

Analysis of hyperspectral reflectance data for monitoring growth and development of lesquerella

K.R. Thorp^{a,*}, D.A. Dierig^b, A.N. French^a, D.J. Hunsaker^a

^a USDA-ARS, U.S. Arid Land Agricultural Research Center, 21881 N Cardon Ln, Maricopa, AZ 85138, United States

^b USDA-ARS, National Center for Genetic Resources Preservation, 1111 S Mason St, Fort Collins, CO 80521, United States

ARTICLE INFO

Article history:

Received 3 June 2010

Received in revised form

29 September 2010

Accepted 4 October 2010

Available online 21 December 2010

Keywords:

Remote sensing

Hyperspectral

Spectroradiometer

Partial least squares regression

Biomass

Flower

Silique

Management

Arizona

Agriculture

Oilseed

ABSTRACT

Seed oil from lesquerella (*Lesquerella fendleri* (Gray) Wats.) is currently being developed as a biorenewable petroleum substitute, but several issues related to crop management and breeding must be resolved before the crop will be commercially viable. Due particularly to the prominent yellow flowers exhibited by lesquerella canopies, remote sensing may be a useful tool for monitoring and managing the crop. In this study, we used a hand-held spectroradiometer to measure spectral reflectance over lesquerella canopies in 512 narrow wavebands from 268 to 1095 nm over two growing seasons at Maricopa, Arizona. Biomass samples were also regularly collected and processed to obtain aboveground dry weight, flower counts, and silique counts. Partial least squares regression was used to develop predictive models for estimating the three lesquerella biophysical variables from canopy spectral reflectance. For model fitting and model testing, the root mean squared prediction errors between measured and modeled aboveground dry weight, flower counts, and silique counts were 2.1 and 2.3 Mg ha⁻¹, 251 and 304 flowers, and 1018 and 1019 siliques, respectively. Analysis of partial least squares regression coefficients and loadings highlighted the most sensitive spectral wavebands for estimating each biophysical variable. For example, the flower count model heavily emphasized the reflectance of yellow light at 583 nm, and contrasted that with reflectance in the blue (483 nm) and at the red edge (721 nm). Because of the indeterminate nature of lesquerella flowering patterns, remote sensing methods that monitor flowering progression may aid management decisions related to the timing of irrigations, desiccant application, and crop harvest.

Published by Elsevier B.V.

1. Introduction

Lesquerella (*Lesquerella fendleri* (Gray) Wats.) has been studied as a potential source of industrial seed oil since 1960 (Smith et al., 1961). High concentrations of hydroxylated fatty acids can be found in the seeds of the lesquerella plant, and these compounds are useful as an additive to improve the lubricity of petroleum-based diesel fuels (Geller and Goodrum, 2004; Moser et al., 2008). Due to environmental concerns about air quality, governmental regulations currently limit the sulfur content of commercial petroleum diesel fuels. Ultra-low sulfur diesel fuels have been developed to address this concern; however, the decreased lubricity of these fuels can reduce the lifespan of diesel engines and their fuel injection systems. Seed oil from lesquerella plants may provide a renewable and environmentally sensitive solution to this problem. Hydroxylated fatty acids from lesquerella seed also has use as a petroleum substitute in the production of many other products, including greases, lubricants, cosmetics, paints, inks, and coatings.

Development of robust cultivars and reliable agronomic practices are required before lesquerella can become a viable commercial crop. Many field investigations have been conducted to address crop management issues, such as optimum sowing date and density (Garcia et al., 2007; Nelson et al., 1996), fertilization requirements (Nelson et al., 1996, 1999), irrigation water management (Hunsaker et al., 1998; Puppala et al., 2005), and herbicide tolerance (Roseberg, 1996). To define regions for optimum production, Dierig et al. (2006) evaluated crop growth, development, and yield at four Arizona sites having different elevations and temperature regimes. Lesquerella is native to the southwestern United States and northern Mexico and is thus naturally adapted to arid environments. Efforts are currently focused on domesticating and commercializing the crop for production as a winter annual in this region.

Remote sensing has been widely investigated as a tool to characterize various aspects of crop production, including crop species identification (Jakubauskas et al., 2002), crop yield (Idso et al., 1977), crop water stress (Jackson et al., 1981), nitrogen stress (Blackmer et al., 1996), evapotranspiration (Hunsaker et al., 2005), and plant stand density (Thorp et al., 2008). More recently, remote sensing has been proposed as a viable tool to aid breeding

* Corresponding author. Tel.: +1 520 316 6375; fax: +1 520 316 6330.

E-mail address: kelly.thorp@ars.usda.gov (K.R. Thorp).

Table 1
Summary of the 2007–2008 and 2008–2009 lesquerella experiments.

	Planting 1	Planting 2	Planting 3
2007–2008 Experiment			
Planting date	9/28/2007	2/15/2008	3/10/2008
Replications	3	3	3
Sample areas per plot	20	12	0
Biomass samples per plot	16	7	0
RS data collection dates	19	11	0
2008–2009 Experiment			
Planting date	10/6/2008	1/8/2009	2/6/2009
Replications	3	3	3
Sample areas per plot	12	8	8
Biomass samples per plot	8	4	3
RS data collection dates	11	7	5

efforts through its implementation on high-throughput phenotyping platforms (Montes et al., 2007). The latter application may be particularly important for speeding the development of new crops such as lesquerella.

