

# Biomass Modeling of Four Leading World Crops Using Hyperspectral Narrowbands in Support of HyspIRI Mission

Michael Marshall and Prasad Thenkabail

## Abstract

New satellite missions are expected to record high spectral resolution information globally and consistently for the first time, so it is important to identify modeling techniques that take advantage of these new data. In this paper, we estimate biomass for four major crops using ground-based hyperspectral narrowbands. The spectra and their derivatives are evaluated using three modeling techniques: two-band hyperspectral vegetation indices (HVI), multiple band-HVI (MB-HVI) developed from Sequential Search Methods (SSM), and MB-HVI developed from Principal Component Regression. Overall, the two-band HVI and MB-HVI developed from SSMs using first derivative transformed spectra in the visible blue and green and NIR explained more biomass variability and had lower error than the other approaches or transformations; however a better search criterion needs to be developed in order to reflect the true ability of the two-band HVI approach. Short-Wave Infrared 1 (1000 to 1700 nm) proved less effective, but still important in the final models.

## Introduction

California leads the United States (US) in agriculture, receiving more than 40 billion dollars in revenue or approximately 12 percent of US agriculture revenues (CDFA, 2013). Seven of the top ten agricultural producing counties (Fresno, Tulare, Kern, Merced, Stanislaus, San Joaquin, and King) are in the Central Valley of California. More than 60 percent of crops by area are irrigated, which accounts for 75 to 80 percent of the state's annual water budget (USDA, 2009). Growing and competing urban and domestic use, declines in non-local sources (e.g., Colorado River), and environmental legislation in the Sacramento-San Joaquin Delta have put considerable strain on agricultural water resources (Faunt, 2009). Alfalfa, the top water user by crop type, accounts for nearly a quarter of the state's irrigation water, while yielding only 4 percent in crop revenues. The other large water users include maize, rice, and cotton, and generate an additional 10 percent total in crop revenues. Other factors notwithstanding, the disparity between crop water use and yield affords an opportunity for tremendous water savings that incurs a small cost to the state's

economy (Umbach, 1997). Increasing crop water productivity (WP), defined here as the ratio of actual marketable crop yield to actual seasonal crop water consumption or evapotranspiration (Zwart and Bastiaanssen, 2004), involves integrated water management strategies that focus on irrigation techniques and scheduling, soil amendments, seed preparation, tilling practices, water harvesting, crop species, and variety selection (Ali and Talukder, 2008), as well as economic incentives, such as water markets and crop subsidies (Sunding, 2000). Remote sensing-based models of WP facilitate the design, implementation, and assessment of these strategies at low cost and over large areas (Bastiaanssen *et al.*, 2000).

Pinter *et al.* (2003) reviews remote sensing approaches in agriculture, focusing on broad-band, hyperspectral narrow bands (HNBS), and thermal band based vegetation indices and their application to crop yield estimation and the management of water, nutrients, and pests. Actual marketable crop yield can be estimated empirically by remote sensing-based indices that are sensitive to several crop biophysical and biochemical properties. Biomass production is the net gain in carbon and energy through assimilation, which is a measure of physiological efficiency and is therefore a common determinant of empirically-based yield models. The harvest index, for example, is a crop-specific coefficient developed from field studies that reduces crop biomass to yield (Hay, 1995). Given the sensitivity of the index to crop variety, management practices, and external environmental factors, physiologically-based crop growth models where biomass is an integral model component have been developed (Doraiswamy *et al.*, 2003).

Field measurements of crop biomass are time-consuming, destructive, and difficult to extrapolate over large areas, so remote sensing biomass techniques have been developed to overcome these limitations. Early attempts used the Normalized Difference Vegetation Index (NDVI) to estimate biomass, as the index is sensitive to plant "greenness," however, results were less accurate and consistent compared to NDVI estimates of primary productivity (Box *et al.*, 1989). Lu (2006) provides a review of empirical and semi-process based biomass modeling techniques using broad-band multi-spectral remote sensors, radar, and lidar, with an emphasis on forest land cover. Empirical models of biomass are typically measured

---

Michael Marshall is with Climate Change, Agriculture, and Food Security, World Agroforestry Centre, P.O. Box 30677-00100, Nairobi, Kenya, and the Southwestern Geographic Center, United States Geological Survey (USGS), 2255 Gemini Dr, Flagstaff, AZ (m\_marshall@cgiar.org).

Prasad Thenkabail is with the Southwestern Geographic Center, United States Geological Survey (USGS), 2255 Gemini Dr, Flagstaff, AZ.

---

Photogrammetric Engineering & Remote Sensing  
Vol. 80, No. 8, August 2014, pp. 757–772.  
0099-1112/14/8008-757

© 2014 American Society for Photogrammetry  
and Remote Sensing  
doi: 10.14358/PERS.80.8.757

over specific regions using linear and nonlinear regression models, neural networks, K-nearest-neighbor, texture analysis, or vegetation indices like NDVI. Semi-process-based approaches infer biomass from canopy parameters estimated by inverting vegetation canopy models. Plants absorb strongly in the visible, most notably the visible red, due to photosynthetic and accessory pigments in the plant tissue. Plants reflect strongly in the near infrared (NIR) due to little absorbance of plant material in this region and scatter by plant cell walls. Whether empirical or semi-process-based, biomass models are typically parameterized by wavelengths around the transition between red absorption and NIR reflectance, termed the “red-edge” (Horler *et al.*, 1983). The major limitations of these biomass estimation techniques in general are red-edge bands may not account for important physical and physiological processes related to biomass (Baret *et al.*, 1989), varying soil conditions, or crop-specific traits (Haboudane *et al.*, 2004). Hyperspectral remote sensing analysis, which commonly involves hundreds or thousands of HNBs in the optical range (400 to 2500 nm), can reveal spectral regions that account for crop-specific physiological traits and minimize soil background effects.

Ustin *et al.* (2004) gives a brief review of hyperspectral analysis using current airborne (e.g., AVIRIS) and spaceborne (e.g., Hyperion on board the EO-1 satellite) remote sensing sensors. The review shows that hyperspectral image analysis has been used to measure several biochemical plant properties related to crop biomass: Leaf Area Index (LAI), fraction of photosynthetically active radiation, moisture status, and stress. Carotenoids, anthocyanin, and chlorophyll increase the range over which light energy can be absorbed for photosynthesis, as well as increase light-use efficiency and provide protection from temperature extremes and ultraviolet radiation (Ustin *et al.*, 2009). These

pigments are sensitive to light from 510 to 520, 540 to 560, and 700 to 730 nm, respectively (Gitelson *et al.*, 2006). Chlorophyll concentration, for example, decreases as crops reach senescence, which causes a blue shift in the red-edge (680 to 730 nm). Crop stress related to nutrient (nitrogen) deficiencies is also sensitive to the light characteristics in this region (Perry and Roberts, 2008). Leaf water absorbs strongly in select bands across the NIR (700 to 1000 nm), Short-Wave Infrared 1 (SWIR1: 1000 to 1700 nm), and Short-Wave Infrared 2 (SWIR2: 1700 to 2500 nm). Crop water content, which is related to biomass, is particularly sensitive to light at 970 and 1180 nm (Champagne *et al.*, 2003), while water stress, a physiological constraint, is better estimated using metrics sensitive to pigment concentration (Perry and Roberts, 2008). Structural carbon (lignin-cellulose) contained primarily in dry plant residues, absorbs strongly in the SWIR2 (Asner, 1998). Lignin-cellulose bands are therefore used primarily to estimate dry plant matter biomass. Given the number of wavelengths involved in hyperspectral analysis, several data reduction and mining techniques have been explored to develop empirical NB biophysical crop models (Thenkabail *et al.*, 2004; Thenkabail *et al.*, 2002). These techniques are empirical and therefore may not be transferable due to a small sample size and limited research area, yet produce better estimates compared to crop biophysical estimates derived from model inversion, an approach that will be more comparable as the physical detail of process-based models improve (Casa and Jones, 2004).

In this study, we use three common spectral transformations and empirically-based biophysical modeling techniques to identify important HNBs sensitive to aboveground fresh crop biomass ( $\text{gm}^{-2}$ ). The study uses a large two-year dataset of ground-based spectroradiometric and aboveground fresh crop biomass data, which is divided into independent subsets for model calibration and validation. The dataset spans the Central Valley of California and includes measurements from the four largest water users (alfalfa, cotton, maize, and rice).

## Methods

### Study Site and Data Collection

A large field campaign in the Central Valley was conducted during the warm crop season (boreal summer) in 2011 and 2012, in which crop biophysical and spectroradiometric data were collected to develop a remote sensing-based crop water productivity model for the entire state. The Central Valley stretches over 700 km and covers an area of over 100,000  $\text{km}^2$  between the Coast range to the west and Cascade, Sierra Nevada, and Tehachapi ranges to the east (Plate 1).

Above-ground fresh biomass spot samples and spectra were collected for California's four largest water users (alfalfa, maize, cotton, and rice) during important phenological stages (sprouting, flowering/silking, and grain/bud-filling). The samples were

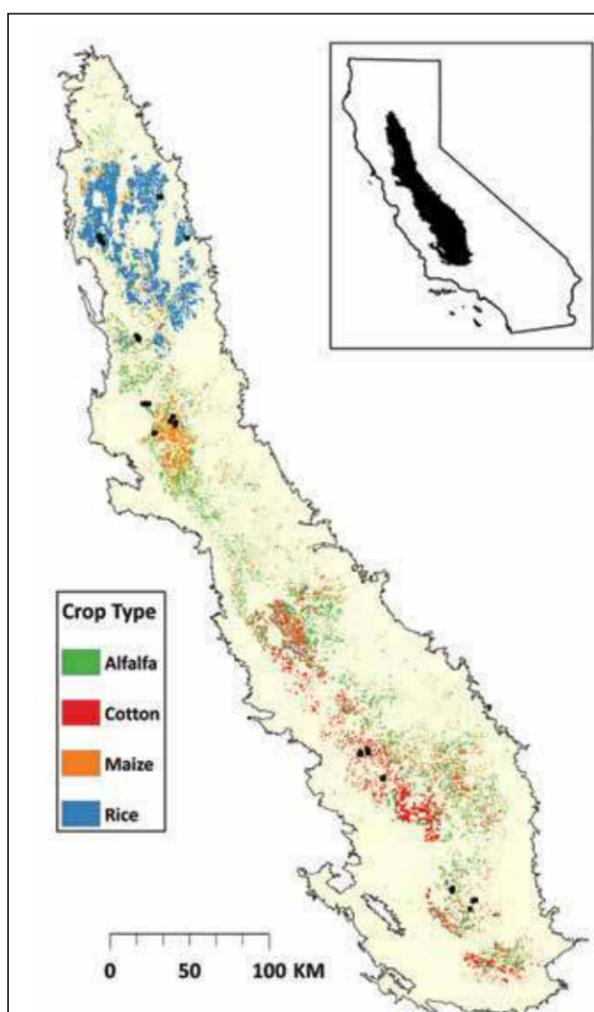


Plate 1. The extent of major water users (alfalfa, cotton, maize, and rice) in California defined by the National Agricultural Statistics Service Cropland Data Layer (<http://nassgeodata.gmu.edu/>) in 2012 overlaid with the five hundred and twenty-seven field spectra and above-ground wet biomass sample pairs in black taken during major phenological phases over two growing seasons (2011 and 2012) in the Central Valley of California.  $N = 44, 65, 84$ , and 85 for rice, alfalfa, cotton, and maize, respectively.

