data over multispectral data to predict LAI within four diverse biome types*. By comparing regression models derived from individual AVIRIS channels with those derived from simulated ETM+ and MODIS channels using the AVIRIS data, the effects of bandwidth, band placement, and number of bands were largely isolated from radiometric quality. Models with selected AVIRIS channels performed better to predict LAI than the other datasets, with the best models having up to 23 AVIRIS channels. However, these

latter models may have been overfit. When the number of channels was limited to seven, model strengths were intermediate between those derived from seven broad bands, and those having more, narrow channels. This suggests some advantage of using narrow, select channels over broad bands. Nonetheless, models based on actual ETM+ data were generally stronger than those based on simulated ETM+ data, suggesting that ETM+ data suffer no penalty for having lower radiometric quality than the AVIRIS-simulated ETM+ data for predicting LAI.

- *If number of bands is the primary advantage of hyperspectral data over multispectral data for predicting LAI, a considerable amount of work remains to determine when a predictive model is overfit.* Studies like those of Thenkabail et al. (2000) that use neutral models to compare random and actual datasets are useful, but they are only one approach. Testing of models with independent datasets will also help determine if a given model is overfit.
- *NDVI was generally not sensitive to LAI at the four sites*. Contrasting agricultural types produced apparent sensitivity at one site, but within types, only the very low LAI type had a useful relationship. This suggests that the ubiquitous near-IR-based indices are not capitalizing on important spectral information of available datasets.
- *Spectral channels in the red-edge and SWIR regions are generally more important than those in the near-IR for predicting LAI*. Because of its narrow wavelength characteristic, the red edge is only detectable with hyperspectral sensors, whereas the SWIR region is observable with many multispectral sensors.
- *Band placement of broad-band sensors (e.g., simulated ETM+ and MODIS) did not affect relationships with LAI*. This suggests that there is no inherent advantage to MODIS spectral properties over those of ETM+ for estimating LAI. However, other characteristics of the sensors, such as spatial and temporal resolution, may be very important.

Acknowledgements

This research was funded by NASA's Terrestrial Ecology Program. We greatly thank all of those that helped collect the field data, including Al Kirschbaum, John Lennon Campbell, Drew Feldkirschner, and Sean Burrows. We owe much appreciation to Jeff Morisette and Jaime Nickeson for their help in so effectively coordinating MODIS validation activities.

References

AIG [Analytical Imaging and Geophysics] (2000). *ACORN 4.0 User's Guide* (Boulder, CO, USA).

Asner, G. P. (1998). Biophysical and biochemical sources of variability in canopy reflectance. *Remote Sensing of Environment*, *64*, 234–253.

Asner, G. P., & Heidebrecht, K. B. (2002). Spectral unmixing of vegetation, soil and dry carbon cover in arid regions: Comparing multispectral and hyperspectral observations. *International Journal of Remote Sensing*, *23*, 3939–3958.

Badhwar, G. D., MacDonald, R. B., & Mehta, N. C. (1986). Satellite-derived leaf area index and vegetation maps as input to global carbon cycle models: A hierarchical approach. *International Journal of Remote Sensing*, *7*, 265–281.

Bonan, G. (1993). Importance of leaf area index and forest type when estimating photosynthesis in boreal forests. *Remote Sensing of Environment*, *43*, 303–314.

Broge, N. H., & Leblanc, E. (2001). Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sensing of Environment*, *76*, 156–172.

Brown, L., Chen, J., Leblanc, S., & Cihlar, J. (2000). A shortwave infrared modification to the simple ratio for LAI retrieval in boreal forests: An image and model analysis. *Remote Sensing of Environment*, *71*, 16–25.

Burrows, S., Gower, S., Clayton, M., Mackay, D., Ahl, D., Norman, J., & Diak, G. (2002). Application of geostatistics to characterize leaf area index (LAI) from flux tower to landscape scales using a cyclic sampling design. *Ecosystems*, *5*, 667–679.