Very few reports of remote sensing studies over lesquerella canopies exist in literature. However, the indeterminate and vibrant yellow flowering patterns of lesquerella canopies make the crop spectrally intriguing. For example, Adamsen et al. (2000) used a digital camera to photograph lesquerella canopies and analyzed the images to estimate flower counts. Additionally, remote sensing studies over oilseed rape (*Brassica napus* L.) canopies, which also produce vibrant yellow flowers at anthesis, have shown that spectral indices can be sensitive to flowering time (Mogensen et al., 1996). Since the yield of many agricultural crops is sensitive to plant stress at anthesis and during grain fill, a tool to identify the temporal flowering patterns of lesquerella canopies may aid crop management decisions. Similarly, a tool for rapid estimation of flower counts and other biophysical properties may facilitate breeding efforts for lesquerella. To investigate the use of remote sensing as a tool for lesquerella management and breeding, our objectives were to (1) collect hyperspectral remote sensing data of lesquerella canopies in a field setting with different planting dates and (2) use partial least squares regression (PLSR) to relate hyperspectral information to field measurements of several biophysical properties of the lesquerella canopy, including total aboveground dry weight, flower count, and silique count. PLSR methods have been used in previous field studies to relate hyperspectral canopy reflectance data to leaf nitrogen concentration in cotton (*Gossypium hirsutum* L.) (Fridgen and Varco, 2004), plant nitrogen concentration in rice (*Oryza sativa* L.) (Bajwa, 2006), and yield in citrus (*Citrus unshiu* Marc.) (Ye et al., 2009).

2. Materials and methods

2.1. Field experiments

Lesquerella was grown at the University of Arizona's Maricopa Agricultural Center (MAC) near Maricopa, Arizona (33.067547° N, 111.97146° W) over the winters of 2007–2008 and 2008–2009. The soil type at the site was a Casa Grande sandy loam, classified as fine-loamy, mixed, hyperthermic, Typic Natrargids. In both growing seasons, the field layout consisted of nine experimental plots, each 20 × 180 m and hydrologically isolated with border dikes. Three planting date treatments were replicated three times over the nine plots (Table 1). In the 2007–2008 experiment, the first and second planting dates were September 28 and February 15, respectively. The third treatment was planted in March, but poor stand density prevented any useful data from being collected from this treatment. In the 2008–2009 experiment, planting dates were October 6, January 8, and February 6. All plots were broadcast

planted at a rate of 12 kg ha⁻¹ using a Brillion planter with a roller ring. Plots were flood irrigated by siphoning water from a canal along the southern edge of the field. After crop emergence, multiple locations within each plot were randomly selected and flagged for biomass sampling, flower counting, silique counting, and recurrent remote sensing data collection. The marked areas were each 0.125 m². In 2007–2008, 20 sample areas were marked for the first planting while 12 areas were marked for the second planting. In 2008–2009, the total number of marked areas was twelve, eight, and eight for the first, second, and third plantings, respectively (Table 1).

2.2. Field measurements

Ground-based radiometric measurements were collected at each of the 0.125 m² sampling locations from emergence until biomass was destructively sampled at that location. Additional radiometric measurements (approximately 24 per plot) were collected while walking along a 180 m linear transect on the western edge of each plot. Spectral data collection occurred on a weekly basis during the 2007–2008 experiment and at a two-week interval during the 2008–2009 experiment. The total number of remote sensing data collection dates for each treatment is given in Table 1.

A hand-held field spectroradiometer (GER 1500, Spectra Vista Corp., Poughkeepsie, New York) was used to collect uncalibrated digital numbers related to light radiance from the crop canopy. The instrument collected information in 512 narrow wavebands from 268 to 1095 nm with bandwidth ranging from 1.5 to 2.1 nm. The instrument was equipped with an 18° field-of-view fiber optic. A wand constructed from PVC tubing was used to position the fiber optic at a nadir view angle over each target. For scans over the biomass sampling areas, the fiber optic was positioned approximately 1.0 m above the soil surface. For scans along the linear transects, data was collected approximately 2.0 m above the soil surface. Radiometric data collection consistently occurred in the morning around the time of a 57° solar zenith angle. Frequent radiometric observations of a calibrated, 0.6 m², 99% Spectralon panel (Labsphere, Inc., North Sutton, New Hampshire) were used to characterize solar irradiance throughout the data collection period. Canopy reflectance factors in each waveband were computed as the ratio of the canopy radiance over the corresponding time-interpolated value for solar irradiance. Reflectance factors from three radiometric measurements over each biomass sampling area were averaged to estimate the spectral reflectance of the sample area on each measurement date. Further data analysis was based on the spectra from 356 to 945 nm only, since the instrument's signal to noise ratio was highest at these wavelengths.

Biomass was destructively sampled at one of the 0.125 m² sampling locations in each plot on a weekly basis during the 2007–2008 experiment and every two weeks during the 2008–2009 experiment. In 2008, the first planting date treatment was sampled 16 times per plot from January 15 through May 21, and the second planting date treatment was sampled seven times per plot from May 1 to June 19. In 2009, the first treatment was sampled eight times per plot from February 3 to May 12, and the second treatment was sampled four times per plot from April 30 to June 11. The third treatment in 2009 was sampled three times per plot from May 13 to June 12. Typically, we overestimated the number of sampling areas needed to document lesquerella growth and development over the growing season, and some of the predefined sample areas remained unsampled at crop maturity (Table 1). To collect the samples, square frames of white PVC tubing were constructed to delineate each 0.125 m² sample area. Samples were typically collected in the early morning hours. Plant material was then immediately processed in the laboratory to obtain aboveground dry weight, flower count,

silique count, and several other biophysical properties of the lesquerella canopy.

2.3. Statistical analysis

PLSR was used to assess the relationship between the measured crop biophysical properties and the canopy spectral reflectance data. PLSR is an extension of the standard multiple linear regression (MLR) procedure, the latter of which is defined by:

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon \quad (1)$$

where \mathbf{Y} is an $n \times 1$ vector of responses (crop biophysical measurements), \mathbf{X} is an n -observation by p -variable matrix of predictors (hyperspectral reflectance measurements in p wavebands), β is a $p \times 1$ vector of regression coefficients, and ε is an $n \times 1$ vector of error terms with each element following an independent, normal distribution. The ordinary least squares estimate of β is given by:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad (2)$$

The MLR approach is commonly used to develop linear statistical models for predicting \mathbf{Y} from \mathbf{X} . However, when the number of predictor variables (p) is greater than the number of observations (n) or when multicollinearity exists among the predictor variables, the matrix $\mathbf{X}'\mathbf{X}$ is often singular and unable to be inverted, as required by Eq. (2). Since these conditions are certainly likely for the highly correlated, narrow-band reflectance data from the field spectroradiometer used in these experiments, an alternative approach such as PLSR was required.

