Prediction of senescent rangeland canopy structural attributes with airborne hyperspectral imagery

Rebecca Phillips¹, Mark West², Nicanor Saliendra¹, Brad Rundquist³ & Duane Pool⁴

¹ USDA-ARS, Northern Great Plains Research Laboratory, PO Box 459, 1701-10th Avenue SW Mandan, ND, 58554, USA
² USDA-ARS, 2150 Centre Avenue, Fort Collins, CO, 80526-8119, USA
³ University of North Dakota, 221 Centennial Drive, Stop 9020, Grand Forks, ND, 58202, USA
⁴ Rocky Mountain Bird Observatory, 1500 Edwards Avenue, Bismarck, ND, 58506, USA

Published online: 14 Jun 2013.

To cite this article: Rebecca Phillips, Mark West, Nicanor Saliendra, Brad Rundquist & Duane Pool (2013) Prediction of senescent rangeland canopy structural attributes with airborne hyperspectral imagery, GIScience & Remote Sensing, 50:2, 133-153

To link to this article: http://dx.doi.org/10.1080/15481603.2013.793469

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing,
Prediction of senescent rangeland canopy structural attributes with airborne hyperspectral imagery

Rebecca Phillipsa*, Mark Westb, Nicanor Saliendraa*, Brad Rundquistc and Duane Poold

aUSDA-ARS, Northern Great Plains Research Laboratory, PO Box 459, 1701-10th Avenue SW Mandan, ND 58554, USA; bUSDA-ARS, 2150 Centre Avenue, Fort Collins, CO 80526-8119, USA; cUniversity of North Dakota, 221 Centennial Drive, Stop 9020, Grand Forks, ND 58202, USA; dRocky Mountain Bird Observatory, 1500 Edwards Avenue, Bismarck, ND 58506, USA

(Received 23 October 2012; final version received 17 April 2013)

Canopy structural and chemical data are needed for senescent, mixed-grass prairie landscapes in autumn; yet data-driven models are lacking for rangelands dominated by non-photosynthetically active vegetation (NPV). We report how field data and aerial hyperspectral imagery were modeled to predict canopy attributes post growing-season using two approaches: (1) application of narrow spectral regions with Vegetation Indices (VIs) and (2) application of the full spectrum with Partial Least Squares Regression (PLSR). Analyses of the full spectrum using PLSR resulted in slightly lower root-mean-square error of prediction, as compared to VIs, which represent reflectance ratios for specific spectral bands.

Keywords: remote sensing; grassland; landscape; Partial Least Squares Regression (PLSR); Vegetation Indices (VIs); Short-Wave InfraRed (SWIR)

Introduction

Adaptive rangeland management plans require data at local and landscape scales to assess effects of climate and disturbances on canopy structural and chemical characteristics, such as photosynthetically active vegetation (PV), non-photosynthetically active vegetation (NPV), total standing crop (TSC), canopy nitrogen (N) content, percent bare ground (%BG), and canopy height (Moran et al. 1997; Hunt et al. 2003). Furthermore, agencies tasked with managing rangelands for multiple ecosystem functions often require these data pre, post, and during the growing season, particularly for wildlife habitat assessment (Dechant et al. 2002; Fisher and Davis 2010; Herrick et al. 2012). Managers in North and South Dakota, USA, typically manually measure canopy height in autumn after cattle are removed to evaluate the remaining canopy structure important to rare and threatened avian species. Another, perhaps complementary approach includes synoptic mapping of canopy structural attributes with remote sensing-based data (Pickup et al. 1994; Toevs et al. 2011). Remote sensing techniques have been developed when plants are mostly green to help meet rangeland forage information needs (Tucker et al. 1983; Guerschman et al. 2009; Brinkmann, Dickhoefer, et al. 2011; Brinkmann, Patzelt, et al. 2011), but techniques are lacking for senescent canopies post-growing-season (Tucker 1978; Roberts et al. 1993; Roberts et al. 1998; Rundquist 2002). The need to assess rangelands following grazer removal in autumn (Bement 1969) and the potential for hyperspectral data applications are clear; however, specific approaches to
understanding and modeling senescent mixed-grass prairie need to be evaluated for expansive rangeland landscapes in the northern Great Plains.

Rangeland canopies near the 100th meridian are often mixtures of PV and NPV (Frank and Aase 1994; Beeri et al. 2007), and remote sensing techniques for canopies dominated by PV have been developed. Several groups have investigated how vegetation indices (VIs) derived from airborne hyperspectral data can be modeled to estimate biomass, N content, and leaf area during the growing season (Jacquemoud et al. 1995; Kokaly and Clark 1999; Beeri et al. 2007). For example, the best predictor for savannah and mixed-grass prairie PV during the growing season was the Normalized Difference Vegetation Index (NDVI) derived from hyperspectral imagery (Gamon et al. 1997; Beeri et al. 2007; Guerschman et al. 2009; Brinkmann, Dickhoefer, et al. 2011; Brinkmann, Patzelt, et al. 2011). Guerschmann et al. (2009) also found that spectral data available from the Moderate Infrared Spectrometer (MODIS), the simple ratio \( \rho_{2130}/\rho_{1640} \), was the optimum VI for estimating fractional cover of NPV. Airborne hyperspectral imagery has also been applied post growing-season to estimate annual crop cover (corn, wheat and alfalfa residues) with the Cellulose Absorption Index (Daughtry et al. 2005). However, residues for monocultures are more homogeneous than senescent native rangelands, and these natural environments comprised of mixed-materials present challenges for biomass estimations using VIs (Guo et al. 2003; Fava et al. 2009; Guerschman et al. 2009; Brinkmann, Dickhoefer, et al. 2011; Brinkmann, Patzelt, et al. 2011). Consequently, this work required analysis of not only several well-established VIs, but also analysis of the entire spectrum with PLSR.