evenly distributed across crop type and phenological stages, as the same fields were visited during each trip. Spot samples of aboveground fresh biomass were measured with a spring scale (precision = 0.1 g). A square-meter quadrant was constructed and used to define the area over which aboveground fresh biomass spot samples were extrapolated, yielding units of  $\text{gm}^{-2}$ . Approximately 2 to 4 spot samples were collected for each of the more than 520 frames used in the analysis. The frames (Plate 1) were distributed across the Central Valley's diverse soil types and climatology. Spot samples of aboveground fresh biomass were extrapolated by multiplying the average mass of the spot samples by the number of plants in the corresponding square meter area. The number of plants in each frame was counted manually from photos cropped to the sample frame and taken during the sprouting phase using a red-green-blue camera. Each photo was collected at a fixed height (1.5 m for alfalfa, cotton, and rice, and 2.5 m for maize) at nadir during  $\pm 2$  hours of solar noon on near cloudless days to minimize solar illumination effects caused by sun angle and shadowing. Height was measured for alfalfa, cotton, and rice using a meter stick and telescoping measuring rod for maize. The biomass of alfalfa was measured differently, because it is a perennial and grows in clusters. Spot measurements of the ground length and width of clusters were taken, averaged, and used to calculate the area of a cluster, assuming the clusters form ellipses. From the photos, the vegetation fraction was estimated and multiplied by a square meter and then divided by the area of a cluster to estimate the number of alfalfa plants, which was then multiplied by the average mass of spot samples taken. This procedure was deemed appropriate to avoid the unnecessary destruction of crops in each field. The mean and standard deviation of the biomass square-meter samples taken across crop type and phenological stage are shown in Table 1.

TABLE 1. SUMMARY STATISTICS OF ABOVEGROUND BIOMASS SAMPLES TAKEN PER CROP TYPE AND VISIT (1 = SPROUTING, 2 = FLOWERING/SILKING, AND 3 = GRAIN/BUD-FILLING) FOR COTTON, MAIZE, AND RICE. BECAUSE ALFALFA IS HARVESTED THROUGHOUT THE GROWING SEASON, THE VISIT DOES NOT NECESSARILY MATCH THE PHENOLOGICAL STAGE. N IS THE NUMBER OF SAMPLES AND  $\sigma$  IS THE STANDARD DEVIATION.

Crop	Visit	N	Mean ( $\text{gm}^{-2}$ )	$\sigma$ ( $\text{gm}^{-2}$ )
Alfalfa	1	54	7694.46	11117.34
	2	57	3187.17	3241.97
	3	60	14647.11	13715.95
Cotton	1	58	800.95	759.89
	2	59	7908.44	3590.35
	3	57	9713.79	5924.20
Maize	1	54	8148.00	3646.59
	2	60	13824.53	2973.18
	3	59	12186.12	2942.21
Rice	1	44	800.39	487.94
	2	50	2703.15	1089.45
	3	47	2574.64	1006.03

Field spectra were collected for each sample frame with an Analytical Spectral Devices (ASD: [www.asdi.com](http://www.asdi.com)) portable spectroradiometer (Field Spec Pro 3), which has an optical range of 350 to 2500 nm resampled to 1nm resolution. A pistol grip and  $18^\circ$  field of view (FOV) fore-optic were attached to the fiber optic cable bundle, which records light from the crop canopy. The  $18^\circ$  FOV fore-optic was selected, because it shows greater spatial and spectral uniformity over the area collected, compared to smaller fore-optics employed with the

ASD (MacArthur *et al.*, 2012). As with the photos, each series of spectra were collected at a fixed height (1.5 m for alfalfa, cotton, and rice, and 2.5 m for maize) above the surface at nadir between  $\pm 2$  hours of solar noon on near cloudless days to minimize solar illumination effects caused by sun angle and shadowing. The Field Spec Pro 3 records raw radiance values, which were collected 30 times or more under sub-optimal field conditions (e.g., high winds) and averaged internally. Raw radiance ( $\text{Wsr}^{-1}\text{m}^{-2}$ ) was converted internally to percent reflectance using calibration spectra collected every 2 to 10 minutes depending on field conditions and near each sample frame with a white standard panel composed of  $\text{BaSO}_4$ . Over the course of each visit, current generated by the detector (i.e., "dark current") or passing the fore-optic from dark objects to bright objects or vice versa typically saturated the signal, which necessitated detector re-optimization. The internally processed reflectance spectra were collected at five different locations within each square-meter frame to reduce uncertainties in spectroradiometric FOV and mixed effects from crop, soil, and shadow. Five reflectance spectra collected randomly within each frame where biomass was estimated, yielded nearly 3,600 recorded spectra over the two-year period.

#### Data Processing

The spectra underwent additional preprocessing steps, before statistical models relating biomass to spectral reflectance were developed. The ASD Field Spec Pro 3 fiber optic cable collects light and diverts it to three detectors: 350 to 1050, 900 to 1850 nm, and 1700 to 2500 nm. The spectroradiometer automatically rectifies the overlap between each sensor, but slight differences across detectors were observed, so inter-sensor normalization was performed by multiplying visible/NIR and SWIR2 detector reflectance by near-edge SWIR1 detector ratios. This had a minimal impact on the spectra and was used primarily for visualization perhaps. The standard deviation of the five sample spectra was computed in the SWIR2, which was typically the most spurious, and spectra with values greater or less than one standard deviation were omitted, before final averaging. Once the spectra were averaged, so that one spectrum represented one sample, wavelengths where strong greenhouse gas absorption occurs, were omitted. These included 350 to 390nm ( $\text{O}_3$ ), 1350 to 1450 nm ( $\text{H}_2\text{O}$  and  $\text{CO}_2$ ), 1790 to 2000 nm ( $\text{H}_2\text{O}$  and  $\text{CO}_2$ ), and 2300 to 2500 nm ( $\text{H}_2\text{O}$  and  $\text{CO}_2$ ). In order to reduce the number of wavelengths analyzed, the 1 nm spectra collected by the spectroradiometer were averaged to 10 nm HNBs considering the full width at half maximum, but matching the 400 to 2500 nm 10 nm HNBs detected by the Hyperion sensor. In order to reduce potential data redundancies and make computations more efficient, previous studies have used similar 10 nm HNBs to build accurate empirical hyperspectral biophysical models (Thenkabail *et al.*, 2004; Thenkabail *et al.*, 2002). In addition, in a process known as the Hughes effect (Hughes, 1968), biophysical simulation accuracy may decrease with the initial introduction of more predictors during model-building. The aggregations yielded 196 10 nm matching Hyperion bands, as some of the 242 Hyperion bands were erroneous and were not included in the analysis. For the remainder of this paper, the 196 HNBs are expressed using the 10 nm wavelength centroids.

Spectra were inherently mixed, because leaf litter, soil, and other background features interfere with the vegetation signal. In remote sensing, methods such as band ratioing reduce the effect of background features, but assume that these features vary consistently across samples and wavelengths (Hall *et al.*, 1990). First (Demetriades-Shah *et al.*, 1990), second (Hall *et al.*, 1990), and both first and second (Elvidge and Chen, 1995) derivatives were used to transform spectra in order to reduce the effects of soil background, as the rate of change of the background signal tends to change more gradually than

vegetation at major spectral inflection points. Inflection points (e.g., red-edge) are wavelengths at which dramatic changes in absorption by and reflectance from the canopy occur. Cubic spline functions were fit to each spectrum using the “stats” package available in R® ([www.r-project.org](http://www.r-project.org)) and the first and second derivative of the function at each wavelength was computed and used along with the untransformed spectra for model-building. Other transformations commonly used in hyperspectral analysis, such as inverse-log, which generates a pseudoabsorbance curve (Serrano *et al.*, 2002), were also compared, but showed no advantage over the untransformed or derivative spectra during the model-building process.

Additional preprocessing was performed by comparing HNBs (predictors) to the biomass data itself. In this phase, the biomass was log-transformed, and a 70 to 30 percent split random sample stratified by crop type was created for model calibration and validation, respectively. This yielded 368 samples for calibration and 159 samples for validation. Outliers were detected from scatter and studentized residual plots of biomass for each crop type versus two-band HVIS identified to be sensitive to crop biomass in Thenkabail *et al.* (2013) and from biomass and two-band cross correlation matrices detailed in the *Single and Two-band HNB Band Relationship Subsection*. In each case, the best two-band HVI using this method, one red and NIR HVI (651 and 855 nm), one green HVI (550 and 651 nm), and one SWIR1 HVI (651 and 1649 nm) were chosen to generate the plots. Potential outliers were flagged as data points with absolute studentized residuals greater than two. The outliers were typically common to all plots for each crop type, however the outliers were further scrutinized before a final decision was made on omission. In the calibration dataset, five outliers were detected for alfalfa, cotton, and rice, while only three were detected for maize. In the valida-

tion dataset, four outliers were detected for rice, alfalfa, and maize, while only one outlier was detected for cotton. Outliers were due to corrupted data or irregular sampling, namely large biomass when vegetation was sparse (Figure 1a and 1b) and anomalously high or low spectra across wavelengths, which over-predicted or under-predicted biomass, respectively (Figure 1c and 1d). The former occurred when a biomass sample was taken outside the spectroradiometer FOV, because biomass was insufficient within the spectroradiometer FOV to take a sample. The latter occurred when white reflectance calibration was infrequent in the presence of clouds, sun angle was low, or the fore-optic was off-nadir.

#### Band Selection for Biomass Prediction

The development of an empirical HNB biophysical model is complex; several bands must be analyzed together, which often exhibit strong multicollinearity; the number of predictors often exceeds the number of dependent samples, so that rules must be applied to reduce the Hughes effect; and model parsimony must be achieved to prevent over-fitting. From the data mining techniques employed by the Thenkabail studies highlighted in the introduction, we selected three to develop empirically-based predictions of crop biomass: two-band NB HVIS selected from lambda-lambda ( $R^2$ ) contour plots, multiple band-HVIS (MB-HVIS) developed from Sequential Search Methods (SSM), and MB-HVIS developed from Principal Component Regression (PCR). Each method has unique advantages and disadvantages, which will be discussed below. Single band correlation plots were used to show which spectral ranges (visible, NIR, SWIR1, and SWIR2) had the strongest relationship with biomass and to detect important inflection points.