The first step in PLSR is to center \mathbf{X} and \mathbf{Y} by subtracting from each data value its respective column mean, giving \mathbf{X}_0 and \mathbf{Y}_0 . PLSR then addresses the singularity problem by decomposing \mathbf{X}_0 into a set of A orthogonal scores (which are weighted sums of the predictor variables in \mathbf{X}_0) with corresponding scores for \mathbf{Y}_0 and loadings for \mathbf{X}_0 and \mathbf{Y}_0 :

- X-scores: $\mathbf{T} = [\mathbf{t}_1, \dots, \mathbf{t}_A]$
- X-loadings: $\mathbf{P} = [\mathbf{p}_1, \dots, \mathbf{p}_A]$
- X-weights: $\mathbf{R} = [\mathbf{r}_1, \dots, \mathbf{r}_A]$
- Y-scores: $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_A]$
- Y-loadings: $\mathbf{Q} = [\mathbf{q}_1, \dots, \mathbf{q}_A]$

The X-scores, \mathbf{t}_a , are determined successively by maximizing their covariance with the corresponding Y-scores, \mathbf{u}_a , for $a = \{1, 2, \dots, A\}$. This is accomplished by first computing the singular value decomposition of the cross product matrix $\mathbf{S}_0 = \mathbf{X}'_0\mathbf{Y}_0$, thereby incorporating information on the variance of both \mathbf{X}_0 and \mathbf{Y}_0 and the covariance between them. Scores and loadings are then successively computed according to:

$$\mathbf{t}_a = \mathbf{X}_0\mathbf{r}_a, \quad a = 1, 2, \dots, A \quad (3)$$

$$\mathbf{p}_a = \mathbf{X}'_0\mathbf{t}_a, \quad a = 1, 2, \dots, A \quad (4)$$

$$\mathbf{q}_a = \mathbf{Y}'_0\mathbf{t}_a, \quad a = 1, 2, \dots, A \quad (5)$$

$$\mathbf{u}_a = \mathbf{Y}_0\mathbf{q}_a, \quad a = 1, 2, \dots, A \quad (6)$$

where \mathbf{r}_a is the first left singular vector of successively deflated cross product matrices, \mathbf{S}_{a-1} . Each succeeding \mathbf{S}_a is determined from its predecessor \mathbf{S}_{a-1} by projecting \mathbf{S}_{a-1} onto the column space of the X-loadings, \mathbf{P}_a , thereby removing the information previously extracted. The solution to Eqs. (3)–(6) is constrained by several conditions, including:

- maximization of covariance: $\mathbf{u}'_a\mathbf{t}_a = \text{maximum}$
- normalization of scores \mathbf{t}_a : $\mathbf{t}'_a\mathbf{t}_a = 1$
- orthogonality of all \mathbf{t} scores: $\mathbf{t}'_b\mathbf{t}_a = 0$ for $a > b$

It can be shown that the partial least squares regression coefficients, $\hat{\beta}_{\text{PLS}}$, are estimated as:

$$\hat{\beta}_{\text{PLS}} = \mathbf{R}\mathbf{Q}' \quad (7)$$

where \mathbf{R} and \mathbf{Q} are the X-weights and Y-loadings as defined previously. In an equation similar to (1), stable predictions of \mathbf{Y}_0 can then be obtained in spite of the multicollinearity in \mathbf{X}_0 according to:

$$\mathbf{Y}_0 = \mathbf{X}_0\hat{\beta}_{\text{PLS}} + \varepsilon \quad (8)$$

To validate or apply the model for new datasets, one simply subtracts from the validation dataset the column means of the calibration dataset used to fit the model.

There are many variations of PLSR methods and their respective computational implementations. The preceding discussion describes the PLSR method of de Jong (1993), commonly known as SIMPLS. We have implemented this method using the 'pls' package (Mevik and Wehrens, 2007) within the R Project for Statistical Computing (<http://www.r-project.org/>). The 'pls' package includes several useful features for fitting and validating PLSR models. For example, choosing an appropriate number of factors (A from above) is an important decision during modeling fitting. Cross validation is typically used for this purpose, and the 'pls' package is fully equipped to utilize this procedure during model fitting. We used leave-one-out cross validation to select the appropriate number of factors to be included in each PLSR model.

Three PLSR models were developed in this study, one each for predicting aboveground dry weight, flower count, and silique count. Since the lesquerella crop was sampled more frequently in 2007–2008, we used data from this season to build the models, and leave-one-out cross validation on the 2007–2008 dataset was implemented to select the appropriate number of factors. Data from the 2008–2009 season were then used to independently test each PLSR model. To build and test the PLSR models, we only included the spectral information from the measurement date that immediately preceded biomass sampling at each location. Originally, we attempted to build separate PLSR models for each planting date treatment. However, this lowered the sample number used for model building, and the resulting predictions from these models were often poor. Therefore, the three PLSR models in this study were built from a combined dataset from all the treatments in the 2007–2008 growing season. Data were included in the analysis only if the biophysical measurement was greater than zero. Mainly, this prevented early season changes in reflectance from affecting the development of PLSR models for reproductive characteristics, such as flower count and silique count. Interpretation of each PLSR model was accomplished by analysis of the X-loadings, \mathbf{p}_a , for each factor in the model and by analysis of the estimates of the regression coefficients, $\hat{\beta}_{\text{PLS}}$.

2.4. Application

The spectral reflectance data collected along the 180 m transects were used to demonstrate the ability of the PLSR models to temporally monitor lesquerella growth and development for the three planting dates in each experiment. Each of the evaluated PLSR models was applied to estimate aboveground dry weight, flower count, and silique count at each location along the transect on each measurement date. Estimates were then averaged according to the measurement date and planting date treatment to generate time-series plots of each crop biophysical variable. These plots demonstrated the ability of the remote sensing approach to temporally monitor lesquerella growth and development over each growing season.