PLSR is a common method for analyzing airborne hyperspectral data, particularly for images collected during the growing season. Huang et al. (2004) demonstrated that PLSR was a better alternative to conventional stepwise regression for estimating foliar N in eucalypt tree canopies. Schmidtlein and Sassin (2004) used PLSR for mapping floristic gradients in grasslands, and PLSR predicted wheat biomass with reflectance measurements obtained with a field spectrometer (Hansen and Schjoerring, 2003). Cho et al. (2007) also concluded that modeling airborne hyperspectral data with PLSR led to more accurate prediction of herbaceous green biomass than VIs. The predictive potential of PLSR using airborne hyperspectral imagery in senescent rangelands post growing-season still requires investigation.

The primary objectives of this research were to (i) assess the utility of hyperspectral VIs derived from airborne hyperspectral imagery to predict PV, NPV, TSC, N content, and %BG for a herbaceous rangeland landscape in autumn, (ii) determine whether PLSR results in lower prediction errors for canopy attributes when compared to VI-based models, and (iii) determine which of the four spectral regions [visible (VIS), near infrared (NIR), short-wave infrared one (SWIR1), or short-wave infrared two (SWIR2)] contributed most to predictive performance. To achieve these objectives, field data and airborne images collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) were collected over the Grand River National Grassland, South Dakota. Predictions using VI model selection and PLSR procedures were determined and compared using calibration and test data sets.

**Materials and methods**

**Study area description**

Our research site is located at the Grand River National Grassland (GRNG), South Dakota, USA (45.7°N, 102.5°W) within the Northern Great Plains ecoregion (Omernik
1987). About 75% of the annual precipitation (350 mm) occurs during the growing season (April–September). Average monthly temperature is highest in July (21°C) and lowest in January (−9°C). Topography ranges from open plains to rolling grassland prairie, with elevations from 670–880 m. Soils are predominantly well-drained, moderately deep, moderately permeable, fine-loamy, mixed, superactive, and frigid Typic Argiustolls. The GRNG is a mixed-grass prairie ecosystem characterized by blue grama [Bouteloua gracilis (H.B.K.) Lag. Ex Griffiths] and western wheatgrass [Pascopyron smithii (Rybd.) Löve]. Many of the GRNG lowlands were farmed in the early twentieth century and are now dominated by crested wheatgrass [Agropyron cristatum (L) Gaertn.]. The GRNG is seasonally grazed by cattle, and stocking rates are approximately one animal unit per hectare between May and October each year.

Field data collection site details are fully described in Phillips et al. (2012). Briefly, a 100,000 ha landscape of interest (LOI) in the center of the GRNG was selected for this study and stratified to ensure field data were representative of the landscape for herbaceous vegetation (Pickup et al. 1994; Toevs et al. 2011). We identified herbaceous vegetation areas using a Landsat 5 image and geo-located species data in the Definiens eCognition Developer (v. 7™) object-based classification software. Further, we modeled Digital Elevation Model (DEM) data [http://ned.usgs.gov/(Gesch et al. 2002)] and classified the landscape into three topographic positions: summits, mid-slopes, or toeslopes (Qin et al. 2009). We also evaluated historical (10 y) reflectance data collected by the Moderate Resolution Imaging Spectrometer (MODIS) sensor using the Enhanced Vegetation Index (EVI) collected in June and July from 2000 to 2009. We identified four classes (from low to high EVI values) using an unsupervised classification in geographic information systems (GIS) that comprised more than 98% of the herbaceous vegetation LOI. All of the herbaceous landscape was assigned into one of the four groups, which we refer to as a Historic Vegetation Index (HVI). We randomly selected six plots within each HVI class. Within each plot, the south facing slope nearest to the plot center was selected and divided into summit, mid-slope, and toeslope positions for a total of 72 sample sites. Consistent trends in species covered by topographic position groups were previously reported for all sites (Phillips et al. 2012), where mid and toeslopes were dominated by mid-grass species, such as P. smithii and summits were dominated by short-grass species, such as B. gracilis. Data collected in July 2010 indicated PV, NPV, TSC, N, and height did not vary with HVI class but did vary with topographic position group (Phillips et al. 2012).

**Field data collection**

At each of the 72 sites, vegetation height was measured 3 m from center in each of the four cardinal directions with a Robel pole (Robel et al. 1970; Uresk and Benson 2007). After recording the height, a 0.187-m² frame was placed in front of the Robel pole. A visual estimate of the percent vegetation and bare ground cover within the frame was recorded using the Daubenmire method (Daubenmire, 1959). All vegetation within the frame was clipped to 2 cm above the soil surface, separated into PV and NPV groups, and dried for 48 h at 60°C. Total standing crop biomass (TSC, kg ha⁻¹) was calculated as the sum of PV and NPV. Vegetation for PV was ground separately from NPV through a 1-mm mesh screen, and analyzed for total N using dry combustion on a Carlo Erba Model NA 1500 Series 2 N/C/S analyzer (CE Elantech, Lakewood, NJ). Canopy N content (kg N ha⁻¹) was calculated using N content and mass for both PV and NPV vegetation. Average PV, NPV, TSC, Canopy N, and percent vegetation cover by point were used in all
subsequent analyses. Each point was precisely geo-located (<1 m spatial resolution) using a Global Positioning System (Trimble Geo XT).

**AVIRIS data collection and correction**

Airborne AVIRIS data were collected on 21 October 2010 over the study area during the field campaign. The flight was carried out by NASA’s Jet Propulsion Laboratory (JPL), Pasadena, CA on a Twin Otter aircraft. The AVIRIS sensor operates over the spectral range of 360–2500 nm using four on-board spectrometers (VIS, NIR, SWIR1, and SWIR2). Average spectral and spatial resolutions for the data collected were 10 nm and 4 m, respectively. Data were acquired over the 72 sample sites in six image strips under clear, cloud-free conditions. Field sites were located near the center of each image strip. Image acquisition was close to solar noon and the solar zenith and azimuth angles for the six image strips ranged from 57 to 60 and from 170 to 205, respectively. The images were processed by NASA and then atmospherically corrected using the Fast Line-of-Sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) module in the ENVI/IDL® software (Atmospheric Correction Module 4.7) and converted to units of reflectance (scaled between 0% and 100%).