#### Single and Two-band HNB Band Relationship

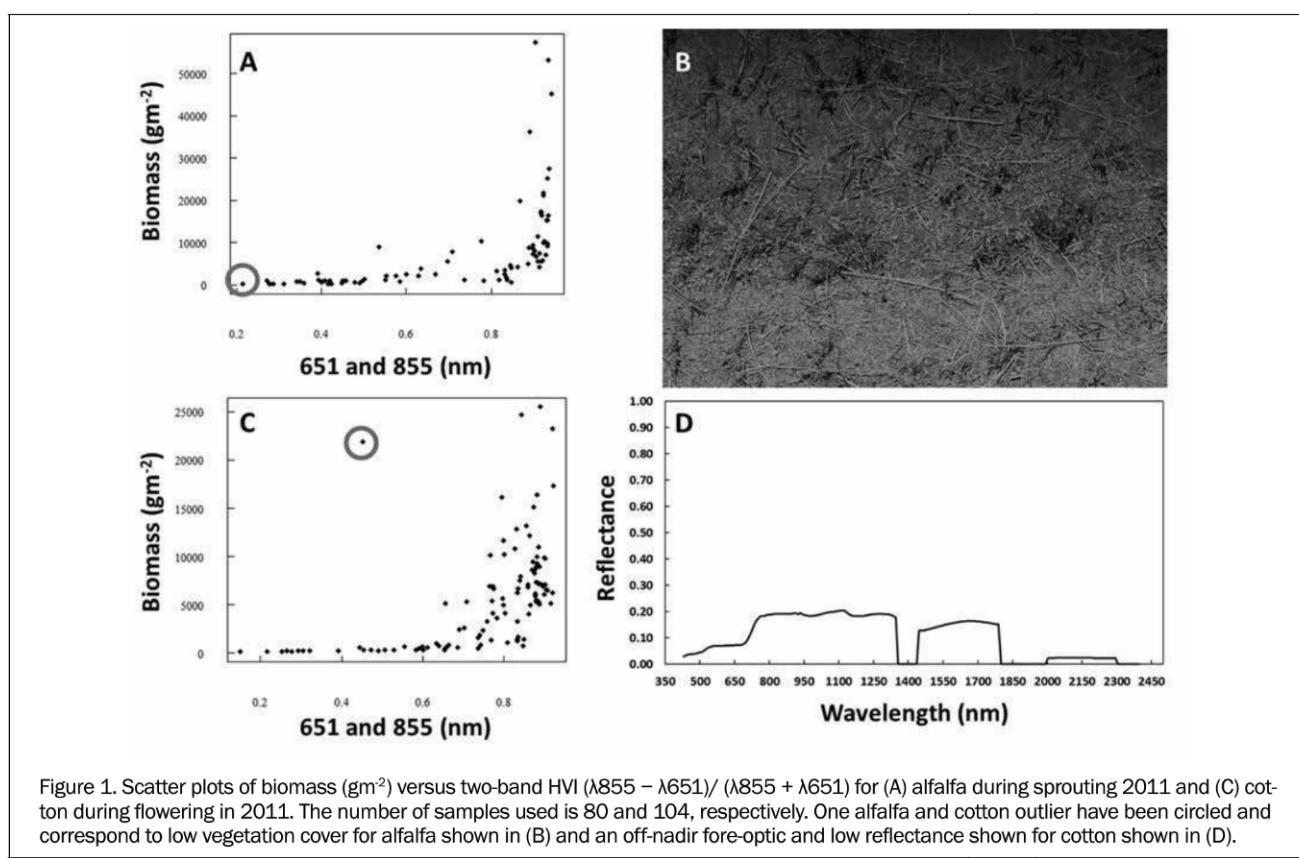


Figure 1. Scatter plots of biomass ( $\text{gm}^{-2}$ ) versus two-band HVI ( $\lambda 855 - \lambda 651$ )/ ( $\lambda 855 + \lambda 651$ ) for (A) alfalfa during sprouting 2011 and (C) cotton during flowering in 2011. The number of samples used is 80 and 104, respectively. One alfalfa and cotton outlier have been circled and correspond to low vegetation cover for alfalfa shown in (B) and an off-nadir fore-optic and low reflectance shown for cotton shown in (D).

For the single HNB band analysis, Pearson R correlations were made with each 10 nm HNB, showing both the strength and direction of the relationship with biomass. The correlations, as well as other statistical procedures discussed in the remainder of the paper were performed in R. Plots were made for each transformed dataset (untransformed, first derivative, and second derivative) and for every crop type. Initially, the analysis was performed for one mega dataset that included all crop types, but the results of the analysis were poorer than per crop type, so no further analysis was performed on the mega dataset.

Further processing was performed on the second derivative transformed data, as the correlations between the second derivative transformed spectra and biomass were highly variable. In order to smooth out inconsistencies between second derivative transformed spectra, the absolute values of HNB reflectance were integrated over important inflection points per Elvidge and Chen (1995). Various integrating windows were evaluated (25 to 300 nm at 25 nm intervals). The correlations were consistently lower than with the non-integrated second derivative transformed spectra and were therefore not included in model building.

For the two-band NB band analysis, Lambda-lambda ( $R^2$ ) contour plots for each transformation and crop type were created by correlating every possible two-band combination of the 196 bands with biomass. Analogous to NDVI the vegetation indices were calculated as follows:

$$HVI = \frac{R_2 - R_1}{R_2 + R_1} \quad (1)$$

where HVI is the two-band HVI,  $R_1$  is the untransformed or transformed reflectance band one, and  $R_2$  is the untransformed or transformed reflectance band two.

#### *Multiple Band HVI Using Sequential Search Methods*

Sequential search methods, like stepwise and forward addition (FA) regression, iteratively select the best predictor and then incrementally include predictors successively having lower, but significant partial correlation with the residuals (Hair, Jr. et al., 1998). In this study, we used FA, because the number of predictors exceeded the number of samples for each crop type. A subset of the data used to calibrate two-band HVIs was used for SSM, because SSM requires that missing wavelengths are consistent across all samples. A common rule of thumb in statistics to prevent over-fitting is to build a model with less than  $20 \times (\text{number of predictors})$ ; this criterion was used during the model building process. Several methods for testing the significance of the partial correlations exist. We used the Akaike Information Criterion (AIC), which is a function of both the maximized log-likelihood and number of predictors at each incremental stage of model-building. Lower AIC values, therefore, not only indicate better model fit, but greater model parsimony as well. Each step was scrutinized and a final model was selected that had low AIC, significant predictors at the 99.9 percent confidence band, and explained at least an additional two percent of the biomass variance ( $\Delta R^2 \geq 0.02$ ).

#### *Multiple Band HVI Using Principal Components Regression*

Principal component regression uses component scores derived from a Principal Component Analysis (PCA) as predictor variables (Rodarmel and Shan, 2002). The purpose of PCA is to transform large collinear datasets to reduce interdependencies and yield factors (linear combinations) of input variables that explain unique proportions of the total variance. This is carried out in one step, which makes it contrary to SSMs, which analyze predictors individually and may remove important, but collinear predictors. The first component explains the most variance, while subsequent components explain successively less variance. The smallest components typically ex-

plain the error variance. The transformation is performed by the eigenvalue decomposition of the correlation or covariance matrix. Transformations on the correlation matrix are typically performed to standardize the data. The linear transformation is defined as follows:

$$y_{k(i)} = w_k \cdot x_i$$

where  $y_{k(i)}$  is the score for sample  $i$ ,  $w_k$  is the loading for variable  $k$ , and  $x_i$  is the sample value at variable  $k$ . The square of the loading, as the name implies, is the contribution of the variable to the factor total variance. The scores are the values of the component for each sample and are used to determine the regression coefficients in single or multiple PCR analysis.

The PCA was performed on a subset of the wavelengths and biomass was then regressed against the scores. Principal Components Analysis, like SSM, requires that each sample has the same missing wavelengths, so some of the samples with missing wavelengths inconsistent with the full calibration dataset used to develop the two-band HVIs could not be used. The data was mean centered and the transformation was performed on the covariance matrix, because the reflectance ranged from 0 to 1 at all wavelengths. Orthogonal rotations maximize variable loading on a single factor, which makes predictor loadings more interpretable, so orthogonal (VARIMAX) transformations were analyzed alongside the unrotated data. Oblique transformations were not used, because the procedures are less developed and subject to debate (Hair Jr. et al., 1998). Scree plots, which show the contribution of each component to the total variance, along with correlations and significance tests ( $<0.001$ ) from the PCR, were used to determine the number of components to include in the final PCR models.

#### **Model Validation**

Validation data for each model category was handled differently, however all approaches yielded scatterplots of actual versus predicted biomass and summary statistics. In order to validate the two-band HVI approach, the HVI that ranked first in the calibration subset was used for model validation. Many of the two-band HVIs, however had similar correlations and could have been used alternatively. The SSM approach was the most straightforward: the equations derived from the calibration dataset parameterized the validation HNBs. For PCR, the loadings were multiplied by the validation HNBs and summed to yield a score across components, which was then used to create a linear model with log-transformed biomass.

## **Results**

#### **Single HNB Relationship**

Pearson correlations (R) reveal the strength and direction of the HNB-biomass relationships when all the crops are considered and for crop type subsets. Generally, the patterns of the single HNB analysis for the crop type subsets shown in Figure 2 are consistent with the “all crops” results; however, the correlations are lower for the latter (maximum  $R < 0.41$ ). Because of the poor results of the all crops analysis, no further analyses are performed on these data. The number of samples ( $N$ ) used to develop the correlation plots by crop type are 60, 80, 104, and 106 for rice, alfalfa, cotton, and maize, respectively. Figure 2 also includes the maximum and minimum Pearson correlation and the wavelength centroid at which it occurs. Pearson correlations per crop type are highest in the visible and NIR, near the red-edge. Correlations between log-transformed biomass and reflectance outside the red-edge region are also high across much of the visible and NIR, while relatively lower correlations exist in the SWIR. Rice, alfalfa, and cotton generally show high positive correlations with

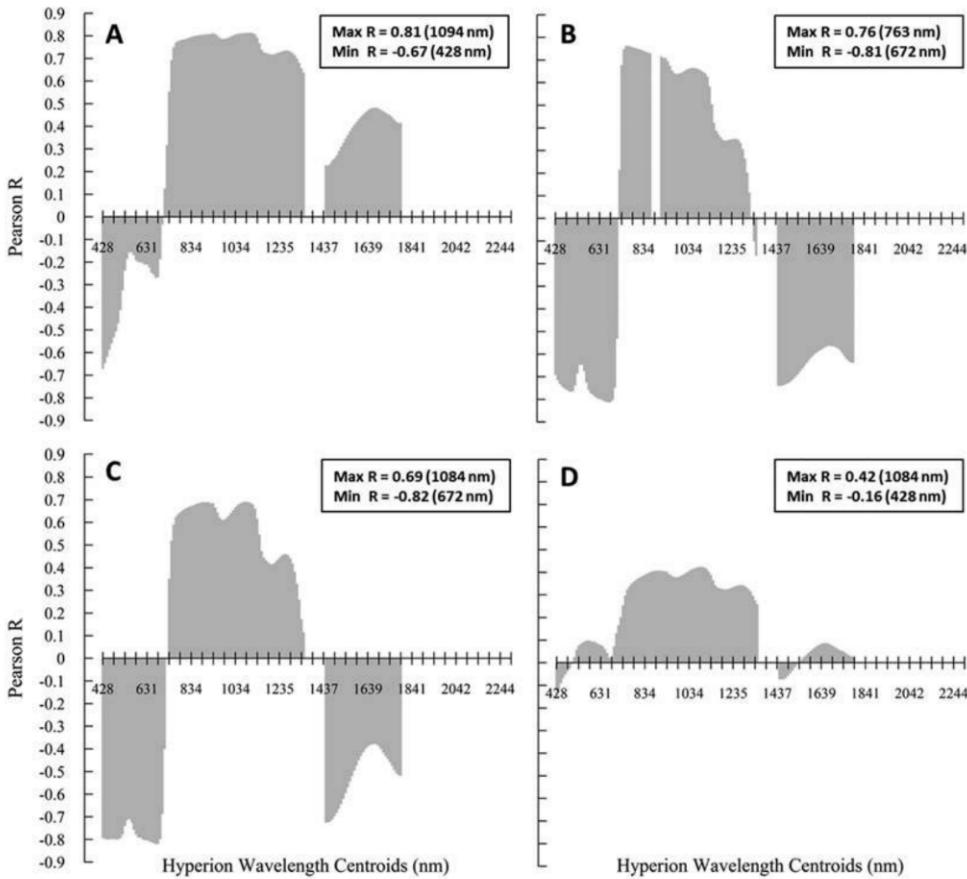


Figure 2. Pearson correlation coefficients ( $R$ ) between 186 untransformed and discrete (10 nm) channel reflectance and crop biomass for (A) rice, (B) alfalfa, (C) cotton, and (D) maize. Ten of the bands at the end of the SWIR2 have been removed, because of irregularities between the samples.