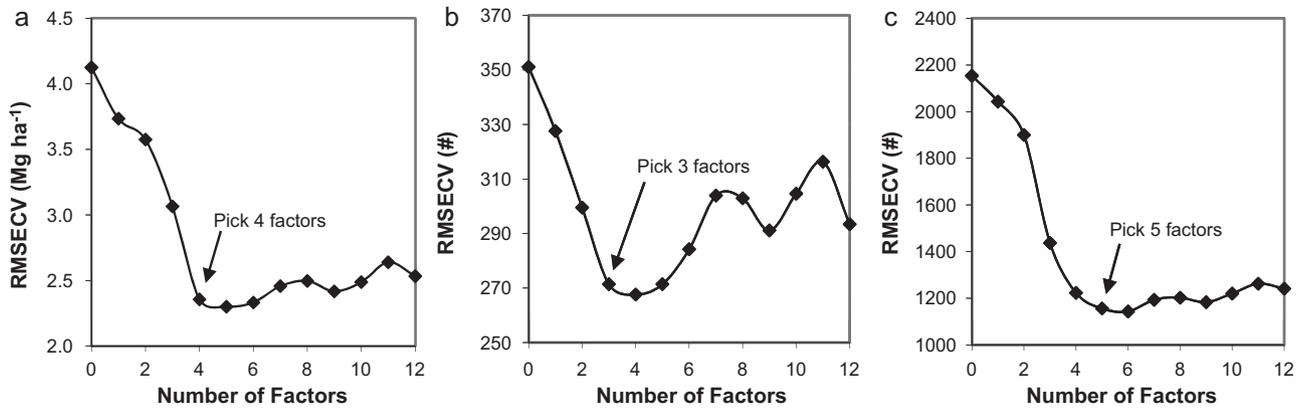


Fig. 1. Scree plots showing the root mean squared error of cross validation (RMSECV) when using partial least squares regression to predict (a) aboveground dry weight, (b) flower count, and (c) silique count from the spectral reflectance of the lesquerella canopy in the 2007–2008 growing season.

3. Results and discussion

3.1. Model fitting

Scree plots demonstrated the performance of each PLSR model as additional PLS factors were included in the fit (Fig. 1). These plots were used to select the number of factors (A) needed to build the most effective PLSR model for each biophysical property. Selection of the appropriate number of factors is somewhat subjective, which is perhaps a weakness of PLSR methods. We have made our selections roughly at the first local minimum of each scree plot, where the addition of another factor did not substantially improve model predictions of the biophysical property. Using this criterion, PLSR models for aboveground dry weight, flower count, and silique count were developed using the first four, three, and five PLS factors, respectively.

Results of the model fitting procedure using the 2007–2008 experimental data demonstrated a moderate relationship between the crop canopy spectral reflectance data and the three crop biophysical properties (Fig. 2). The fitted PLSR model for aboveground dry weight performed with a 2.1 Mg ha^{-1} root mean squared error of prediction (RMSEP). Flower counts were estimated with a RMSEP of 251 flowers, and silique counts were estimated with a RMSEP of 1018 siliques. Modeling results demonstrated moderate trends in the relationship between canopy spectral reflectance and the crop biophysical variables, but the method may not be suitable if very precise estimates of these quantities are needed. Some of the error between modeled and measured values is likely due to the time difference between remote sensing measurements and biomass samples, which could be up to several days in length. We

took great care to coordinate the remote sensing measurements and biomass samples spatially, but improved coordination in the timing of remote sensing measurements and biomass samples is warranted for future studies.

3.2. Model testing

Independent model testing using the 2008–2009 experimental data gave similar results regarding the ability of the PLSR models to predict aboveground dry weight, flower counts, and silique counts from canopy spectral reflectance (Fig. 3). For aboveground dry weight, the RMSEP between modeled and measured values was 2.3 Mg ha^{-1} . Also, flower counts and silique counts were estimated with RMSEPs of 304 flowers and 1019 siliques, respectively. These error values were all slightly higher than that obtained during the model fitting phase. However, they were quite similar in magnitude, indicating that the models were reasonably predicting the three biophysical variables for this independent dataset.

3.3. Model interpretation

According to Eq. (4), the set of X-loadings, \mathbf{P} , provide the vectors that relate the X-scores, \mathbf{T} , back to the original mean-centered reflectance data, \mathbf{X}_0 . As such, the X-loadings (Fig. 4) can be used to understand the relative contribution of each spectral waveband to the corresponding X-scores. For aboveground dry weight, the X-loadings demonstrated that reflectance in the near-infrared (NIR) region from 750 to 900 nm contributed quite strongly to the X-scores of the first factor, while reflectance in the green region around 550 nm provided an additional contribution. Positive

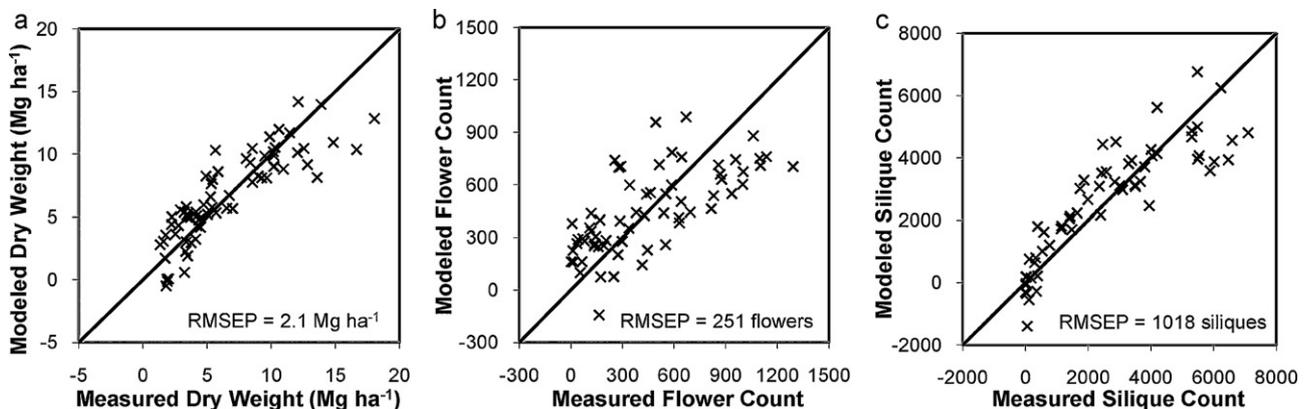


Fig. 2. Modeled versus measured (a) aboveground dry weight, (b) flower count, and (c) silique count when fitting partial least squares regression models to relate each biophysical variable to canopy spectral reflectance for the 2007–2008 lesquerella experiment.