Dataset accuracy and quality were evaluated for each image according to Beeri et al. (2007). Briefly, we geometrically corrected each image in the projection UTM Zone 13 North, Datum WGS 84 with 1 m resolution aerial photography from the National Agriculture Imagery Program. Twelve ground control points were identified that were spatially distributed throughout our AOI. We found geo-location root-mean-square-error (RMSE) was smaller than one-half pixel. To evaluate signal-to-noise ratio (SNR), spectra were collected from uniform areas in each image. Low SNRs for wavelengths <500, 1353–1423, 1812–1937, and >2470 were eliminated from the dataset, for a total of 180 bands remaining. We evaluated image homogeneity for all bands using spectral signatures from roads that cross each image. We confirmed that image reflectance values were similar to ground-based reflectance values recorded at uniform areas. We extracted data for each sample point using a 2 × 2 pixel window (i.e., 8 × 8 m) centered on each GPS waypoint to include the full (6 × 6 m) sample plot area. Vegetation indices (Table 1) were calculated for each pixel, and mean values for each of the 72-pixel windows were compared with field data.

**Statistical analyses**

We investigated the potential for AVIRIS reflectance values to predict PV, NPV, TSC, Canopy N content, and %BG by 1) analyzing the entire spectrum and 2) analyzing only specific spectral regions with VIs. For the first, we fit PLSR (Garthwaite, 1994) models for each dependent variable using the reflectance data collected for all spectral bands. For the second, we applied model selection procedures using a suite of established VIs as candidate predictors (Table 1). We aimed to build linear prediction models that included spectral, topographic position and HVI class data. The model selected for each prediction method was tested iteratively by partitioning the available data at random 1000 times into training and test sets. For each partition, models were fitted to each training set and these were used to predict observations in each test set. The ratio of training samples to test samples in each of the 1000 partitions was chosen as 2:1. Goodness-of-fit for each model was reported in terms of average $R^2$ across the random
Table 1. Vegetation Indices (VIs) calculated using narrow-band Airborne Visible-Infrared Imaging Spectrometer (AVIRIS) reflectance data collected on 21 October 2010 over the Grand River National Grassland (GRNG). Original reference wavelengths ($\rho$) were modified to match those collected at the AVIRIS sensor.

<table>
<thead>
<tr>
<th>Index acronym</th>
<th>Algorithm (modified for AVIRIS)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARI1, anthocyanin reflectance 1</td>
<td>$ARI1 = \left(\frac{1}{\rho_{550}}\right) - \left(\frac{1}{\rho_{705}}\right)$</td>
<td>(Gitelson et al. 2001)</td>
</tr>
<tr>
<td>ARI2, anthocyanin reflectance 2</td>
<td>$ARI2 = \rho_{802} \left(\frac{1}{\rho_{550}} - \frac{1}{\rho_{705}}\right)$</td>
<td>(Gitelson et al. 2001)</td>
</tr>
<tr>
<td>CAI, cellulose absorption</td>
<td>CAI = 0.5(\rho_{1997} + \rho_{2198}) - \rho_{2098}</td>
<td>(Daughtry 2001)</td>
</tr>
<tr>
<td>MSI, moisture stress</td>
<td>$MSI = \rho_{1602}/\rho_{822}$</td>
<td>(Hunt and Rock 1989)</td>
</tr>
<tr>
<td>MTVI1, modified transformed 1</td>
<td>$MTVI1 = 1.2[1.2(\rho_{802} - \rho_{550}) - 2.5(\rho_{638} - \rho_{550})]$</td>
<td>(Haboudane et al. 2004)</td>
</tr>
<tr>
<td>MTVI2, modified transformed 2</td>
<td>$MTVI2 = \frac{1.5[1.2(\rho_{802} - \rho_{550}) - 2.5(\rho_{638} - \rho_{550})]}{\sqrt{(2\rho_{802} + 1)^2 - (6\rho_{802} - 5\sqrt{\rho_{638} - 0.5})}}$</td>
<td>(Haboudane et al. 2004)</td>
</tr>
<tr>
<td>NDII, normalized delta infrared</td>
<td>$NDII = (\rho_{822} - \rho_{1652})/(\rho_{822} + \rho_{1652})$</td>
<td>(Hardisky et al. 1983)</td>
</tr>
<tr>
<td>NDLI, normalized delta lignin</td>
<td>$NDLI = \frac{\log\left(\frac{1}{\rho_{1752}}\right) - \log\left(\frac{1}{\rho_{1682}}\right)}{\log\left(\frac{1}{\rho_{1752}}\right) + \log\left(\frac{1}{\rho_{1682}}\right)}$</td>
<td>(Serrano et al. 2002)</td>
</tr>
<tr>
<td>NDNI, normalized delta nitrogen</td>
<td>$NDNI = \frac{\log\left(\frac{1}{\rho_{1512}}\right) - \log\left(\frac{1}{\rho_{1682}}\right)}{\log\left(\frac{1}{\rho_{1512}}\right) + \log\left(\frac{1}{\rho_{1682}}\right)}$</td>
<td>(Serrano et al. 2002)</td>
</tr>
<tr>
<td>NDSVI, normalized senescent</td>
<td>$NDSVI = (\rho_{1652} - \rho_{658})/(\rho_{1652} + \rho_{658})$</td>
<td>(Marsett et al. 2006)</td>
</tr>
<tr>
<td>NDVI, normalized delta</td>
<td>$NDVI = (\rho_{802} - \rho_{638})/(\rho_{802} + \rho_{638})$</td>
<td>(Rouse et al. 1977)</td>
</tr>
</tbody>
</table>
Table 1. (Continued).