Campbell, J. C., Burrows, S., Gower, S. T., & Cohen, W. B. (1999). BigFoot: Characterizing land cover, LAI, and NPP at the Landscape Scale for EOS/MODIS Validation. Field Manual 2.1. Environmental Science Division Pub. No. 4937. Oak Ridge National Laboratory, Oak Ridge, TN.

Chen, J. M., & Cihlar, J. (1996). Retrieving leaf area index of boreal conifer forests using Landsat TM images. *Remote Sensing of Environment*, *55*, 153–162.

Cohen, W., & Justice, C. (1999). Validating MODIS terrestrial ecology products: Linking in situ and satellite measurements. *Remote Sensing of Environment*, *70*, 1–3.

Cohen, W. B., & Goward, S. N. (2004). Landsat's role in ecological applications of remote sensing. *BioScience*, *54*, 535–545.

Cohen, W. B., Maierpserger, T. K., Gower, S. T., & Turner, D. P. (2003a). An improved strategy for regression of biophysical variables and Landsat ETM+ data. *Remote Sensing of Environment*, *84*, 561–571.

Cohen, W. B., Maiersperger, T. K., Yang, Z., Gower, S. T., Turner, D. P., Ritts, W. D., Berterretche, M., & Running, S. W. (2003b). Comparisons of land cover and LAI estimates derived from ETM+ and MODIS for four sites in north America: A quality assessment of 2000/2001 provisional MODIS products. *Remote Sensing of Environment*, *88*, 233–255.

Curran, P. J., Dungan, J., & Gholz, H. L. (1992). Seasonal LAI measurements in slash pine using Landsat TM. *Remote Sensing of Environment*, *39*, 3–13.

Danson, F. M. (1996). Developments in the remote sensing of forest canopy structure. In Danson, F. M. & Plummer, S. E. (Eds.), *Advances in environmental remote sensing* (pp. 53–69). New York, USA: Wiley.

Eklundh, L., Harrie, L., & Kuusk, A. (2001). Investigating relationships between Landsat ETM+ sensor data and leaf area index in a boreal conifer forest. *Remote Sensing of Environment*, *78*, 239–251.

Goetz, A. F. H., Vane, G., Solomon, J. E., & Rock, B. N. (1985). Imaging spectrometry for Earth remote sensing. *Science*, *228*, 1147–1153.

Gong, P., Pu, R., Biging, G. S., & Larrieu, M. R. (2003). Estimation of forest leaf area index using vegetation indices derived from hyperion hyperspectral Data. *IEEE Transactions on Geoscience and Remote Sensing*, *41*, 1355–1362.

Gower, S., Kucharik, C., & Norman, J. (1999). Direct and indirect estimation of leaf area index, fAPAR, and net primary production of terrestrial ecosystems. *Remote Sensing of Environment*, *70*, 29–51.

Gower, S. T., Vogel, J., Norman, J. M., Kucharik, C. J., Steele, S. J., & Stow, T. K. (1997). Carbon distribution and net primary production of aspen, jack pine and black spruce BOREAS forests. *Journal of Geophysical Research*, *102*(D24), 29029–29041.

Green, R. O., Eastwood, M. L., & Williams, O. (1998). Imaging spectroscopy and the Airborne Visible/infrared Imaging Spectrometer (AVIRIS). *Remote Sensing of Environment*, *65*, 227–240.

Jacquemond, S., Baret, F., Andrieu, B., Danson, F. M., & Jaggard, K. (1995). Extraction of vegetation biophysical parameters by inversion of the PROSPECT+SAIL models on sugar beet canopy reflectance data. Application to the TM and AVIRIS sensors. *Remote Sensing of Environment*, *52*, 163–172.

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

Kuusk, A. (1998). Monitoring of vegetation parameters on large areas by the inversion of a canopy reflectance model. *International Journal of Remote Sensing*, *19*, 2893–2905.