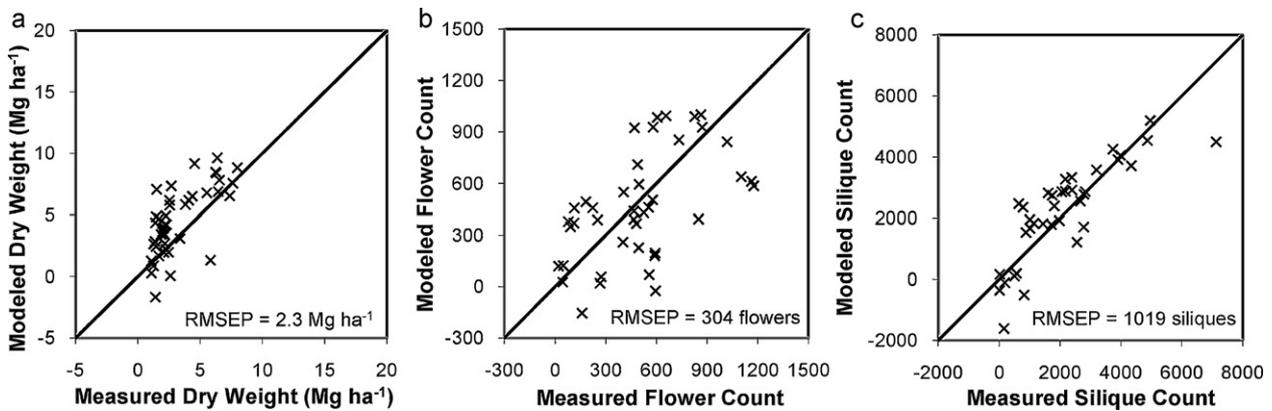


Fig. 3. Modeled versus measured (a) aboveground dry weight, (b) flower count, and (c) silique count when testing the partial least squares regression models, developed using the 2007–2008 experimental data, against the data collected during the 2008–2009 lesquerella experiment.

X-loading values in these regions indicated that the X-scores of the first factor should increase with increasing reflectance of green and NIR radiation. It is widely known that NIR radiation is readily scattered from healthy plant leaves and that vegetation reflects green light more strongly than other visible light (Knippling, 1970). Thus, the X-scores from the first factor were directly related to the increases in green and NIR reflectance that would be expected with increases in biomass as the season progresses. The X-loading values for the second factor were strongly negative in the red region from 600 to 700 nm, which relates to absorption of red light by chlorophyll in plant leaves. Negative X-loading values in the red region indicated that the X-scores of the second factor should increase as red reflectance decreases with increasing biomass. Other factors of the PLSR model for aboveground dry weight contrasted the reflectance in several regions where the curvature of the canopy spectra was quite pronounced, such as the green peak at 550 nm and the wavebands at 700 and 750 nm which straddle the red edge (Horler et al., 1983). The purpose of these factors was likely to identify subtle reflectance differences in these regions of the spectrum where vegetative reflectance changes rapidly with wavelength.

Similar to the PLSR model for dry weight, the first factor of the PLSR model for flower count focused on the positive correlation of green and NIR reflectance with increasing vegetative biomass. Thus, the primary purpose of the first factor was to characterize the general vegetative spectral signal in the data. The second factor of the flower count model more directly identified the aspects of the spectral signal that were related to flowering. With positive X-loading values that peaked around 600 nm, the X-scores of the second factor should increase with increasing reflectance of visible yellow light from the canopy, which is directly related to

the increase in the number of yellow flowers present. The third factor of the PLSR model for flower counts is similar to the second factor in the model for dry weight, which characterizes the decrease in red reflectance due to increases in chlorophyll. Since lesquerella exhibits indeterminate flowering patterns, increases in plant biomass are expected to occur concurrently with increases in flower number. Thus, the third component was likely characterizing chlorophyll in the growing canopy rather than directly monitoring the progression of flowering. The PLSR models know nothing of the physiology of lesquerella growth and only aim to relate variability in spectral responses to variability in the measured crop biophysical variables.

As opposed to dry weight and flower count, there is no direct spectral explanation for the relationship between silique numbers and canopy spectral reflectance. Likely, the remarkable performance of the PLSR model for silique count (Figs. 2 and 3) is related to other aspects of the crop canopy that can be directly correlated to crop yield, such as dry weight and flower numbers. Similar to the PLSR model for flower counts, the first factor of the PLSR model for silique counts also focuses on the positive correlation of green and NIR reflectance with increasing vegetative biomass, while the second factor characterizes the aspects of the spectral signal related to increases in yellow reflectance due to flowering in the canopy. These first two factors are likely characterizing spectral responses that are correlated with silique counts but not related directly. The third factor in the PLSR model for silique counts highlights reduction in red reflectance, similar to the second factor in the dry weight model and the third factor in the flower count model. However, the magnitude of the third factor X-loadings are somewhat smaller for the silique count model. This may be related