<table>
<thead>
<tr>
<th>Index acronym</th>
<th>Algorithm (modified for AVIRIS)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDVI705, red edge normalized delta</td>
<td>$\text{NDVI705} = (\rho_{861} - \rho_{1245})/(\rho_{861} + \rho_{1245})$</td>
<td>(Sims and Gamon 2002)</td>
</tr>
<tr>
<td>NDWI, normalized difference water</td>
<td>$\text{NDWI} = (\rho_{1652} - \rho_{832})/(\rho_{1652} + \rho_{832})$</td>
<td>(Gao 1996)</td>
</tr>
<tr>
<td>OSAVI, optimized soil adjusted</td>
<td>$\text{OSAVI} = 1.16 \times (\rho_{802} - \rho_{675})/(\rho_{802} + \rho_{675} + 0.16)$</td>
<td>(Rondeaux et al. 1996)</td>
</tr>
<tr>
<td>PRI, photochemical reflectance</td>
<td>$\text{PRI} = (\rho_{531} - \rho_{570})/(\rho_{570} + \rho_{531})$</td>
<td>(Gamon et al. 1997)</td>
</tr>
<tr>
<td>PSRI, plant senescence reflectance</td>
<td>$\text{PSRI} = (\rho_{678} - \rho_{502})/\rho_{754}$</td>
<td>(Merzlyak et al. 1999)</td>
</tr>
<tr>
<td>REI, red edge</td>
<td>$\text{REI} = \rho_{705}/\rho_{754}$</td>
<td>(Gitelson et al. 1996)</td>
</tr>
<tr>
<td>SATVI, soil adjusted total VI</td>
<td>$\text{SATVI} = 1.5 \left[ \frac{(\rho_{1642} - \rho_{658})}{\rho_{1652} + \rho_{658} + 0.5} \right] - \left[ \frac{(\rho_{2218})}{2} \right]$</td>
<td>(Marsett et al. 2006)</td>
</tr>
<tr>
<td>SWIR-SR</td>
<td>$\text{SWIR-SR} = \rho_{2128}/\rho_{1642}$</td>
<td>(Guerschman et al. 2009)</td>
</tr>
<tr>
<td>TCARI, transformed chlorophyll absorption</td>
<td>$\text{TCARI} = 3 \left[ (\rho_{705} - \rho_{675}) - 0.2 \times (\rho_{705} - \rho_{550}) \times \left( \frac{\rho_{550}}{\rho_{675}} \right) \right]$</td>
<td>(Haboudane et al. 2002)</td>
</tr>
<tr>
<td>TCARI/OSAVI</td>
<td>$\text{TCARI}/\text{OSAVI}$</td>
<td>(Haboudane et al. 2002)</td>
</tr>
<tr>
<td>WBI, water band</td>
<td>$\text{WBI} = \rho_{899}/\rho_{967}$</td>
<td>(Penuelas et al. 1993)</td>
</tr>
</tbody>
</table>
partitions, which was calculated by the square of the Pearson correlation between
observed and predicted response. We also computed root mean square error of prediction
(RMSEP) to provide a direct estimate of the modeling error expressed in original
measurement units:

\[
\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n} (M_i - P_i)^2}{n}}
\]  

(1)

where \( P_i \) and \( M_i \) were the predicted and measured crop variables, respectively, and \( n \) the
number of samples \( (n = 72) \). RMSEP provides a direct estimate of the modeling error
expressed in original measurement units. Statistical computations were conducted using
NC, USA) and \( R \) statistical package (http://R-Project.org).

**Partial least squares regression approach**

We developed PLSR prediction models (one for each dependent variable) in the form
\( \hat{Y} = R_x \beta \) using the method of de Jong (de Jong, 1993) where \( \hat{Y} \) is an \( n \times 1 \) vector of
predictions of vegetative growth and \( R_x \) is an \( n \times p \) matrix of reflectance values. Each row
vector of \( R \) consists of \( p \) discrete entries of reflectance values \( R_x \) representing the amount

![Graphs showing R² values for different canopy attributes](image-url)

Figure 1. \( R^2 \) computed from the PRESS statistic is plotted against the number of factors used in
PLS prediction models for models including up to 10 factors. These plots were used as a guide for
determining the least number of factors to use for PLS regression for each canopy attribute. \( R^2 \) is
greater when prediction is based on five factors for PV, NPV, TSC and %BG, but \( R^2 \) is greater when
three factors are used for predicting Canopy N.
Reflectance was measured for \( p = 180 \) distinct bands ranging from 511 nm to 2457 nm (i.e., \( 511 \leq x \leq 2457 \)). Leave-one-out cross validation was used on the entire data set to determine the number of factors to be used for PLSR regression. We chose the number of PLSR factors based on the PRESS statistic and \( R^2 \) values for each response variable (Figure 1). The latter is computed as \( R^2 = 1 - \frac{SSE}{SST} \) where \( SSE \) is the residual sum of squares for the cross-validated predictions and \( SST \) is the (corrected) total sums of squares for the response.

We evaluated the ability of PLSR to predict canopy attributes using Monte Carlo sampling. Sixteen plots (four plots of six possible from each of four HVI classes) were randomly sampled for the training set for each of 1000 Monte Carlo trials. Data from the remaining eight plots not used in the training dataset comprised the test data set in each Monte Carlo trial. The fitted PLSR for each Monte Carlo trial was used to predict observations of canopy response in both the trial and test data sets and the squared correlation coefficient between fitted and observed was computed for all trials. These were used to test how well the fitted regressions could predict observations not used in fitting the models. Monte Carlo evaluations of the models were conducted separately from the Monte Carlo trials used for model selection.

An additional step was taken to further constrain spectral regions most influenced by measured canopy attributes using principal components. The data were analyzed with respect to the spectral regions, or banks of detectors, collected at the AVIRIS sensor. The AVIRIS is comprised of four spectrometers, and each collect VIS (360–670 nm), NIR (660–1280 nm), SWIR1 (1260–1880 nm), or SWIR2 (1880–2500 nm) data (Green et al. 1999). For each spectral region, only two principal components (on the variance–covariance, unscaled data) were needed to capture most (~99%) of the reflectance information within each region (data not shown). Therefore, two components for each region were used to determine the spectral region most highly influenced by a given canopy attribute. These components were named VIS-1, VIS-2, NIR-1, NIR-2, SWIR1-1, SWIR1-2, SWIR2-1, and SWIR2-2. The first principal component for each region can be interpreted as a size component because the principal component loadings (the correlations between reflectance at each band and the principal component scores) are approximately the same size and sign for all bands. The second principal component is a contrast of spectral regions, where one region correlates bands with higher than average reflectance and the other region correlates bands with lower than average reflectance. These components loadings vacillate between negative and positive at most only two times, allowing for a straightforward interpretation of the components.