NIR reflectance and high negative correlations with visible reflectance. Correlations in the red-edge are high, but the direction is less consistent across crop types. Rice had the highest correlations in the NIR, followed by alfalfa, cotton, and maize. Unlike alfalfa and cotton, rice biomass is sensitive to a much narrower range of HNBs in the visible and correlations are lower. In addition, biomass is positively correlated to SWIR reflectance. Alfalfa shows high correlations particularly near the red-edge and the highest correlations in the SWIR compared to other crops. The results for maize are less comparable with other crop types and correlations are lower. Maize shows much higher correlations around important inflection points (468, 560, 672, 1013, 1145, 1215, 1225, 1235, and 1326 nm) after the first derivative transformation of HNBs (Figure 3). Correlations for the other crops are also higher after the first derivative transformation, but over a much narrower range of HNBs. This is particularly the case around major water absorption bands in the SWIR. In addition, unlike the untransformed spectra, the direction of correlations is consistent across crop types in the SWIR. Second derivative correlations (not shown) are more variable than the first derivative transformed spectra. Correlations overall are typically lower than the first derivative transformed or untransformed spectra.

#### Two-band HVIs

The strength of two-band HVIs is illustrated with lambda-lambda ( $R^2$ ) contour plots in Plate 2. The dark red regions in Plate 2 pinpoint where two NBs combine in an NBVI that correlates highly with biomass. As in the 1D plots, the correlations between reflectance ratios and biomass are the highest for rice, but over a much narrower range compared to alfalfa and cotton. Again, maize shows the lowest correlations, with none of the 2-band ratios having an  $R^2 > 0.5$  with biomass. The best predictor of biomass for rice is at 1205 and 1256 nm ( $R^2 = 0.83$ ), which is within a narrow range of the NIR from 1185 to 1276 nm. Strong, but relatively lower correlations occur across the NIR centered at 743 nm and visible at 550 nm. A wider range of relatively high correlations occurs between SWIR (1326 to 1740 nm) and NIR bands. Alfalfa two-band ratios show high correlations with biomass across the visible, NIR, and SWIR. Alfalfa is the only crop to show significant correlations in the SWIR (2032 to 2274 nm). Red-edge ratios (611 to 712 nm) are the highest correlated with alfalfa biomass- the best being at 641 and 702 nm ( $R^2 = 0.80$ ). These are followed by red-edge and NIR, NIR and SWIR, SWIR and SWIR, and lastly SWIR2 and SWIR2 ratios. As with alfalfa, cotton shows high correlations across the full measured spectrum, except for the SWIR2-SWIR2 ratios. The highest correlations occur between the visible and NIR, particularly around 550 nm and 722 to 1326 nm. The highest is at 550 and 1124 nm ( $R^2 = 0.83$ ).

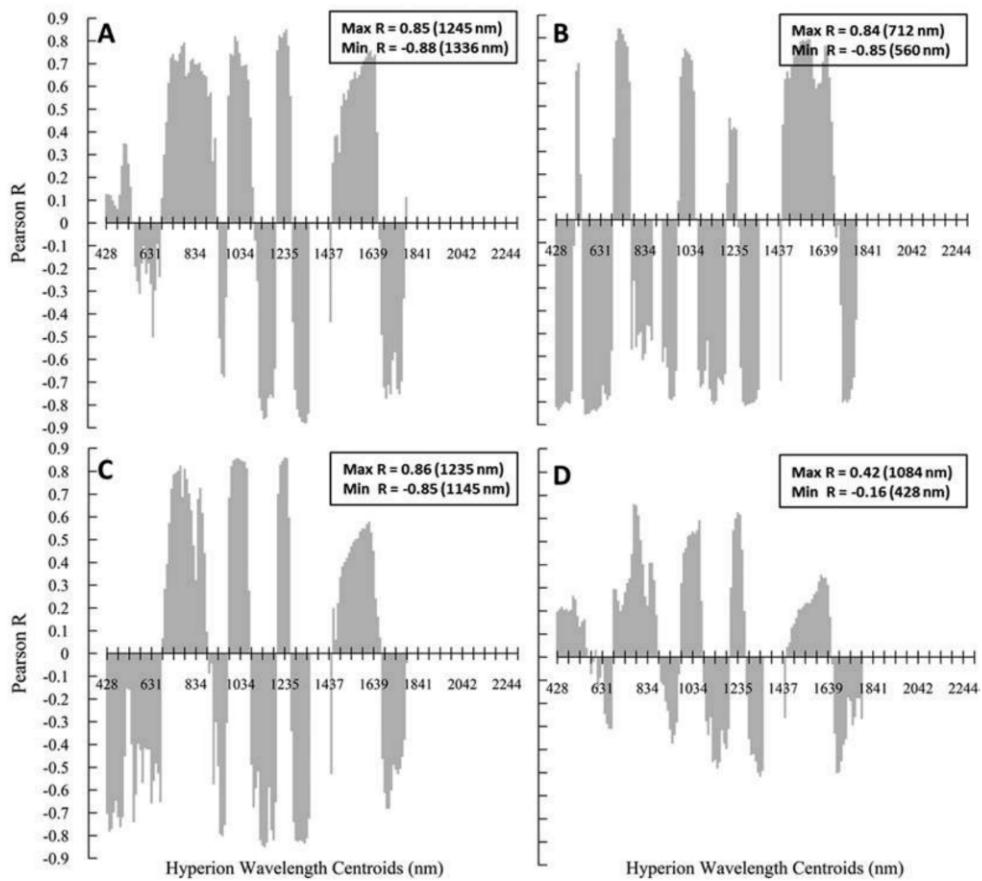


Figure 3. Pearson correlation coefficients ( $R$ ) between 186 first derivative transformed and discrete (10 nm) channel reflectance and crop biomass for (A) rice, (B) alfalfa, (C) cotton, and (D) maize. Ten of the bands at the end of the SWIR2 have been removed, because of irregularities between the samples.

Weaker correlations are found between the NIR and SWIR, followed by the SWIR and SWIR. Unlike the other crop types, maize shows the highest correlations between SWIR and SWIR bands (1629 to 1730 nm). The highest correlation is at 1669 and 1699 nm ( $R^2 = 0.49$ ). As with the single band plots, the first (Plate 3) and second derivative transformed spectra correlated over a much narrower range of bands than the untransformed spectra. Correlations between biomass and first derivative transformed spectra are comparable to the untransformed spectra correlations, while second derivative transformed spectra correlations (not shown) are lower. Table 2 shows the top ten ranked two-band predictors using first derivative transformed spectra from the calibration subset. As with the untransformed data, rice shows high correlations, but over a much narrower range of wavelengths compared to cotton and alfalfa. The highest correlations for rice occur between the visible green and red-edge and NIR around 1225 nm. The highest correlations for alfalfa and cotton occur between the visible blue and red-edge/NIR bands. Correlations between the red-edge and NIR are not as high for cotton as for alfalfa, nor are the correlations between visible and SWIR bands. Maize, again, did not have any ratios with  $R^2 > 0.5$ . The highest correlations are in the NIR.

#### Multiple Band-HVIs (Sequential Search Method)

Multiple band-HVIs that correlated strongly with biomass for each crop type and explained the most unique variance, while meeting the sample size criterion are shown in Table 3. The HNBs are combined additively using the FA method. The table shows the wavelength centroids used in the FA model for each crop and transformation used (untransformed, first derivative, and second derivative). No one transformation consistently shows the highest correlations with biomass across crop type, however, the second derivative transformation is the best for two of the crops (rice and alfalfa). For rice, the best predictors are in the visible and NIR. In the case of the untransformed data, the 973 nm and 2052 nm HNBs for the untransformed and 1<sup>st</sup> derivative transformed spectra would have explained an additional 2 percent of biomass variance, but were not included, because the sample size was too small or the predictor was not significant, respectively. For cotton, the best predictors are found across the spectrum, the most significant in the visible and red-edge. In the case of the untransformed and first derivative transformed data, the addition of wavelength centroids would have met the  $R^2$  criteria, but risked over-fitting. For maize, the best predictors are found primarily in the NIR, with the exception of the second derivative transformed data, which included an important SWIR band (2042 nm). Additional bands could have been added that account for significant unexplained variance by the second derivative