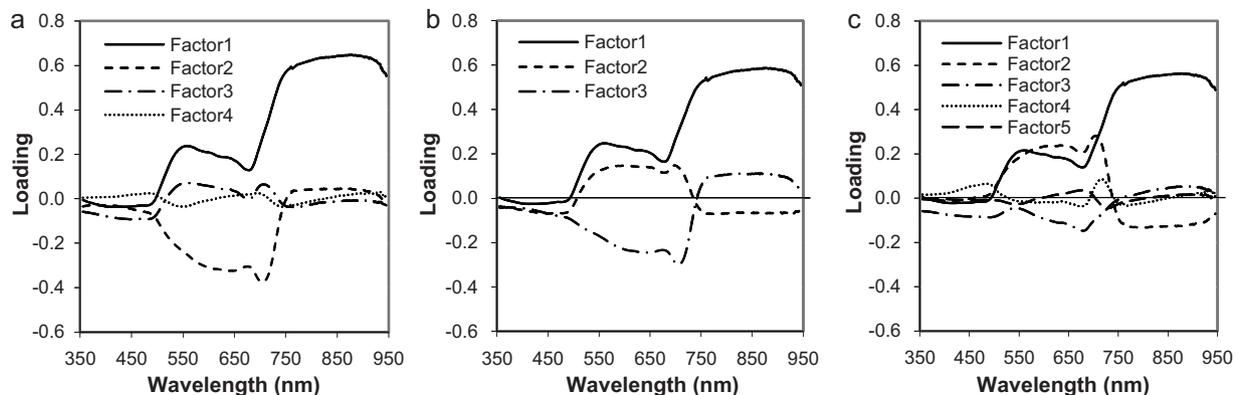


Fig. 4. The X-loading vectors of the fitted partial least squares regression models for (a) aboveground dry weight, (b) flower count, and (c) silique count.

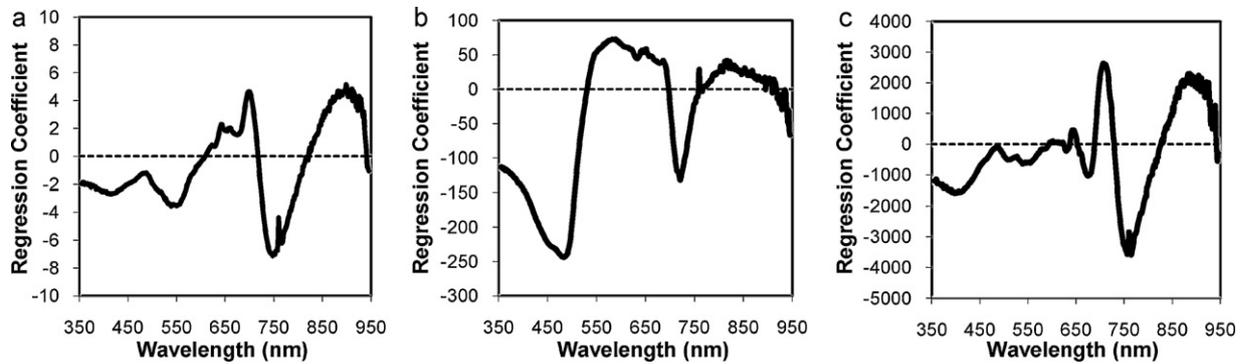


Fig. 5. Regression coefficients of the fitted partial least squares regression models for (a) aboveground dry weight, (b) flower count, and (c) silique count.

to the nutrient translocation processes that begin to occur as the crop increases its silique load, because gradual increases in red reflectance are expected as the plant leaves lose chlorophyll and overall vegetative vigor before maturity. Positive X-loadings for the fourth factor at 475 nm in the blue and at 700 nm in the far red may also be characterizing declines in canopy chlorophyll content that accompany the increase in silique load as the crop matures. The X-loadings may indicate a non-causal explanation for the relationship between spectral reflectance and silique number. Nonetheless, the PLSR model was able to capture variability in silique counts based on canopy spectral reflectance information.

The X-loading plots indicated how each spectral waveband contributed to the X-scores of each PLSR model. On the other hand, the regression coefficients, β_{PLS} , for each PLSR model highlighted the contribution of each spectral waveband to the overall prediction of each biophysical variable (Fig. 5). In the PLSR model for dry weight, local extrema in regression coefficients were found at 539, 699, 748, and 911 nm, which corresponds to visible green light, far red light at the foot of the red edge, NIR radiation at the shoulder of the red edge, and middle NIR radiation. These four wavebands effectively characterized the shape of the vegetation spectral signal and were thus the most important wavelengths to consider in the prediction of lesquerella dry weight. In the PLSR model for flower count, local extrema in the regression coefficients were found at 483, 583, 721, and 817 nm, corresponding to far blue light, near yellow light, the red edge inflection band, and the middle NIR. These wavebands demonstrated the contrast necessary to distinguish yellow light reflected from lesquerella flowers from light reflected in other regions of the visible spectrum. These wavebands for estimating flower count were also quite different from those found for dry weight, which demonstrates the strength of the PLSR method for selecting wavebands most critically related to the biophysical variable of interest. In the PLSR model for silique count, local

extrema in regression coefficients were found at 401, 708, 757, and 886 nm. Inclusion of a band in the near blue may help to estimate changes in canopy chlorophyll content that may occur concurrently with increasing silique number. It is interesting to note that the two bands for characterizing the red edge in the silique count model are 9 nm greater than that found for the dry weight model. These bands may also be more readily able to capture declining chlorophyll levels in the plant canopy. The regression coefficient results highlight the usefulness of expensive spectroradiometers and PLSR approaches to select key wavebands for detection of crop canopy biophysical responses. This information may be helpful in the construction of more inexpensive multispectral sensors that use information in several key wavebands for practical in-field estimation of these responses.