Both the \( R \) package “pls” and the SAS procedure PROC PLS were used to fit the PLS regressions. Both software packages provided the same regression model fits but the \( R \) package “pls” was used to facilitate simulation for evaluating goodness of fit. Coefficients from the fitted PLS equations were plotted against corresponding bands with the intention of identifying bands most important for prediction. Standard errors for the PLSR coefficients were bootstrapped using “random X sampling” and were used to standardize the coefficients that were plotted.

**Model selection approach**

Our second approach involved model selection procedures aimed toward identifying the VI, or combination of VIs (Table 1) that were most predictive of canopy attributes computed from AVIRIS spectral band data. The procedures included historical reflectance
class (HVI) and topographic position group for each sample site and VIs as candidate predictors. In addition, we allowed for the squared term for each VI to be included as a predictor as well as any two-way interaction among the VIs, HVI class, and topographic position group. Thus, 349 regression effects were entered as possible predictors to be included in a model for each canopy attribute. Our approach for model selection was based on re-sampling and model selection frequency and incorporated both Akaike and cross-validation criteria. To carry out the re-sampling model selection, the following algorithm was used.

Re-Sampling Model Selection Algorithm

Do the following steps 1000 times.

1. Randomly partition the data to form two sets.
   a. From each HVI class select four plots at random. This partitioned the data set into a training set and a validation set. The remaining two plots not selected in each HVI class are used to create a validation set. Partitioning in this manner ensures equal representation of both HVI class and topographic position group.

2. Use stepwise regression on the partitioned data to choose a prediction model.
   a. Order in which effects enter model based on the Akaike Information Criterion.
   b. Significance level for effects to enter set at 0.05.
   c. Significance level for effects to stay set at 0.10.
   d. Choose the model among the steps of model selection with the smallest average mean squared prediction error (AMSE) from the validation sample.
   e. Require model hierarchy, so that when interactions between two variables are entered, each of these variables is also included in the model.

3. Record the chosen model for each partition (i.e., each Monte Carlo sample)
   a. Report selected models and model selection frequencies

The model selection described above was carried out in SAS using the SURVEYSELECT procedure to randomly partition the data and the GLMSELECT procedure with options choose = validate, select = aicc, sle = 0.05, sls = 0.10, stop = sl and hier = single to carry out model selection for each partition.

The statistic described above avoided over-fitting by using random partitioning and re-sampling to identify the most reliable model or models. We compared these results with the more conventional stepwise regression approach using the same suite of VIs (Table 1). The conventional stepwise regression selected the model that minimized the PRESS statistic (leave one out cross-validation) but did not randomly partition and resample the data.

Field data analyses

Total standing crop (kg ha\(^{-1}\)), NPV (kg ha\(^{-1}\)), PV (kg ha\(^{-1}\)), Canopy N (kg ha\(^{-1}\)), and canopy height (cm) were analyzed separately by fitting a generalized linear mixed model (SAS System for Windows, copyright © 2002–2008, SAS Institute Inc., Cary,
NC, USA). The model-fixed effects included topographic position, HVI class, and the interaction of topographic position and HVI group, and random effects included plot and plot × topographic position. A Tukey–Kramer test was used to determine if least-square mean values at specific topographic positions or specific HVI groups were significantly different from each other. TSC was regressed on canopy height to determine the relationship between TSC and height for October 2010 (Vermiere et al. 2002; Uresk and Benson 2007).

Results

Field data

Similar to data collected at these points in July, measured canopy attributes varied significantly with topographic position (Table 2), including PV ($F_{2,40} = 9.30; p < 0.0005$), TSC ($F_{2,40} = 18.05; p < 0.0001$), NPV ($F_{2,40} = 15.24; p < 0.0001$), canopy N ($F_{2,40} = 35.50; p < 0.0001$), bare ground ($F_{2,40} = 23.78; p < 0.0001$), and canopy height ($F_{2,40} = 29.14; p < 0.0001$). Neither HVI class nor HVI class × topographic position interaction was significant for any of these variables. The effects of topographic position were clear, with TSC, PV, NPV, height, and Canopy N lowest at summits and %BG highest at summits.

The TSC and canopy height data were fitted with a linear regression ($R^2 = 0.62$), where

$$\text{TSC} = 50.36 + 262(\text{Canopy Height})$$

The proportion of PV by topographic position group ranged from <0.01 at summits to 0.12 at toeslopes (Table 2). Average (± standard error) green leaf area index calculated using PV mass (Phillips et al. 2012) was 0.09 (0.02) for all 72 sites.

Partial least square regression (PLSR)

The $R^2$ value (based on the PRESS statistic for PLSR using leave-one-out cross-validation) was plotted against the number of factors used (Figure 1). The number of factors that maximized this statistic was five for all canopy variables except for Canopy N which was three. The fitted PLS regression for each Monte Carlo trial predicted observations of canopy responses for both the training and test datasets. The average $R^2$ and associated

### Table 2. Average (± standard error) for canopy mass, height, percent bare ground, and proportion of photosynthetically active vegetation (PV) collected 21 October 2010 for each topographic position group ($n = 24$).