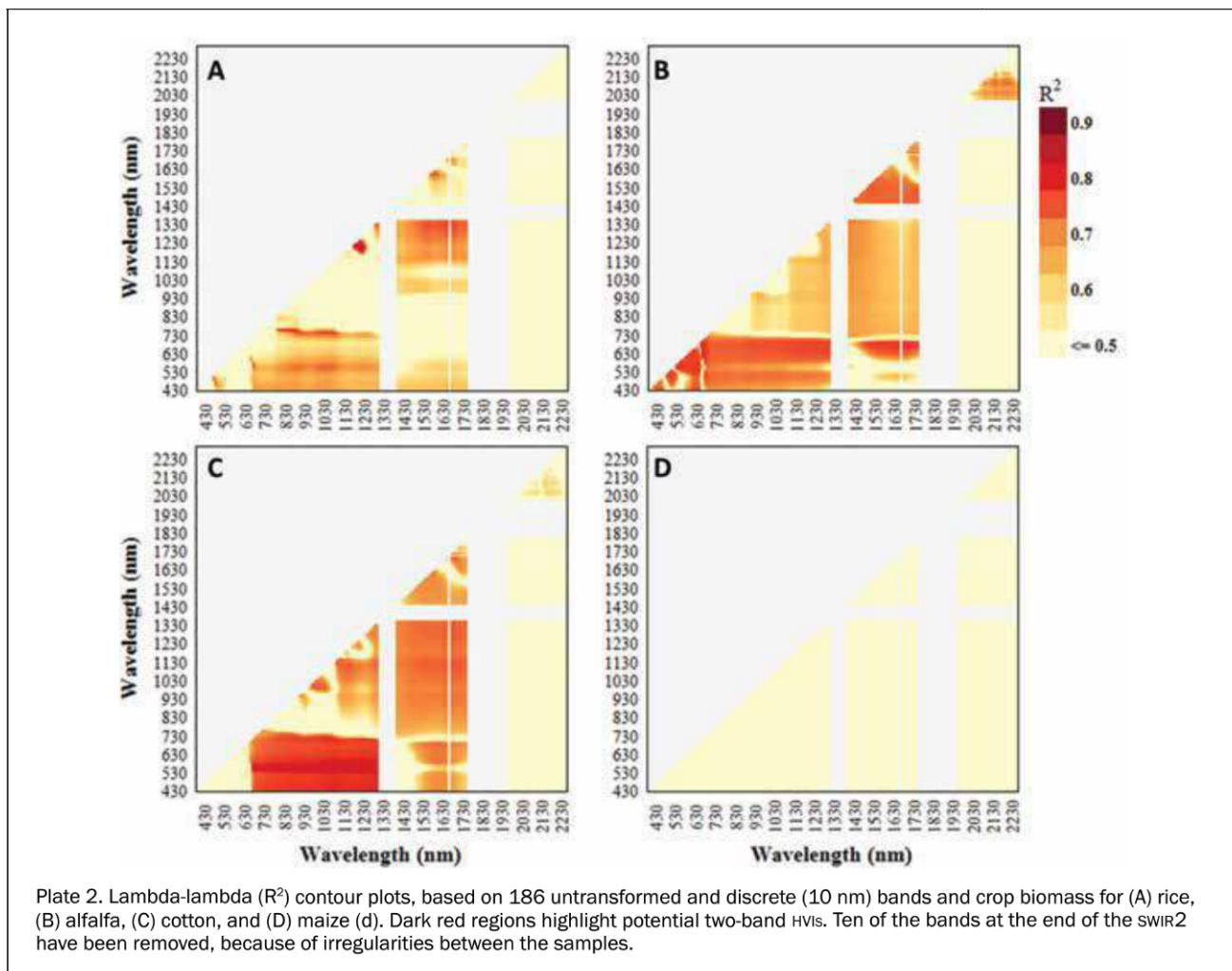


Plate 2. Lambda-lambda ( $R^2$ ) contour plots, based on 186 untransformed and discrete (10 nm) bands and crop biomass for (A) rice, (B) alfalfa, (C) cotton, and (D) maize (d). Dark red regions highlight potential two-band HVIs. Ten of the bands at the end of the SWIR2 have been removed, because of irregularities between the samples.

transformed data, but is restricted by the sample size.

#### Multiple Band-HVIs (Principal Component Regression)

Like FA, PCA is performed on a subset of samples containing consistent missing HNBs. This yielded the same sample subset as with FA. Principal Components Analysis requires that the number of predictors is less than the number of samples, so for each crop, predictors with Pearson correlations less than a given threshold are omitted from the PCA to meet this criterion. For rice, alfalfa, cotton, and maize, these thresholds are |0.768|, |0.658|, |0.561|, and |0.117|, respectively. The data is also rotated to aid in interpretation of the component loadings. The rotated loadings, which describe the relative contribution of each HNB to total component variance, yield correlations with biomass similar to the unrotated scores. The orthogonally rotated loadings for each HNB greater than |0.7| are shown in Table 4. Since the data in Table 4 is rotated orthogonally, the first two components of each crop type explain nearly half of the total variance. Components not shown in Table 4, are either significant in the final PCR model, but explained a small proportion of biomass variability, or are insignificant at the 99.9 percent confidence band. The first component for rice explains 50 percent of the total variance, while the second component explains just less than 50 percent of the total variance. The first component consists of wavelengths exclusively in the NIR (753 to 1276 nm).

Loadings that explain 50 percent or more of the total factor variance (weight > |0.7|) are from 773 to 794 nm on the first component and from 943 to 1276 nm on the second component. The first component of alfalfa explains 54 percent of the total variance, while the second component explains 44 percent of the total variance. The first component shows high loadings in the visible/red-edge (428 to 702 nm) and the SWIR (1477 to 1548 nm), while the second component shows high loadings in the NIR (743 to 1074 nm). The first component of cotton explains 49 percent of the total variance, while the second component of cotton explains 48 percent of the total variance. The first component has the highest loadings in the NIR (743 to 1134 nm), while the second component has the highest loadings in the visible/red-edge (428 to 702 nm) and the SWIR (1447 to 2012 nm). For maize, the first component explains 70.5 percent of the total variance, while the second component explains 22.0 percent of the variance. The first component loads high in the NIR (722 to 1346 nm), while the second component loads high in the SWIR (2002 to 2295 nm). The first two components are selected for rice, alfalfa, and cotton, while the first three components are selected for maize to compute scores, which are used to develop PCR models.

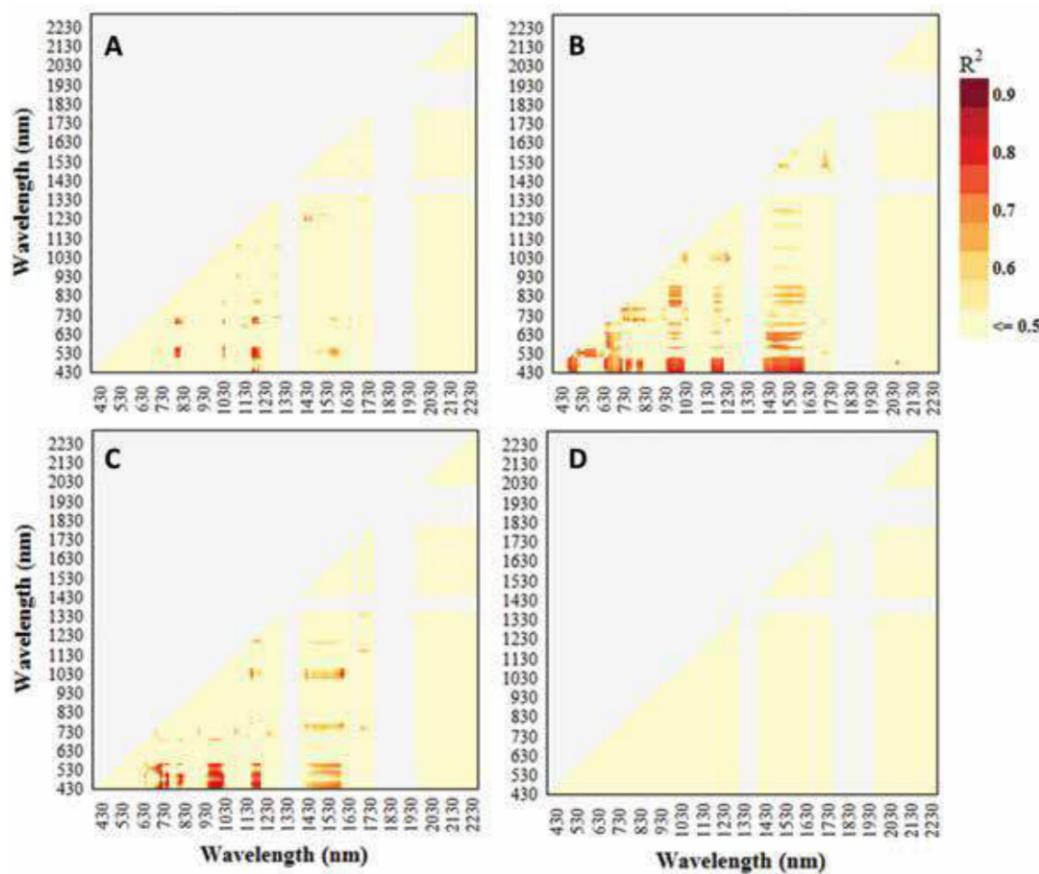


Plate 3. Lambda-lambda ( $R^2$ ) contour plots, based on 186 first derivative transformed and discrete (10 nm) bands and crop biomass for (A) rice, (B) alfalfa, (C) cotton, and (D) maize. Dark red regions highlight potential two-band NBVs. Ten of the bands at the end of the SWIR2 have been removed, because of irregularities between the samples.

#### Two-band and Multiple Band Model Validation

The models selected from the calibration subset using 2-band NBVs, multi-band FA, and multi-band PCR for the untransformed and first derivative transformed spectra are used to evaluate the models with the validation subset. The first derivative transformed NBs have higher correlations with the validation subset using all three approaches. When correlating NBVs calibration and validation biomass subsets, the ranks of such correlations are more consistent for the first derivative transformed spectra as well. Overall, the predicted biomass derived from the three approaches tends to underestimate actual biomass when biomass is low, while higher biomass predictions tend to be more variable.

The scatterplots of actual biomass versus predicted biomass using 2-band NBVs are shown in Figure 4. With the exception of maize, the top NBVI from the calibration subsets for each crop type performed well, but worse than during the calibration analysis. Many of the NBVs ranked below the top NBVI shown in Table 2, performed as well, if not better on the validation subset. These were primarily in the visible green and NIR for alfalfa and visible blue and NIR for cotton. Rice was more divergent, showing high performing NBVs on the validation subset between the visible green and NIR, as well as the NIR and SWIR1.

The correlations between the FA models shown in Figure 5 and actual biomass are comparable to the 2-band NBVI's. As

with the 2-band NBVs, low biomass tends to be underpredicted, while higher biomass predictions tend to be more variable. Overall, the FA approach performs better than the selected NBVs on the validation subset. The deviations from the 1:1 line (residuals), indicated by the Root Mean Squared Error (RMSE), are higher than the NBVI approach for rice and alfalfa, but lower than the NBVI approach for cotton and maize. Unlike the two-band approach, rice only uses one predictor near the SWIR (1336 nm), while the other crops use three predictors in the FA models. Alfalfa includes two SWIR predictors (1447 and 1770 nm), but NIR HNBs tend to enter the model first. The PCR models shown in Figure 6 yield the lowest correlations with actual biomass and have higher RMSE.