3.4. Model application

Application of the PLSR models using the canopy reflectance data collected along the 180 m transect in each plot demonstrated the ability of the models to track lesquerella growth and development for each of the planting date treatments. Of particular interest is the ability of PLSR model for flower count to identify the day of year (DOY) or days after planting (DAP) that peak flowering occurred. For the 2007–2008 experiment, the average peak flowering date was April 24 (DOY 115; DAP 209) for the first planting, while peak flowering was May 16 (DOY 137; DAP 91) for the second planting (Fig. 6). Similarly for the 2008–2009 experiment, the average peak flowering date was April 20 (DOY 110; DAP 196) for the first planting, and peak flowering for the second and third planting dates both occurred on June 1 (DOY 152; DAP 144 and 115) (Fig. 7). Because lesquerella flowering is indeterminate in nature, remote sensing may be a useful tool to track flowering patterns and identify when the crop begins to flower less vigorously. In this way, remote

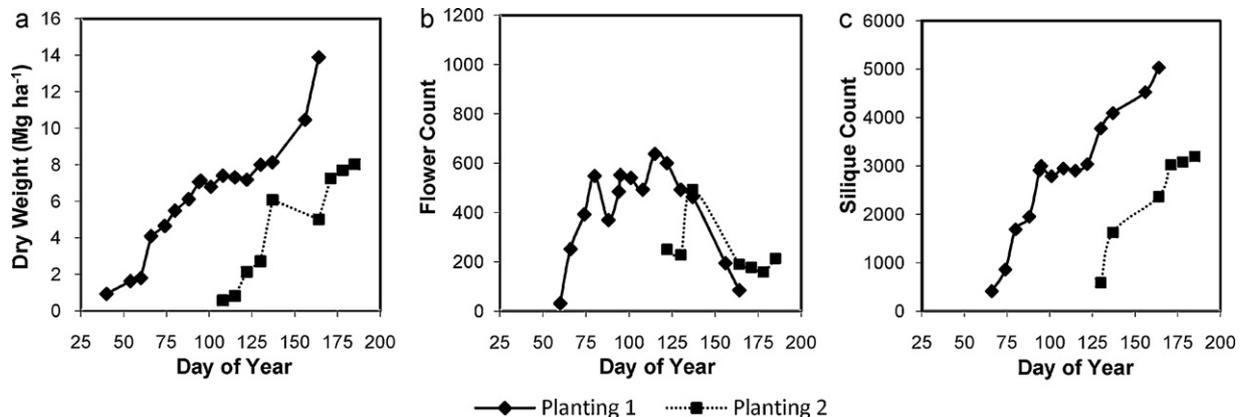


Fig. 6. Average modeled (a) aboveground dry weight, (b) flower count, and (c) silique count using the transect reflectance data collected from February to July of 2008.

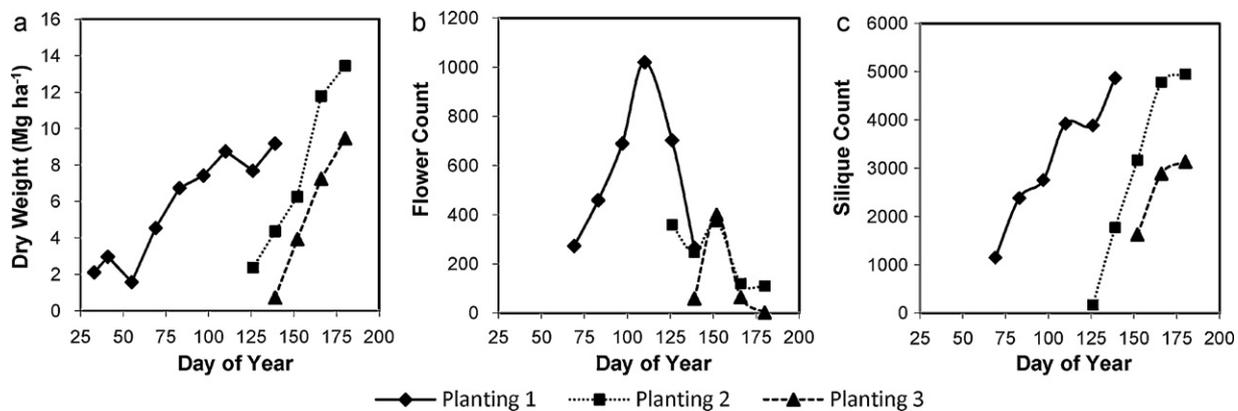


Fig. 7. Average modeled (a) aboveground dry weight, (b) flower count, and (c) silique count using the transect reflectance data collected from February to July of 2009.

sensing techniques may aid a grower's decision to cut off irrigation water in the late season or to apply desiccant in preparation for crop harvest. For lesquerella crops in Arizona, these decisions are particularly important, since the crop may likely be grown in rotation with cotton, and earlier lesquerella harvest dates would allow more time to prepare the field for the cotton crop or allow more flexibility to add a different summer crop, such as sorghum, into the rotation scheme. Mature siliques have also been known to easily shatter in heavy winds or rain, so quantification of optimum harvest dates may help prevent yield loss due to adverse weather. In addition to remote sensing estimates of flower count, the silique count estimates may also be useful for deciding the optimum time to harvest lesquerella. Plots given in Figs. 6 and 7 also demonstrate how the remote sensing techniques may be useful for estimating crop traits of interest to lesquerella breeders.

Estimates from the PLSR model for silique count demonstrate how remote sensing could be developed as a yield prediction tool for lesquerella. For example, in the 2007–2008 growing season, average grain yields for the first and second plantings were 1689 and 462 kg ha⁻¹, respectively. Based on the difference in silique count predictions from remote sensing (Fig. 6c) in this growing season, this difference in final yield between the planting date treatments could be reasonably expected.

4. Conclusions

Partial least squares regression is a useful statistical tool for analyzing hyperspectral datasets and relating crop canopy spectra to measured biophysical variables. Due to the mathematical complexity of the technique, its value lies mainly in the reduction of hyperspectral datasets to determine the spectral wavebands that are most greatly related to a crop biophysical variable of interest. Likely, the remote sensing methods used in this study will not be practically implemented in a production field setting. Rather the results may be used to develop more practical, more inexpensive radiometers that are specifically tailored to estimate a particular crop trait, such as lesquerella flower count. The remote sensing methods implemented in this study have wide potential applicability to other crop species and other crop canopy traits. The methods may also be applied to improve remote sensing technologies in several application areas, including precision crop management and high-throughput phenotyping efforts. Perhaps the most important consideration is that the methods are used to more clearly understand the mechanism for reflectance of light from crop canopies and why it is possible to estimate some canopy traits from spectral measurements. This approach is more preferable than the alternative one that simply seeks high statistical correlation between crop canopy traits and spectral reflectance.