Lefsky, M., Cohen, W., & Spies, T. (2001). An evaluation of alternative remote sensing products for forest inventory, monitoring, and mapping of Douglas-fir forests in western Oregon. *Canadian Journal of Forest Research*, *31*, 78–87.

Liang, S. (2001). Atmospheric correction of landsat ETM+ landsurface imagery-part I: Methods. *IEEE Transactions on Geoscience and Remote Sensing*, *39*, 2490–2498.

Martin, M., Newman, S., Aber, J., & Congalton, R. (1998). Determining forest species composition using high resolution remote sensing data. *Remote Sensing of Environment*, *65*, 249–254.

Nemani, R. R., Pierce, L., Running, S., & Band, L. (1993). Forest ecosystem processes at the watershed scale: Sensitivity to remotely-sensed leaf area index estimates. *International Journal of Remote Sensing*, *14*, 2519–2534.

Peddle, D. R., Hall, F. R., & LeDrew, E. F. (1999). Spectral mixture analysis and geometric-optical reflectance modeling of boreal forest biophysical structure. *Remote Sensing of Environment*, *67*, 288–297.

Peterson, D. L., Spanner, M. A., Running, S. W., & Teuber, K. (1987). Relationship of thematic mapper data to leaf area index of temperate coniferous forests. *Remote Sensing of Environment*, *22*, 323–341.

Pierce, L. L., & Running, S. W. (1988). Rapid estimation of coniferous forest leaf area index using a portable integrating radiometer. *Ecology*, *69*, 1762–1767.

Price, J. C. (1997). Spectral band selection for visible-near infrared remote sensing: Spectral-spatial resolution tradeoffs. *IEEE Transactions on Geoscience and Remote Sensing*, *35*, 1277–1285.

Reich, P. B., Turner, D. P., & Bolstad, P. (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.

Roberts, D. A., Smith, M. O., & Adams, J. B. (1993). Green vegetation, nonphotosynthetic vegetation, and soils in AVIRIS. *Remote Sensing of Environment*, *44*, 255–269.

Smith, J. A. (1993). LAI Inversion using a back-propagation neural network trained with a multiple scattering model. *IEEE Transactions on Geoscience and Remote Sensing*, *31*(5), 1102–1106.

Spanner, M. A., Johnson, L., Miller, J., McCreight, R., Freemantle, J., Runyon, J., & Gong, P. (1994). Remote sensing of seasonal leaf area index across the Oregon transect. *Ecological Applications*, *4*, 258–271.

Tabachnick, B., & Fidell, L. (1989). *Using multivariate statistics* (2nd edition). United Kingdom: Harper Collins.

Teillet, P. M., Staenz, K., & Williams, D. J. (1997). Effects of spectral, spatial, and radiometric characteristics on remote sensing vegetation indices of forested regions. *Remote Sensing of Environment*, *61*, 139–149.

Thenkabail, P. S., Smith, R. B., & De Pauw, E. (2000). Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sensing of Environment*, *71*, 158–182.

Turner, D., Cohen, W., Kennedy, R., Fassnacht, K., & Briggs, J. (1999). Relationships between leaf area index and Landsat TM spectral vegetation indices across three temperate zone sites. *Remote Sensing of Environment*, *70*, 2–68.

Turner, D. P., Ritts, W., Cohen, W. B., Gower, S. T., Zhao, M., Running, S. W., Urbanski, S., Dunn, A., & Munger, J. W. (2003). Scaling Gross Primary Production (GPP) over boreal and deciduous forest landscapes in support of MODIS GPP product validation. *Remote Sensing of Environment*, *88*, 256–270.

Warner, T. A., Steinmaus, K., & Foote, H. (1999). An evaluation of spatial autocorrelation feature selection. *International Journal of Remote Sensing*, *20*, 1601–1616.

Wessman, C. A., Aber, J. D., Peterson, D. L., & Melilo, J. M. (1988). Remote sensing of canopy chemistry and nitrogen cycling in temperature forest ecosystems. *Nature*, *335*, 154–156.