#### Discussion

The study makes three important contributions to the design of upcoming satellite research missions with the intent of monitoring water-intensive crops across the globe: (a) biomass, a key component of crop water productivity, can be simulated using crop-specific spectral properties determined across various crop varieties, soil types, climatology, and growth stage; (b) first derivative transformed spectra minimize mixed spectra effects and yield higher correlations particularly at longer wavelengths; and (c) sequential search methods are more robust and explain more biomass variance than

TABLE 2. THE TOP TEN TWO-BAND HVIS FOR THE CALIBRATION SUBSET, RANKED USING  $R^2$ ; R1 AND R2 CORRESPOND TO EQUATION 1

Crop	R1	R2	$R^2$
Alfalfa	682	458	0.81
	1256	438	0.81
	1205	438	0.81
	1629	438	0.81
	993	438	0.81
	1619	438	0.80
	993	428	0.80
	1629	428	0.80
Cotton	1619	428	0.80
	794	458	0.86
	763	438	0.86
	763	428	0.86
	794	489	0.84
	1023	438	0.84
	1013	438	0.84
	1034	438	0.84
Maize	1044	438	0.84
	1054	448	0.84
	1205	773	0.46
	1205	794	0.45
	1215	773	0.43
	1225	773	0.42
	1649	773	0.42
	1629	773	0.41
Rice	1639	773	0.41
	1619	773	0.40
	1235	773	0.40
	1225	529	0.82
	1225	519	0.81
	1225	509	0.81
	1225	539	0.80
	845	539	0.79

other statistical techniques, but are less flexible than two-band HVIs. Given the success of these techniques compared to previous ground-based HNB studies, further investigation is required with space-borne hyperspectral sensors to prove their validity across scales.

The single band analysis of the untransformed spectra shows that biomass is negatively (positively) correlated to a large portion of the visible (NIR) spectrum, reflecting the preferential absorption (scatter) of plant chlorophyll as crops mature through the growing season. Correlations vary across each crop type, perhaps reflecting their unique canopy architecture and physiology. And this undoubtedly impacts the significance of various bands when transformed and/or analyzed using two-band HVIs, SSMs, or PCA. Rice biomass, for example, correlated with HNBs over a much narrow range than alfalfa and cotton. Rice has an erectophile (vertical) leaf area distribution (LAD), especially during midday and under well-watered (irrigated) conditions (Moran *et al.*, 1989)

TABLE 3. RESULTS OF THE FORWARD ADDITION SSM RESULTS FOR UNTRANSFORMED, FIRST DERIVATIVE (1), AND SECOND DERIVATIVE (2) REFLECTANCE AND EXPLAINED VARIANCE OF BIOMASS FROM THE CALIBRATION SUBSET FOR EACH CROP TYPE

Crop Type	Transformations	Band Centers (nm)	$R^2$
Alfalfa (N=65)	Untransformed	672; 733; 1447	0.74
	Derivative(1)	560; 1447; 1770	0.77
	Derivative(2)	448; 763; 1750	0.82
Cotton (N=84)	Untransformed	672; 963; 1114	0.87
	Derivative(1)	438; 672; 1145	0.87
	Derivative(2)	478; 661; 1155	0.83
Maize (N=85)	Untransformed	763; 845; 1023; 1094	0.52
	Derivative(1)	672; 773; 783	0.58
	Derivative(2)	702; 1134; 1155; 2042	0.51
Rice (N=44)	Untransformed	428; 1094	0.78
	Derivative(1)	1336	0.82
	Derivative(2)	1104; 1155	0.85

when spectra were collected, while cotton and alfalfa have a planophile (horizontal) LAD. Light is attenuated much more readily in erectophile canopies, while planophile crops, on the other hand, show greater absorption (scatter) in the visible (NIR) bands, because their leaf area has greater exposure to light during midday under well-watered conditions. Alfalfa and cotton also showed similarities in HNB selection during the model-building process, highlighting the potential of aggregating these crops and those with similar structure in future biomass models. The positive correlation between SWIR bands and rice biomass is surprising, as leaf water content and biomass increase as plants mature before senescence, typically leading to greater absorption in the SWIR. The presence of water in the rice fields could offer an explanation. The rice under study is submerged in 20 cm of water on average when the spectra are collected. When rice is sprouting, water is plainly visible and therefore SWIR absorption by the spectroradiometer is high. As the plant matures and its density increases, however, the canopy obscures the signal, leading to relatively less SWIR absorption. This hypothesis could be explored further, as in Gnyp *et al.* (2014), by performing the analysis at each major development stage and potentially reveal other important SWIR bands that predict growth for each successive stage.

The study was purposely performed over various soil types, irrigation methods, and at different times of the watering regime, as these can influence the performance of visible, NIR, and SWIR spectra in measuring crop biomass. The higher correlations particularly in the NIR and SWIR, after soil and water effects were minimized using the first derivative transformation, reveal the impact background can have on estimating biomass at longer wavelengths. This is reiterated by the higher positive (NIR) correlations, particularly around strong absorption bands, after the transformation for crops with relatively less dense canopies and more exposure to soil (alfalfa and cotton) compared to rice. The second derivative transformation further reduces solar illumination effects (Tsai and Philpot, 1998), particularly for rice in the early growth stage when the background consists mainly of water (Gnyp *et al.*, 2014). Although the second derivative transformed spectra perform well in the additive models, under validation, the correlations are lower, perhaps reflecting the sensitivity of the transformation to over-fitting or poor performance of this transformation

TABLE 4. PRINCIPAL COMPONENTS SELECTED TO BUILD PCR MODELS OF CROP BIOMASS (ALFALFA, COTTON, MAIZE, AND RICE). THE COMPONENTS ARE BASED ON THE FIRST DERIVATIVE TRANSFORMATION OF REFLECTANCE. THE WAVELENGTH CENTROID LOADINGS  $>|0.7|$  ARE HIGHLIGHTED FOR EACH CROP.

Band Centers (nm)	Rice		Alfalfa		Cotton		Maize		
	PCA1	PCA2	PCA1	PCA2	PCA1	PCA2	PCA1	PCA2	PCA3
428		0.89		0.78					
438		0.87		0.78					
448		0.86		0.73					
458		0.87							
468		0.87							
478		0.87							
489		0.83							
499		0.75		0.71			0.72		
509							0.88		
550		0.83		0.75					
560		0.82				0.88			
570		0.82							
580		0.84							
590		0.87				0.89			
600		0.88							
611		0.86							
621		0.87							
631		0.85			0.77				
641		0.76			0.87				
651		0.80							
661		0.83			0.90				
672		0.86							
692							0.80		
702							0.92		
722	0.74								
733		0.70							
743		0.80							
753		0.85							
763		0.89							
773		0.85							
783		0.79							
794							0.73		
804						0.89	0.95		
814						0.83			
824	0.77						0.81		
845							0.89		
855							0.94		
865							0.90		
916				0.80					
943		0.76		0.90			0.90		
953		0.83		0.83			0.92		
963		0.72		0.84			0.81		
993	0.71								
1013	0.90								
1023	0.92								
1084					0.71				
1094						0.79			
1124			0.82	0.84			0.91		
1134			0.84	0.87			0.94		
1145			0.84	0.84			0.94		
1155			0.82	0.83			0.83		
1175			0.83	0.74			0.79		
1185			0.82	0.78			0.79		
1205	0.82								
1215	0.75								
1225	0.72						0.70		
1235									
1276			0.72	0.76					
1286			0.77	0.78			0.91		
1296			0.81	0.74			0.92		
1306			0.82	0.71			0.87		
1316			0.82	0.71			0.87		
1326			0.81	0.70	0.71		0.83		
1336			0.80		0.71		0.80		
1346			0.77		0.74				
1538							0.92		
1548							0.91		
1558							0.89		
1568							0.88		
1578							0.89		
1588	0.91						0.88		
1599	0.92						0.88		
1609	0.92						0.86		
1619	0.90						0.85		
1629	0.89						0.84		
1639	0.86						0.82		
1649	0.74								
1689					0.79		-0.70		
1699					0.80		-0.76		
1710					0.82		-0.79		
1720					0.93		-0.82		
1740		0.75			0.95				
1750		0.72			0.95		-0.82		
1760		0.74			0.95		-0.84		
1770		0.74							

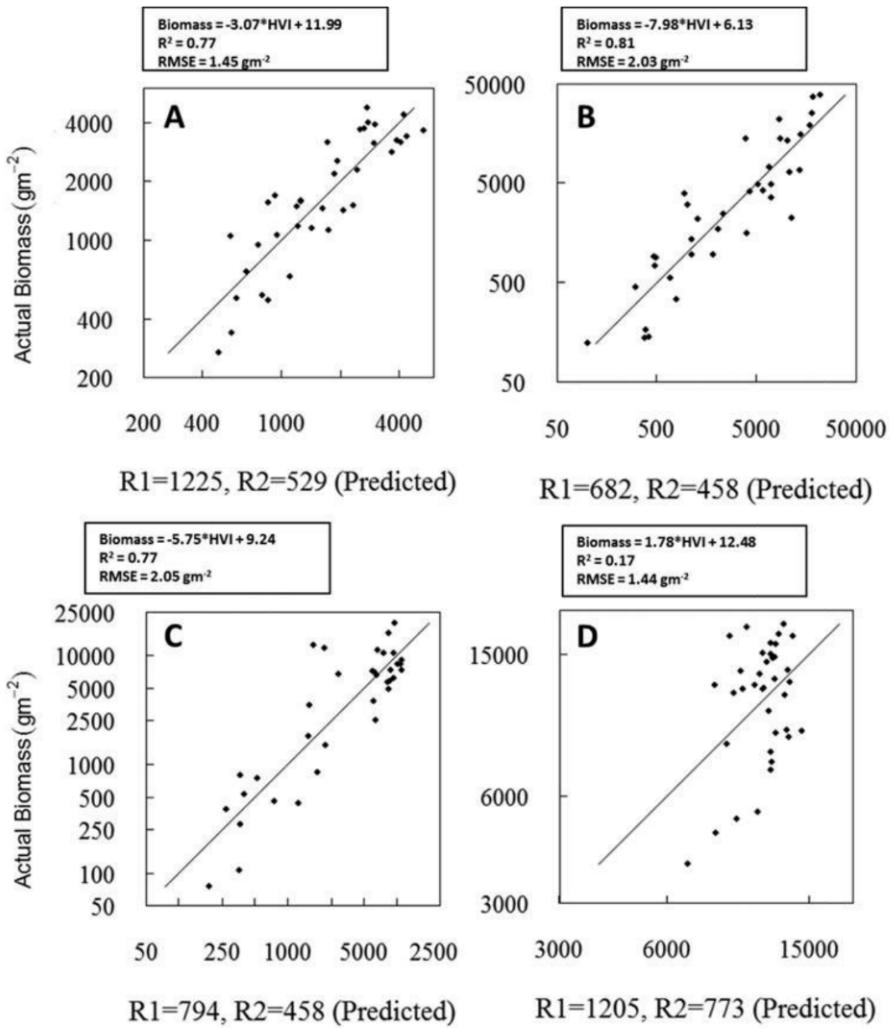


Figure 4. Scatterplots of observed biomass versus predicted biomass for the highest ranked first derivative transformed two-band HVI on the calibration dataset subset for (A) rice ( $N=37$ ), (B) alfalfa ( $N=37$ ), (C) cotton ( $N=35$ ), and (D) maize ( $N=37$ ). The diagonal line represents a 1:1 relationship. The band centroids used to derive the HBVIs are shown on the x-axes. Also shown is the prediction formula, explained variance, and root mean squared error (RMSE).

in later growth stages. With a larger sample size in the future, the analysis should be performed again for each crop at each major phenological phase to resolve this issue.