Acknowledgements

The authors are grateful to the field technicians who conducted the experiments and collected and processed the spectral reflectance data and biomass samples, including Ms. Suzette Maneely, Mr. Rafael Chavez-Alcorta, Ms. Gail Dahlquist, and Ms. Stephanie Johnson.

References

- Adamsen, F.J., Coffelt, T.A., Nelson, J.M., Barnes, E.M., Rice, R.C., 2000. Method for using images from a color digital camera to estimate flower number. *Crop Sci.* 40 (3), 704–709.
- Bajwa, S.G., 2006. Modeling rice plant nitrogen effect on canopy reflectance with partial least square regression (PLSR). *Trans. ASABE* 49 (1), 229–237.
- Blackmer, T.M., Schepers, J.S., Varvel, G.E., Meyer, G.E., 1996. Analysis of aerial photography for nitrogen stress within corn fields. *Agron. J.* 88 (5), 729–733.
- de Jong, S., 1993. SIMPLS—an alternative approach to partial least-squares regression. *Chemometr. Intell. Lab. Syst.* 18 (3), 251–263.
- Dierig, D.A., Adam, N.R., Mackey, B.E., Dahlquist, G.H., Coffelt, T.A., 2006. Temperature and elevation effects on plant growth, development, and seed production of two Lesquerella species. *Ind. Crop. Prod.* 24 (1), 17–25.
- Fridgen, J.L., Varco, J.J., 2004. Dependency of cotton leaf nitrogen, chlorophyll, and reflectance on nitrogen and potassium availability. *Agron. J.* 96 (1), 63–69.
- García, R.R., de Rodriguez, D.J., Angulo Sanchez, J.L., Dierig, D.A., Solis, H.D., De la Rosa Loera, A., 2007. *Lesquerella fendleri* response to different sowing dates in northern Mexico. *Ind. Crop. Prod.* 25 (2), 117–122.
- Geller, D.P., Goodrum, J.W., 2004. Effects of specific fatty acid methyl esters on diesel fuel lubricity. *Fuel* 83 (17–18), 2351–2356.
- Horler, D.N.H., Dockray, M., Barber, J., 1983. The red edge of plant leaf reflectance. *Int. J. Remote Sens.* 4 (2), 273–288.
- Hunsaker, D.J., Nakayama, F.S., Dierig, D.A., Alexander, W.L., 1998. Lesquerella seed production: water requirement and management. *Ind. Crop. Prod.* 8 (2), 167–182.
- Hunsaker, D.J., Pinter, P.J., Kimball, B.A., 2005. Wheat basal crop coefficients determined by normalized difference vegetation index. *Irrigation Sci.* 24 (1), 1–14.
- Idso, S.B., Jackson, R.D., Reginato, R.J., 1977. Remote sensing for agricultural water management and crop yield prediction. *Agric. Water Manage.* 1 (4), 299–310.
- Jackson, R.D., Idso, S.B., Reginato, R.J., Pinter, P.J., 1981. Canopy temperature as a crop water-stress indicator. *Water Resour. Res.* 17 (4), 1133–1138.
- Jakubauskas, M.E., Legates, D.R., Kastens, J.H., 2002. Crop identification using harmonic analysis of time-series AVHRR NDVI data. *Comput. Electron. Agric.* 37 (1–3), 127–139.
- Knipling, E.B., 1970. Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. *Remote Sens. Environ.* 1 (3), 155–159.
- Mevik, B.-H., Wehrens, R., 2007. The pls package: principle component and partial least squares regression in R. *J. Stat. Software* 18 (2), 1–24.
- Mogensen, V.O., Jensen, C.R., Mortensen, G., Thage, J.H., Koribidis, J., Ahmed, A., 1996. Spectral reflectance index as an indicator of drought of field grown oilseed rape (*Brassica napus* L.). *Eur. J. Agron.* 5 (1–2), 125–135.
- Montes, J.M., Melchinger, A.E., Reif, J.C., 2007. Novel throughput phenotyping platforms in plant genetic studies. *Trends Plant Sci.* 12 (10), 433–436.
- Moser, B.R., Cermak, S.C., Isbell, T.A., 2008. Evaluation of castor and lesquerella oil derivatives as additives in biodiesel and ultralow sulfur diesel fuels. *Energy Fuels* 22 (2), 1349–1352.

- Nelson, J.M., Dierig, D.A., Nakayama, F.S., 1996. Planting date and nitrogen fertilization effects on lesquerella production. *Ind. Crop. Prod.* 5 (3), 217–222.
- Nelson, J.M., Watson, J.E., Dierig, D.A., 1999. Nitrogen fertilization effects on lesquerella production. *Ind. Crop. Prod.* 9 (3), 163–170.
- Puppala, N., Fowler, J.L., Jones, T.L., Gutschick, V., Murray, L., 2005. Evapotranspiration, yield, and water-use efficiency responses of *Lesquerella fendleri* at different growth stages. *Ind. Crop. Prod.* 21 (1), 33–47.
- Roseberg, R.J., 1996. Herbicide tolerance and weed control strategies for Lesquerella production. *Ind. Crop. Prod.* 5 (2), 133–139.
- Smith, C.R., Zobel, H., Wolff, I.A., Miwa, T.K., Wilson, T.L., Lohmar, R.L., 1961. Lesquerolic acid—a new hydroxy acid from lesquerella seed oil. *J. Org. Chem.* 26 (8), 2903–2905.
- Thorp, K.R., Steward, B.L., Kaleita, A.L., Batchelor, W.D., 2008. Using aerial hyperspectral remote sensing imagery to estimate corn plant stand density. *Trans. ASABE* 51 (1), 311–320.
- Ye, X.J., Sakai, K., Sasao, A., Asada, S., 2009. Estimation of citrus yield from canopy spectral features determined by airborne hyperspectral imagery. *Int. J. Remote Sens.* 30 (18), 4621–4642.