The Influence of Vegetation Index and Spatial Resolution on a Two-Date Remote Sensing-Derived Relation to C4 Species Coverage

Andrew Davidson* and Ferenc Csillag*

Changes in composition of plant species are expected to accompany a warming climate. In the northern mixed grass prairie, such changes are predicted to take the form of shifts in the relative ground cover of C3 and C4 photosynthetic types. In this study, we explore the feasibility of using two-date remote sensing data as a potential tool for monitoring these shifts. Our approach is based on the well-described asynchronous seasonality of C3 and C4 species. We hypothesize that the ratios of early-season to late-season aboveground live biomass (B_{early}/B_{late}) will decrease as sites become more C4-dominated, and that if B_{early} and B_{late} can be reliably estimated using spectral data, it may be feasible to predict C4 species coverage (%C4) from commercially available satellite information. Using spectral and botanical measurements from three upland communities in the Canadian mixed grass prairie, we (a) examined the relationship between various spectral vegetation indices and aboveground live biomass, (b) investigated the nature of the relationship between remotely sensed estimates of B_{early}/B_{late} and %C4 at multiple sample resolutions (0.5 m, 2.5 m, 10 m, and 50 m), and (c) assessed whether these relationships were dependent on the vegetation index used to estimate biomass. We found a log-linear relationship between each spectral index and aboveground live biomass. Negative linear relationships were found between %C4 and remotely sensed B_{early}/B_{late} at all sampling resolutions. These relationships were strongest at sampling resolutions of 10 m and 50 m. The strengths and forms of relationships were found to be partially vegetation index-dependent. Stronger relationships between variants at coarser resolutions likely result from the smoothing of fine-scale variation in aboveground live biomass and C4 species coverage. Our results suggest that commercially available satellite data at resolutions of 10 m to 50 m (e.g., Landsat Thematic Mapper) may offer the potential for estimating coverage of C4 species and that the choice of vegetation index used to estimate biomass is relatively unimportant. However, we caution that for this technique to be operationally useful, statistical model performance must be strengthened and developed to provide both temporal and spatial generality. Further investigation is needed to examine the applicability of this approach to other growing seasons, community types, and grassland regions.

* Department of Geography, University of Toronto, Toronto
Address correspondence to A. Davidson, University of Toronto, Department of Geography, 100 St. George St., Toronto, ON M5S 3G3, Canada. E-mail: andy.davidson@utoronto.ca

INTRODUCTION

In recent years, there has been a growing interest in the impact of human activities on the biosphere. Of these impacts, the possible consequences of anthropogenically induced climate change have received much attention. While it is generally believed that climatic change will have substantial effects on the functioning of the biosphere, the exact nature of these effects is unclear. There is, however, a general expectation that changes in vegetation patterns will result, and that these changes will not be geographically uniform.

Zones of biological transition are expected to be especially responsive to a changing climate (Neilson, 1993). The northern extreme of the mixed grass prairie is an example of such a region. Here, plants that fix carbon through the C3 photosynthetic pathway are best adapted to the cooler temperatures typical of these latitudes, and contribute most to the floral diversity and net primary productivity of the region. Plants using the C4 photosynthetic pathway are less common and exist at their north-
ern limit, under conditions marginal for growth. These plants are less well adapted to the prevailing climatic conditions. These unfavorable conditions are thought to make C4 species extremely responsive to even subtle changes in climate. The implications of this are twofold. First, shifts in the abundance and distribution of C4 photosynthetic types may indicate a changing climate. Second, because C4 species are able to store more carbon per unit nitrogen than are C3 species, large-scale changes in the relative abundances of these photosynthetic types may have significant feedback effects on carbon sequestration and therefore future atmospheric CO₂ concentrations (Peat, 1997). However, the response of C4 species to climate change is not readily predictable. Few studies have explicitly considered the combined effects of varying multiple climatic variables on C4 production, and those that have attempted to do so have provided results that are somewhat contradictory (see Dahlman, 1993).