It is difficult to select one statistical method over another, but we can say that the two-band HVI and MB-HVI using SSM approaches yield significantly higher  $R^2$  and lower RMSE than the PCR approach. Rice and alfalfa show lower RMSE and modestly higher  $R^2$  using the two-band HVI approach, while cotton and maize perform better using FA. All three methods under-predicted biomass during the sprouting phase and this could be due to the low (high) absorption and reflectance of incoming radiation by the canopy (soil) (Huete, 1988).

The major advantage of FA and additive models in general are that errors are smaller than two-band HVI models due to the way in which variables are combined and the model-building process is simpler to implement and evaluate. The major disadvantage of this approach, as is the case with cotton, is that they typically include more HNBs, which can induce another type of error. With a larger sample size, the number of predictor variables could be increased and/or a stepwise regression could be performed to reduce the probability of eliminating collinear, but highly significant HNBs.

The major advantage and disadvantage of the two-band HVI approach is its flexibility. Many of the two-band HVIs that ranked high in the calibration dataset with a  $R^2 \leq 0.02$  from the top ranked HVI, often performed better on the validation subset. The two-band HVI for rice at 1225 and 539 nm, for alfalfa at 1205 and 438 nm, for cotton at 1023 and 428 nm, and for maize at 1205 and 794 nm, ranked high on the calibration subset and explained 78, 86, 84, and 22 percent of the variability in the validation subset, respectively. Other high correlations are seen at 1205 and 438 nm, as well as at 1629 nm across the visible for alfalfa, while combinations at 428 nm around 763 nm are also strong for cotton. It is recommended when using the two-band HVI approach that a cross-validation is performed for all two-band HVI combinations to minimize errors without the need for a calibration and val-

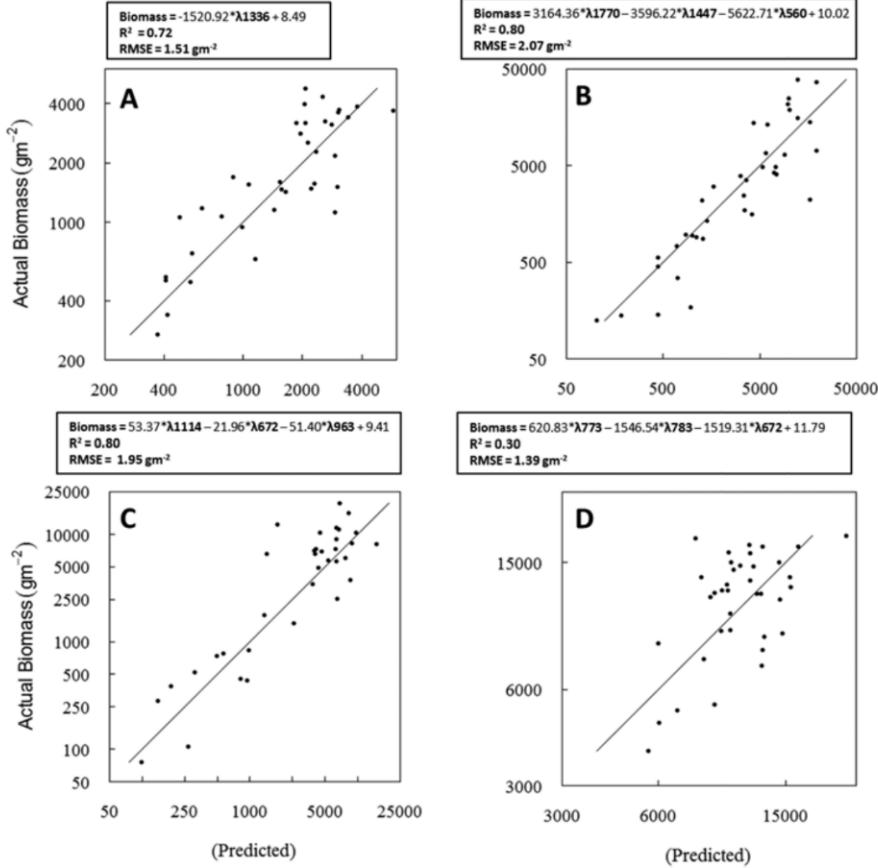


Figure 5. Scatterplots of observed biomass versus predicted biomass for ssm model using the validation subset for (A) rice, (B) alfalfa, (C) cotton, (D) and maize. The diagonal line represents a 1:1 relationship. Band centers in each prediction equation are preceded by  $\lambda$ . With the exception of cotton, which is untransformed, equations are built on first derivative transformed spectra.

dation subsets. This would be computationally intensive, but yield the most robust two-band HVI.

The Principal Component Regression approach is the poorest of the three methods and includes explanatory information from the largest number of HNBs in each component. Its lower performance in the validation dataset reflects the difficulty in choosing components in the final score. The strength of this method may therefore not be in prediction, but in data reduction. Future studies could use PCA as a filter to identify potentially important bands, before implementing SSMs, which initially have restrictions on the number of predictors. But, there would still be the issue in selecting the number of components to include and an additional issue of deciding a loading threshold in which to include HNBs.

Previous ground-based studies confirm many of the results of this study, as well highlight the potential of the approaches developed in this paper for space-borne hyperspectral missions. Nguyen and Lee (2006) use PCR analysis to show that rice growth is highly correlated to pigment absorption in the visible (524 to 534 nm) and red-edge (687 nm), and scatter in the NIR (760 to 1100 nm). Our study reveals in each model-building approach, however that NIR/SWIR bands just beyond 1100 nm (spectral limit of the previous study) are also highly significant. Bai et al. (2007) confirms similar correlations between hyperspectral regions and cotton biomass, with high negative correlations in the visible and relatively lower negative correlations around the green-peak (510 nm), high

positive correlations in the NIR, and negative correlations in the SWIR. In particular, models show high correlations with biomass in the red-edge (672 nm) and NIR (901 nm). Unlike this study, important SWIR1 (1466 nm) and SWIR2 (2096 nm) bands are identified as well. Short-wave channels are not included for cotton in this study, either because the SWIR channels tended to be more noisy or due to methodological differences of the two studies. The previous study selected single bands with the highest biomass variation coefficient correlation in each spectral region and then built single band or two-band HVI models around these bands, whereas in this study, several approaches are evaluated and bands are selected iteratively with minimal bias. Gnyp et al. (2014) took a similar methodological approach as this study for rice. Like this study, the first derivative transformed two-band HVI included visible green (526 nm), but explained a lower amount of variability on their validation subset ( $R^2 = 0.58$ ). Unlike this study, their best SSM model included six bands in the visible green, red, and NIR, yet only accounted for 72 percent of the variability in rice biomass. While our SSM model explained the same amount of variability using only one variable in the longer portion of the NIR. Thenkabail et al. (2000) used an SSM approach to show that the best biomass band for maize is at the red-edge (654 nm), followed by visible bands (410 and 496 nm), and finally NIR (954 nm), yielding an  $R^2$  of 0.78 with four predictors. The relatively strong (poor) performance of maize biomass in the SWIR (red-edge/NIR) in this study, points

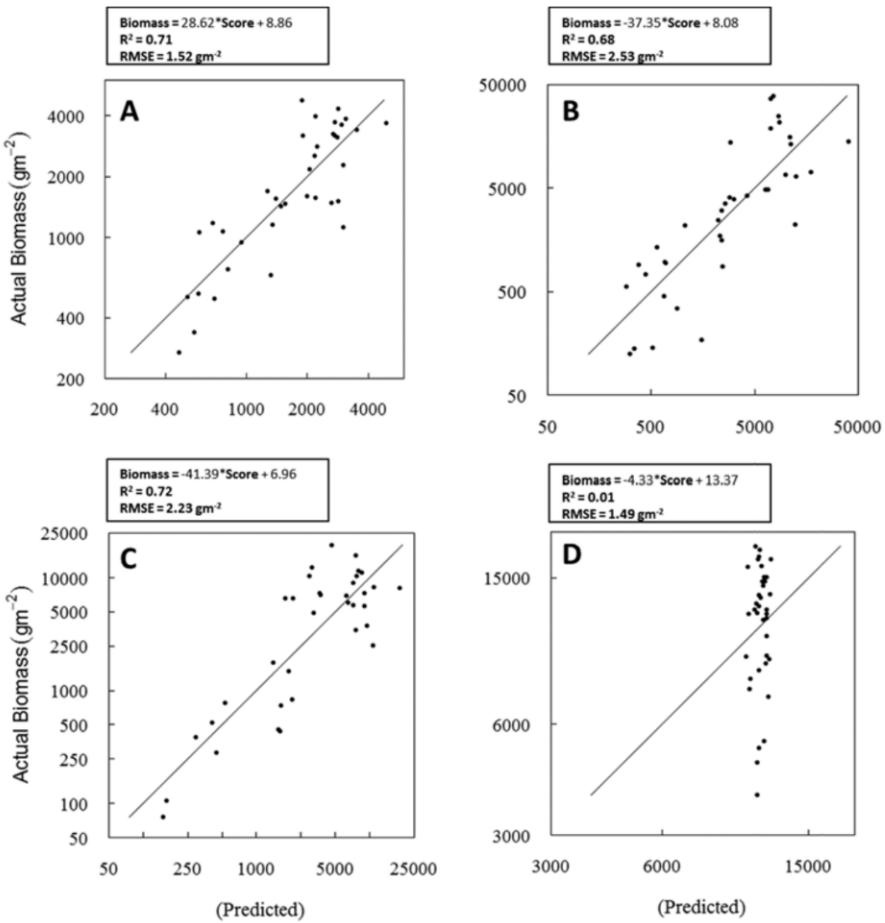


Figure 6. Scatterplots of observed biomass versus predicted biomass derived from PCA scores determined from PCA weights and validation subset for (A) rice, (B) alfalfa, (C) cotton, and (D) maize. The correlations are based on the first derivative transformed spectra for all crop types. Scores were derived from the first two components, except maize, which uses three.

to a potential source of error: the height at which the spectra were collected. The spectra were collected at 2.5 m, which at the beginning growth stage was sufficiently above the canopy. As the maize matured, however the sensor was just above the canopy or in some cases, when the maize unpredictably grew to over 3 m, within the canopy. In these cases, spectra most likely captured a single leaf exposed to sunlight or in shadow, as opposed to the entire canopy, significantly impacting the spectra, yet went undetected during data processing. Mounting the ASD on a cherry-picker could have avoided this, but was not available to us at the time. In a follow-up to this study, multiple hyperspectral and multi-spectral broadband satellite remote sensing data will be analyzed and combined to improve crop biomass estimates over larger areas. Since these sensors are well above the canopy, we expect significantly better correlations in the red-edge and NIR for maize, comparable to Thenkabail et al. (1994) who explained 80 percent of the biomass variability for maize using a combination of NIR, SWIR and thermal Landsat Thematic Mapper bands.