<table>
<thead>
<tr>
<th>Canopy attribute</th>
<th>Summit</th>
<th>Midslope</th>
<th>Toeslope</th>
</tr>
</thead>
<tbody>
<tr>
<td>PV (kg ha$^{-1}$)</td>
<td>4 (4)</td>
<td>118 (50)</td>
<td>260 (59)</td>
</tr>
<tr>
<td>NPV (kg ha$^{-1}$)</td>
<td>650 (52)</td>
<td>1670 (130)</td>
<td>2040 (200)</td>
</tr>
<tr>
<td>TSC (kg ha$^{-1}$)</td>
<td>660 (54)</td>
<td>1790 (170)</td>
<td>2300 (200)</td>
</tr>
<tr>
<td>Canopy N (kg ha$^{-1}$)</td>
<td>53 (2)</td>
<td>51 (2)</td>
<td>45 (4)</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>4 (0.2)</td>
<td>6 (0.5)</td>
<td>8 (0.7)</td>
</tr>
<tr>
<td>Bare Ground (%)</td>
<td>28 (3)</td>
<td>14 (3)</td>
<td>6 (2)</td>
</tr>
<tr>
<td>Proportion PV</td>
<td>0.01 (0.01)</td>
<td>0.05 (0.02)</td>
<td>0.12 (0.03)</td>
</tr>
</tbody>
</table>
standard error across all Monte Carlo samples for both training and test samples is reported in Table 3. The overall fit statistics, based on fitting a PLSR to the entire dataset, is reported in Table 4. Standardized regression coefficients of the PLSR using the entire dataset for prediction are plotted in Figure 2.

Analyses of two components within each of the four banks of detectors collected at the AVIRIS sensor indicated the second component in the first SWIR region (from 1260 to 1880 nm) was an important predictor for all measured canopy attributes (Table 5). For NPV and TSC, the second component in the visible region also contributed to predictions. For %BG, the second component of the second SWIR region (from 1880 to 2500) and the second component in the near infrared region (from 660 to 1280) were also predictive of changes in %BG.

**Re-sampling model selection**

The models selected with the greatest frequencies for each variable are reported in Table 6. The model for each canopy variable is listed under the column with the header “vegetation indices”. For example, the model most frequently selected for predicting Canopy N was NDVI, which was selected 370 out of 1000 times among the Monte Carlo trials. For the case of PV, the MTVI2 was selected 379 times and both MTVI2 + MTVI2^2 were selected 272 times. We report the quadratic model with selection

---

**Table 3.** Average $R^2$ (standard error) for each variable based on PLS fit for both training and test sets for all 1000 Monte Carlo trials.

<table>
<thead>
<tr>
<th>Canopy attribute</th>
<th>Number of factors</th>
<th>$R^2_{train}$</th>
<th>$R^2_{test}$</th>
<th>RMSEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>PV</td>
<td>5</td>
<td>0.73 (0.07)</td>
<td>0.72 (0.15)</td>
<td>118 (28)</td>
</tr>
<tr>
<td>NPV</td>
<td>5</td>
<td>0.55 (0.06)</td>
<td>0.57 (0.12)</td>
<td>587 (127)</td>
</tr>
<tr>
<td>TSC</td>
<td>5</td>
<td>0.62 (0.06)</td>
<td>0.63 (0.11)</td>
<td>611 (123)</td>
</tr>
<tr>
<td>Canopy N</td>
<td>3</td>
<td>0.67 (0.05)</td>
<td>0.68 (0.10)</td>
<td>6 (1)</td>
</tr>
<tr>
<td>%BG</td>
<td>5</td>
<td>0.70 (0.10)</td>
<td>0.58 (0.20)</td>
<td>10 (3)</td>
</tr>
</tbody>
</table>

Note: Root mean square error of prediction (RMSEP) indicates prediction error in kg ha$^{-1}$ for PV, NPV, TSC, and Canopy N and as a percentage for %BG.

**Table 4.** $R^2$ by factor based on PLS fit for each variable using all data.

<table>
<thead>
<tr>
<th>Canopy attribute</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
<th>Factor 4</th>
<th>Factor 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>PV</td>
<td>58.8</td>
<td>60.1</td>
<td>69.6</td>
<td>71.0</td>
<td>73.7</td>
</tr>
<tr>
<td>NPV</td>
<td>33.0</td>
<td>35.6</td>
<td>45.3</td>
<td>49.8</td>
<td>54.0</td>
</tr>
<tr>
<td>TSC</td>
<td>46.8</td>
<td>48.2</td>
<td>54.2</td>
<td>57.3</td>
<td>60.8</td>
</tr>
<tr>
<td>Canopy N</td>
<td>63.2</td>
<td>64.8</td>
<td>66.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>%BG</td>
<td>30.3</td>
<td>49.1</td>
<td>55.2</td>
<td>62.3</td>
<td>66.5</td>
</tr>
</tbody>
</table>

Notes: $R^2$ for reflectance was based on jointly fitted model. Factors are orthogonal linear combinations of predictors that explain variance in both the predictor variables and the response variable(s).
frequency $651 = 379 + 272$ as optimum for prediction of PV. The model selected the SWIR-SR $\rho_{2128}/\rho_{1642}$ and REI for prediction of %BG. Coefficients of determinism for %BG were similar for both PLSR and VI model selection procedures ($R^2 = 0.58$).

Selected VIs and coefficients for each canopy attribute are as follows:

- Canopy N: $\text{Canopy N} (\text{kg ha}^{-1}) = (92.12 \times \text{NDVI}) - 17.32$ \hspace{1cm} (3)
- PV: $\text{PV} (\text{kg ha}^{-1}) = [(254.26 \times \text{MTVI2}) + (14759.57 \times \text{MTVI2}^2)] - 30.37$ \hspace{1cm} (4)

Table 5. Model selection, as performed by partitioning the data into the four spectral regions collected at the AVIRIS sensor, with two components each: VIS1, VIS2, NIR1, NIR2, SWIR11, SWIR12, SWIR21, and SWIR22.