The increasing availability of remotely sensed data at various spatial and spectral resolutions offers the potential to monitor the biophysical characteristics of ecosystems at various landscape scales (Tieszen et al., 1997). This has been particularly well demonstrated by those who have focused their attentions on grasslands. Spectral observations over large tracts of grassland typically have been acquired using National Oceanic and Atmospheric Administration (NOAA) Advanced Very High Resolution Radiometer (AVHRR) data (Paruelo and Laenenroth, 1995; Tieszen et al., 1997). While these spatial resolutions (1–4 km) provide valuable information about climatic and aggregate anthropogenic forcings on vegetation dynamics (Henebry, 1993), they are often at a coarser spatial resolution than some applications require. Finer-resolution satellite observations have been derived from Landsat Multispectral Scanner (MSS) (Pickup et al., 1993) and Landsat Thematic Mapper (TM) sensors (Henebry, 1993) at resolutions of 50 m and 30 m, respectively. At even finer resolutions, information has been provided by airborne (Walthall and Middleton, 1992) and ground-based observations (Weiser et al., 1986; Goodin and Henebry, 1997), whose spatial resolutions are determined by sensor height and field of view (FOV).

One of the greatest challenges in the remote sensing of grasslands has been the reliable estimation of biophysical variables, such as aboveground biomass and productivity, from multisresolution observations. This is largely a consequence of the familiar “mixed pixel” problem, where factors other than the presence and amount of green vegetation (e.g., senescent vegetation, soil, shadow) combine to form composite spectra (see Asner, 1998). This spectral mixing often makes the discrimination of green vegetation difficult and has prompted the development of numerous spectral vegetation indices (VIs). VIs combine two or more spectral bands to enhance the vegetative signal while minimizing background effects. The most common of these indices utilize red and near-infrared canopy reflectances (or radiances) in the form of ratios [e.g., Normalized Difference Vegetation Index (NDVI)] or in linear combination e.g., Perpendicular Vegetation Index (PVI), while others e.g., Modified Soil-Adjusted Vegetation Index (MSAVI) are more complex and also require the derivation of soil correction factors (Chen, 1996). Although many of these VIs have been found to be well correlated with various biophysical parameters, including vegetation cover (Purredorj et al., 1998), aboveground biomass (Boutton and Tieszen, 1983), green leaf area (Baret and Guyot, 1991), photosynthetically active radiation (PAR) (Hatfield et al., 1984), and productivity (Box et al., 1989), some, such as the NDVI, have received more attention than others.

The principal challenge in applying remote sensing observations to monitor the C3/C4 vegetation dynamics of grasslands lies in inferring the relative abundances of C3 and C4 species in mixed canopies (Goodin and Henebry, 1997). This is because pixel reflectances, or reflectance-derived estimates of biomass, are influenced by the relative contributions of both photosynthetic types. We propose that the asynchronous seasonality of C3 and C4 types can be used to “unmix” remotely sensed estimates of total aboveground biomass into their contributing components, particularly the areal coverage of C4 species. Seasonality occurs because C3 species are generally less well adapted to conditions of high light (Tieszen, 1970), elevated temperature (Schuster and Monson, 1990), and limited soil moisture (Barnes et al., 1983) than are C4 species. These factors combine to produce seasonal cycles where C3 species (cool season plants) green up in early spring and are most active under the cooler conditions of spring and fall. In contrast, C4 species (warm season plants) green up later in the growing season and are more active under the hotter and drier summer months. As a result, peak annual biomass tends to occur earlier in the growing season for C3-dominated sites relative to that of C4-dominated sites (see Goodin and Henebry, 1997), a trend that should also be reflected in both early-season ($B_{early}$) and late-season aboveground live (green) biomass measurements ($B_{late}$). If this is generally the case, we may also expect the ratio of early-season aboveground live biomass to late-season aboveground live biomass ($B_{early}/B_{late}$) to decrease as sites become increasingly C4-dominated. Furthermore, if $B_{early}/B_{late}$ can be reliably estimated using some spectral VI (e.g., NDVI), it may then be feasible to predict the percentage of ground surface covered by C4 species ($\%C4$) over wide geographical regions from various remote sensing platforms.

However, to evaluate the potential of this remote sensing approach, two sampling issues must be explicitly
addressed. First, remotely sensed data collection and traditional ecological research normally operate at incongruous spatial resolutions. While remote sensing observations are usually taken at spatial resolutions ranging from 10 m Systeme Pour l'Observation de la Terre-Panchromatic (SPOT-P) to 4 km (AVHRR), biological research has tended to focus on plot level (<1 m) phenomena (Sellers et al., 1990). This disparity is one of the most important limiting factors in the reliable estimation of grassland ecosystem parameters from remotely sensed data and must be narrowed significantly. Second, many studies fail to consider the effects of scale (sample resolution) on the relationship between spatial reflectance and biophysical parameters, and in doing so, ultimately choose a scale of observation that is inappropriate for the task at hand. Relationships often vary with scale, and the way in which pattern and process vary across scales is often not well understood. The correlation between the state of an environmental variable and plant community composition may appear to be significant at some scales, but not at others (Lobo et al., 1998). To identify the most suitable scales of measurement, or to “scale up” plot level relationships to the scales needed for regional studies, changes in the relationship between spectral data and biophysical parameters must be explicitly investigated across a range of spatial scales.