## Conclusions

The study used HBNs and HVIS to model aboveground fresh biomass for four leading world crops: alfalfa, cotton, maize,

and rice. Three distinct model types were investigated using untransformed, first derivative transformed, and second derivative transformed spectra: two-band HVIS, MB-HVIS using SSM and MB-HVIS using PCR. In most cases, one to three HNBs explained more than 70 percent of the biomass variability in the validation subset, while figures are higher in the calibration subset. Both the two-band HVI and MB-HVI (SSM) approaches using first derivative transformed spectra performed the best overall and are comparable. The two-band HVI approach, however, needs to be revisited, as many of the two-band HVIS that performed well in the calibration phase, performed as well, if not better than any other technique on the validation subset. Two-band HVIS tended to include visible green and NIR for alfalfa, visible blue and NIR for cotton, visible green and NIR and NIR and SWIR for rice, and NIR and SWIR for maize.

The need to design water-saving techniques in the Central Valley of California and other semi-arid irrigated agricultural regions of the world is increasing, as population growth, competing sources, and climate change increase the demand for global water supply. Crop biomass is used extensively in a predictive capacity to measure yield and to monitor crop water productivity. The next step in the analysis will involve comparing the technique using multiple broad-band and HNB spaceborne remote sensing platforms, as well as to refine broad-band

approaches to estimating the fraction of absorbed photosynthetically active radiation, a major component of evapotranspiration models. This again will provide improved statewide estimates of another important factor in crop water productivity.

### Acknowledgments

This work was primarily supported by the United States Geological Survey (USGS) Climate and Land Use Mission area's Geographic Analysis and Monitoring and Land Remote Sensing programs under the supervision of the Mendenhall Research Fellowship Program. Funding to support undergraduate/graduate students was granted through a cooperative initiative between the USGS and The National Association of Geoscience Teachers. First, we would like to thank Tony Chang, Jeff Peters, and Bobbi Jean Freeman, who worked long hours under adverse conditions to gather field data required for this project. Second, we would like to thank the agricultural extensionists (Mark Keeley, Chris Greer, Cayle Little, Bob Hutmacher, and Shannon Mueller) from the University of California at Davis (UCD) who put us in contact with growers or farm managers interested in our work and permitted us to collect data on UCD managed research farms (West Side and Shafter Research and Extension Centers). Third, we would like to thank the growers and farm managers (Chuck Mathews, Don Bransford, Jim Casey, Blake Harlan, Mark Boyd, Danny Kirshenmann, John Diener, Marty Roads, Joel Ackerknecht, Scott Schmidt, and Kurt Boeger) who gave us the opportunity to work in their fields and set aside their time to make arrangements and lend us their valuable experience. And to the grower Steve Mellow who welcomed us to a 4<sup>th</sup> of July celebration at his home when we were far from home. Last, we would like to thank Deborah Soltesz, Miguel Velasco, Larry Gaffney, and Lois Hales who dealt with ordering/repairing parts, made travel arrangements, and arranged other field work logistics.

### References

- Ali, M.H., and M.S.U. Talukder, 2008. Increasing water productivity in crop production - A synthesis. *Agricultural Water Management*, 95:1201–1213.
- Asner, G.P., 1998. Biophysical and biochemical sources of variability in canopy reflectance, *Remote Sensing of Environment*, 64:234–253.
- Bai, J., S. Li, K. Wang, X. Sui, B. Chen, and F. Wang, 2007. Estimating aboveground fresh biomass of different cotton canopy types with homogeneity models based on hyper spectrum parameters, *Agricultural Sciences in China*, 6:437–445.
- Baret, F., G. Guyot, and D.J. Major, 1989. Crop biomass evaluation using radiometric measurements, *Photogrammetria*, 43:241–256.
- Bastiaanssen, W.G., D.J. Molden, and I.W. Makin, 2000. Remote sensing for irrigated agriculture: Examples from research and possible applications, *Agricultural Water Management*, 46:137–155.
- Box, E.O., B.N. Holben, and V. Kalb, 1989. Accuracy of the AVHRR vegetation index as a predictor of biomass, primary productivity and net CO<sub>2</sub> flux, *Vegetation*, 80:71–89.
- Casa, R., and H.G. Jones, 2004. Retrieval of crop canopy properties: A comparison between model inversion from hyperspectral data and image classification, *International Journal of Remote Sensing*, 25:1119–1130.
- CDFA, 2013. *California Agricultural Statistics Review 2012-2013*, California Department of Food and Agriculture, Sacramento, California.
- Champagne, C.M., K. Staenz, A. Bannari, H. McNairn, and J.-C. De guise, 2003. Validation of a hyperspectral curve-fitting model for the estimation of plant water content of agricultural canopies, *Remote Sensing of Environment*, 87:148–160.
- Demetriades-Shah, T.H., M.D. Steven, and J.A. Clark, 1990. High resolution derivative spectra in remote sensing, *Remote Sensing of Environment*, 33:55–64.
- Doraishwamy, P.C., S. Moulin, P.W. Cook, and A. Stern, 2003. Crop yield assessment from remote sensing, *Photogrammetric Engineering & Remote Sensing*, 69(7):665–674.
- Elvidge, C.D., and Z. Chen, 1995. Comparison of broad-band and narrow-band red and near-infrared vegetation indices, *Remote Sensing of Environment*, 54:38–48.
- Faunt, C.C., 2009. *Groundwater Availability of the Central Valley Aquifer*, U.S. Geological Survey Professional Paper No. 1766, U.S. Geological Survey, Sacramento, California.
- Gitelson, A.A., G.P. Keydan, and M.N. Merzlyak, 2006. Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves, *Geophysical Research Letters*, 33, DOI 10.1029.
- Gnyp, M.L., Y. Miao, F. Yuan, S.L. Ustin, K. Yu, Y. Yao, S. Huang, and G. Bareth, 2014. Hyperspectral canopy sensing of paddy rice aboveground biomass at different growth stages. *Field Crops Research*, 155:42–55.
- Haboudane, D., J.R. Miller, E. Pattey, P.J. Zarco-Tejada, and I.B. Strachan, I.B., 2004. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture, *Remote Sensing of Environment*, 90:337–352.
- Hair, Jr., J.F., R.E. Anderson, R.E.L. Tatham, and W.C. Black, 1998. *Multivariate Data Analysis*, Fifth edition, Prentice Hall, Upper Saddle River, New Jersey.
- Hall, F.G., K.F. Huemmrich, and S.N. Goward, 1990. Use of narrow-band spectra to estimate the fraction of absorbed photosynthetically active radiation, *Remote Sensing of Environment*, 32:47–54.
- Hay, R.K.M., 1995. Harvest index: A review of its use in plant breeding and crop physiology, *Annals of Applied Biology*, 126:197–216.
- Horler, D.N.H., M. Dockray, and J. Barber, 1983. The red edge of plant leaf reflectance, *International Journal of Remote Sensing*, 4:273–288.
- Huete, A., 1988. A soil-adjusted vegetation index (SAVI), *Remote Sensing of Environment*, 25:295–309.
- Hughes, G., 1968. On the mean accuracy of statistical pattern recognizers, *IEEE Transactions on Information Theory*, 14:55–63.
- Lu, D., 2006. The potential and challenge of remote sensing based biomass estimation, *International Journal of Remote Sensing*, 27:1297–1328.
- MacArthur, A., C.J. MacLellan, and T. Malthus, 2012. The fields of view and directional response functions of two field spectroradiometers, *IEEE Transactions on Geoscience and Remote Sensing*, 50:3892–3907.
- Moran, M.S., P.J. Pinter, Jr., B.E. Clothier, and S.G. Allen, 1989. Effect of water stress on the canopy architecture and spectral indices of irrigated alfalfa, *Remote Sensing of Environment*, 29:251–261.
- Nguyen, H.T., and B.-W. Lee, 2006. Assessment of rice leaf growth and nitrogen status by hyperspectral canopy reflectance and partial least square regression, *European Journal of Agronomy*, 24:349–356.
- Perry, E.M., and D.A. Roberts, 2008. Sensitivity of narrow-band and broad-band indices for assessing nitrogen availability and water stress in an annual crop, *Agronomy Journal*, 100:1211.
- Pinter, P.J., J.L. Hatfield, J.S. Schepers, E.M. Barnes, M.S. Moran, C.S.T. Daughtry, and D.R. Upchurch, 2003. Remote sensing for crop management, *Photogrammetric Engineering & Remote Sensing*, 69(7):647–664.
- Rodarmel, C., and J. Shan, 2002. Principal component analysis for hyperspectral image classification, *Surveying and Land Information Systems*, 62:115–123.
- Serrano, L., J. Peñuelas, and S.L. Ustin, S.L., 2002. Remote sensing of nitrogen and lignin in Mediterranean vegetation from AVIRIS data: Decomposing biochemical from structural signals, *Remote Sensing of Environment*, 81:355–364.
- Sunding, D., 2000. The price of water: Market-based strategies are needed to cope with scarcity, *California Agriculture*, 54:56–63.
- Thenkabail, P.S., E.A. Enclona, M.S. Ashton, and B. Van Der Meer,

2004. Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications, *Remote Sensing of Environment*, 91:354–376.
- Thenkabail, P.S., I. Mariotto, M.K. Gumma, E.M. Middleton, D.R. Landis, and K.F. Huemmrich, 2013. Selection of hyperspectral narrowbands (HNBs) and composition of hyperspectral two-band vegetation indices (HVIIs) for biophysical characterization and discrimination of crop types using field reflectance and Hyperion/EO-1 data, *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*, 6:427–439.
- Thenkabail, P.S., R.B. Smith, and E. De Pauw, 2000. Hyperspectral vegetation indices and their relationships with agricultural crop characteristics, *Remote Sensing of Environment*, 71:158–182.
- Thenkabail, P.S., R.B. Smith, and E. De Pauw, 2002. Evaluation of narrowband and broadband vegetation indices for determining optimal hyperspectral wavebands for agricultural crop characterization, *Photogrammetric Engineering & Remote Sensing*, 68(5):607–621.
- Thenkabail, P.S., A.D. Ward, and J.G. Lyon, 1994. Landsat-5 Thematic Mapper models of soybean and corn crop characteristics, *International Journal of Remote Sensing*, 15:49–61.
- Tsai, F., and W. Philpot, 1998. Derivative analysis of hyperspectral data, *Remote Sensing of Environment*, 66:41–51.
- Umbach, K.W., 1997. *A Statistical Tour of California's Great Central Valley*, California Research Bureau, California State Library, Sacramento, California.
- USDA, 2009. *2007 Census of Agriculture*, U.S. Department of Agriculture, Washington D.C.
- Ustin, S.L., D.A. Roberts, J.A. Gamon, G.P. Asner, and R.O. Green, 2004. Using imaging spectroscopy to study ecosystem processes and properties, *BioScience*, 54:523–533.
- Ustin, S.L., A.A. Gitelson, S. Jacquemoud, M. Schaepman, G.P. Asner, J.A. Gamon, and P. Zarco-Tejada, 2009. Retrieval of foliar information about plant pigment systems from high resolution spectroscopy, *Remote Sensing of Environment* 113(Supplement 1):S67–S77.
- Zwart, S.J., and W.G.M. Bastiaanssen, 2004. Review of measured crop water productivity values for irrigated wheat, rice, cotton and maize, *Agricultural Water Management*, 69:115–133.