<table>
<thead>
<tr>
<th>Canopy Attributes</th>
<th>Indices</th>
<th>Frequency</th>
<th>$R^2_{\text{train}}$</th>
<th>$R^2_{\text{test}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PV</td>
<td>SWIR12 + SWIR12$^2$</td>
<td>744</td>
<td>0.67(0.075)</td>
<td>0.66(0.150)</td>
</tr>
<tr>
<td>NPV</td>
<td>VIS2 + SWIR12</td>
<td>743</td>
<td>0.48(0.079)</td>
<td>0.49(0.149)</td>
</tr>
<tr>
<td>TSC</td>
<td>VIS2 + SWIR12</td>
<td>484</td>
<td>0.57(0.067)</td>
<td>0.57(0.127)</td>
</tr>
<tr>
<td>Canopy N</td>
<td>SWIR12</td>
<td>792</td>
<td>0.64(0.057)</td>
<td>0.66(0.118)</td>
</tr>
<tr>
<td>Bare ground</td>
<td>NIR2 + SWIR12 + SWIR22</td>
<td>495</td>
<td>0.56(0.071)</td>
<td>0.51(0.149)</td>
</tr>
</tbody>
</table>

Notes: Analysis was performed for the purpose of determining the spectral regions that most influenced PLSR predictive performance, $R^2$ (standard error).
NPV (kg ha\(^{-1}\)) = 1401 − \[2003 \times \text{SWIR} - \text{SR}\left(\frac{\rho_{2128}}{\rho_{1642}}\right)\] (5)

\[
TSC (\text{kg ha}^{-1}) = 8032 - \left[10148 \times \text{SWIR} - \text{SR}\left(\frac{\rho_{2128}}{\rho_{1642}}\right)\right]
\] (6)

\[
\%BG = -67.38 - \left[108.06\text{REI}\left(\frac{\rho_{705}}{\rho_{754}}\right)\right] + \left[259.93\text{SWIR} - \text{SR}\left(\frac{\rho_{2128}}{\rho_{1642}}\right)\right]
\] (7)

Results using conventional stepwise regression are listed in Table 7. Compared to re-sampling model selection, conventional stepwise selected more VIs for each variable and \(R^2\) values were higher. Root-mean-square-errors of prediction (RMSEPs), however, were similar (Table 6). For TSC, both re-sampling and conventional stepwise results were the same. Both statistics selected only SWIR-SR \(\left(\frac{\rho_{2128}}{\rho_{1642}}\right)\) for prediction of TSC, with an RMSEP of 692 kg ha\(^{-1}\). This RMSEP for TSC corresponds to a change in canopy height (Equation 2) ranging from approximately 1 to 4 cm and is near the average for summit TSC collected in the field (Table 2).

**Mapping modeled canopy attributes**

Equations 6 and 7 were used to map TSC and %BG using the October 2010 image for an 8 km\(^2\) area previously burned in April 2010. The fire line was mapped by the USFS directly after the burn and is outlined in blue (Figure 3). The map of %BG indicates a higher percentage of bare ground inside the fire line for the entire area burned 6 months earlier. The map of TSC, on the other hand, does not follow the same pattern inside the fire line. Data from the six field sites located inside the burned area indicated higher than average %BG and average TSC, which parallels modeled data in Figure 3. Further, topographic relationships between %BG and TSC are evident in Figure 4. East of the fire line, %BG follows a topographic gradient, with higher %BG at summits than at toeslopes. TSC is lower at summits and higher at toeslopes both inside and outside of the fire line. Finally, Figure 3 depicts a stocktank location and trails to the stocktank, where high percentages of bare ground were mapped. TSC estimates are also low at stocktank and trails, as expected for well-trodden areas.
### Table 7. Average $R^2$ (standard error) for both training and validation sets using conventional stepwise regression procedures.

<table>
<thead>
<tr>
<th>Canopy attributes</th>
<th>Vegetation indices</th>
<th>$R^2_{\text{data}}$</th>
<th>$R^2_{\text{train}}$ (0.07)</th>
<th>$R^2_{\text{test}}$ (0.17)</th>
<th>RMSEP (96)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PV</td>
<td>AR11 + AR11$^2$ + CAI + CAI$^2$ + MTVI2 + MTVI2$^2$ + NDVI</td>
<td>0.81</td>
<td>0.81</td>
<td>0.70</td>
<td>132 (96)</td>
</tr>
<tr>
<td>NPV</td>
<td>CAI + SWIR-SR ($\rho_{2128}/\rho_{1642}$)</td>
<td>0.49</td>
<td>0.50 (0.05)</td>
<td>0.48 (0.11)</td>
<td>664 (473)</td>
</tr>
<tr>
<td>TSC</td>
<td>SWIR-SR ($\rho_{2128}/\rho_{1642}$)</td>
<td>0.53</td>
<td>0.54 (0.05)</td>
<td>0.55 (0.09)</td>
<td>692 (499)</td>
</tr>
<tr>
<td>N</td>
<td>NDLI + NDVI</td>
<td>0.65</td>
<td>0.66 (0.05)</td>
<td>0.64 (0.10)</td>
<td>8 (6)</td>
</tr>
<tr>
<td>%BG</td>
<td>AR12 + NDNI + REI + AR12*REI + SWIR-SR ($\rho_{2128}/\rho_{1642}$) + SWIR-SR ($\rho_{2128}/\rho_{1642}$)$^2$</td>
<td>0.72</td>
<td>0.74 (0.08)</td>
<td>0.64 (0.18)</td>
<td>9 (7)</td>
</tr>
</tbody>
</table>

Note: See Table 1 for a description of the wavelengths comprising each VI.
Discussion

We found lower prediction error for measured canopy structural attributes when using PLSR, as compared to VI model selection analyses (Tables 3, 6, and 7). Similar to Cho...
et al. (2007), prediction performance was greater when all bands of the spectra were employed rather than specific regions of the spectrum. Unlike Cho et al. (2007), our herbaceous canopy was largely senescent, with average green leaf area index < 0.1. Cho

Figure 4. Modeled TSC and %BG draped on topography within the 8 km² area of interest depicted in Figure 3 using AVIRIS data collected 21 Oct 2010. The west side of this area (outlined in blue) was burned in April 2010.
et al. (2007) reported a coefficient of determinism for green standing crop mass during the growing season using HyMap imagery ($R^2 = 0.53$) that was comparable to our total standing crop mass post growing-season ($R^2 = 0.63$). Both studies suggest that PLSR may be more effective for estimating canopy mass in mixed-grass rangelands than VIs. Better predictive performance with PLSR may be attributed to the procedure finding surrogate spectral features for estimating target properties (Gomez et al. 2008). Percent bare ground, on the other hand, was predicted equally well with PLSR and with the combination of both the SWIR-SR ($\rho_{2128}/\rho_{1642}$) and the REI ($\rho_{705}/\rho_{754}$) vegetation indices (Equation 7).

The analyses performed on the four banks of detectors collected at the AVIRIS sensor pointed to the importance of the SWIR spectral region in autumn. All canopy attributes were responsive in the SWIR region, and a small number of components explained most of the variability in each region (Table 5). Others have also noted that selected subsets of hyperspectral data may include almost all of the information in the full spectrum (Fourny and Baret 1997; Darvishzadeh et al. 2011), and selected subsets likely vary seasonally. We found the second component in the SWIR region (1260–1880 nm) contained most of the relevant information (Table 1) at this time of the year for mixed-grass prairie. Fitzgerald and Ustin (1992) also noted correlations between California grassland biomass and AVIRIS reflectance data in the SWIR. Absorption in the SWIR is known to increase with lignin and cellulose (Kokaly et al. 2003; Daughtry et al. 2005), leading to lower values for SWIR-SR ($\rho_{2128}/\rho_{1642}$) as TSC increases. Higher values for the SWIR-SR ($\rho_{2128}/\rho_{1642}$), on the other hand, tend to indicate less TSC and more bare soil (Guerschman et al. 2009). The SWIR, rather than the VIS or NIR regions, was the most important spectral region for prediction of PV, according to PLSR results. The average proportion of PV comprising these canopies was extremely low (<0.1), so we did not see typical PV response in the VIS and NIR regions.

The re-sampling VI model selection procedure indicated the SWIR-SR ($\rho_{2128}/\rho_{1642}$) was the single VI most predictive of TSC and NPV, with generally greater values of TSC and NPV at lower values of SWIR-SR ($\rho_{2128}/\rho_{1642}$). Guerschmann et al. (2009) also found a similar ratio, SWIR-SR ($\rho_{2130}/\rho_{1640}$), could be used to separate the fraction of PV from NPV cover in tropical savannah when used in combination with NDVI. Unlike Guerchmann et al. (2009), NDVI was not selected as an important predictor for senescent TSC, NPV, or PV but was selected for prediction of Canopy N content. Similar to (Haboudane et al. 2004), MTVI2 and MTVI22 were identified as optimum predictors of PV instead of NDVI. Differences between Guerchmann et al. (2009) and this study illustrate how VI selection often depends on ecosystem properties and the time of year imagery are acquired (Cho et al. 2007; Fava et al. 2009).

Delineation of %BG from senescent vegetation is challenging because spectra in the SWIR are often used to estimate both materials. To compensate for this, some use NDVI in conjunction with VIs in the SWIR region to help separate %BG and TSC. Here, our re-sampling model selection procedure selected both SWIR and REI. REI reportedly responds to changes in biomass (Gianelle and Guastella 2007; Fava et al. 2009) and may be less affected by background (Cho et al. 2007). Figures 3 and 4 illustrate how model estimates for %BG compare with TSC. While %BG was clearly homogenously greater inside the fire line, modeled TSC was not homogenously lower. This was corroborated by field data, where the lack of understory in burned areas was clearly evident (higher %BG) but TSC was not anomalously low. Trails to the stocktank are also evident in Figure 3, where %BG is high and TSC is low. This was expected, since travel by livestock increases bare ground and suppresses vegetative growth. Patterns in %BG and TSC were not consistently spatially correlated, suggesting some separation of bare
ground from senescent vegetation. Application of both the SWIR-SR ($\rho_{2128}/\rho_{1642}$) and the REI ($\rho_{705}/\rho_{754}$) to the %BG model likely assisted with spectral separation between biomass and bare soil.

While statistical approaches are fast and easy to implement, the derived relationships are recognized as being sensor-specific and dependent on site and sampling conditions, and are expected to change in space and time (Baret and Guyot 1991; Colombo et al. 2003). The re-sampling VI selection method described here randomly partitioned the data into training and validation sets 1000 times and chose the model with the smallest average mean squared prediction error of the validation set. Alternatives, such as radiative transfer models, rely on the interactions between radiation and photosynthetically active vegetation in a canopy (Darvishzadeh et al. 2011), but radiation and reflectance relationships in senescent rangeland canopies have not been well characterized. For example, we found the slope of relationship between canopy height and TSC (Equation 2) was much steeper in October than in July (Phillips et al. 2012). Consequently, field data are required to train and test statistical models in the absence of accurate biophysical parameters for mixed PV-NPV canopies (Brinkmann, Patzelt, et al. 2011). Prediction of rangeland structural attributes with airborne hyperspectral data is improved when the full spectrum is analyzed in PLSR (Cho et al. 2007). When VIs analyses are needed, the re-sampling approach described here could provide more parsimonious model selection than conventional stepwise regression.

Conclusions

Predictive mapping of canopy structural and chemical attributes are needed for rangeland assessment and monitoring, particularly for lands managed to provide habitat for both wildlife and livestock grazing. Here, we narrowed our target to herbaceous vegetation for the purpose of predicting spatial variation in canopy attributes relevant to avian nesting habitat and rangeland forage quality assessment. We found lower RMSEPs when the full reflectance spectrum available from AVIRIS was modeled with PLSR for canopy structural and chemical properties than modeling specific regions of the spectrum with vegetation indices. The SWIR (1260–1880 nm) was the most informative spectral region for predicting PV, NPV, TSC, Canopy N, and %BG post growing-season. The novel re-sampling VI model selection procedure described here represents an alternative to conventional stepwise regression that avoids over-fitting and produces parsimonious VI models with similar prediction errors.

Acknowledgments

This work would not have been possible without the gracious cooperation of Dakota Prairie Grasslands managers (Phil Sjursen, Dan Svingen, and staff at the Lemmon, SD field office) and GRNG ranchers. The authors heartily thank Moffatt Ngugi, Justin Feld, Sarah Waldron, and Marla Striped-Face Collins for their outstanding teamwork and attention to data quality. Special thanks to Adam Tollefsrud for data quality control and mapping. The authors appreciate comments by Bruce Wylie and anonymous reviewers. Support provided by US Forest Service Support for this project was funded by the USFS Co-operative Agreement No. 15 5445-21310-001-07.

References