This study explores the feasibility of using two-date remote sensing data as a potential tool for the prediction of C4 species coverage over a C3-dominated grassland. To achieve this and to address the above sampling issues directly, we collected C4 species cover information and spectrally derived estimates of \( B_{\text{early}}/B_{\text{late}} \) over plots of comparable ground resolution (approximately 0.5 m) from three field sites in the Canadian mixed grass prairie, then used a spatially nested sampling design to scale each property and the relationships to coarser observational scales (2.5 m, 10 m, and 50 m). Specifically, we seek to (1) compare the forms, strengths, and similarities of eight field-derived univariate models, each describing the statistical relationship between a spectral vegetation index and aboveground live biomass, and (2) examine the form of the functional relationship between C4 species coverage and remotely sensed estimates of \( B_{\text{early}}/B_{\text{late}} \) and describe how this relationship is affected by (a) the scale (sample resolution) of measurement and (b) the remotely sensed vegetation index utilized to estimate biomass. As with many field-based experiments, sampling limitations prevent us from validating empirical relations with independent observations derived from other grassland sites. For this reason, the cross-validation of such relations are restricted to the use of the bootstrap method. In this sense, our study follows similar approaches to estimating biophysical parameters using spectral indices that do not explicitly make use of separate training and validation data sets.

**METHODS**

**Study Region: Grasslands National Park, Saskatchewan, Canada**

Species composition, biophysical, and spectral data were collected during the 1995 growing season at three sites within Grasslands National Park (GNP), Saskatchewan, Canada (49°15’ N, 107°0’ W). GNP was established in 1988 to preserve a representative portion of the Canadian mixed grass prairie ecosystem. The climate of the region is semiarid; winters are long, cold, and dry while summers are short, hot, and comparatively wet (Environment Canada, 1998). Mean daily temperature ranges from 15°C below zero in January to 20°C in July. Total annual precipitation averages 325 mm. Approximately one-third of this total (110 mm) falls as snow in the winter, while the remainder (215 mm) falls as rain throughout the rest of the year, mostly during heavy (but infrequent) summer thunderstorms (165 mm). The growing season in the park is short, averaging 170 days between killing frosts, but low moisture availability often further reduces its effective length (Loveridge and Potyondi, 1983). However, because the average daily bright sunshine is high during the growing season (2400 h annually) (Environment Canada, 1998), growth rates can be great when sufficient moisture is available. The study area is dominated by grasses, of which *Stipa comata* (C3), *Agropyron* spp. (C3), and *Bouteloua gracilis* (C4) are the most common. Sage, mosses, lichens, and cacti make up a significant part of the plant community in drier locations.

**Location and Description of Sample Sites**

Potential sampling locations for our 1995 field campaign were selected based on a 1993 vegetation inventory and classification that was conducted on GNP lands (Michalsky and Ellis, 1994). Using a Geographical Information System (GIS), we identified sites that were representative of native prairie vegetation and the surrounding Park area and were undisturbed. The only vegetation class that met the above requirements was the “Upland Grassland” type, which is characterized by “a dominant cover of grasses or sedges, with shrub cover being low or absent” (Michalsky and Ellis, 1994).

We located three sampling sites in separate regions of Upland Grassland at GNP. These sites were situated along a soil moisture gradient and were chosen to represent areas of low and high total seasonal productivity. The “wet” and “intermediate” sites were located in topographic depressions, providing available soil moisture for plant growth at each site, even during hot summer months (Davidson, unpublished data). In contrast, the “dry” site was located on an exposed plateau, where extreme water limitation was observed during the hottest periods of the growing season. While the dominant plant types at each sampling location were very similar, sites...
varied in the relative abundance of C3 and C4 species, total species richness, and the abundance of rarer species.

**Sampling Design**

Estimating $B_{\text{early}}$ and $B_{\text{late}}$ from Spectral Reflectance Information

Plot-level relationships between aboveground live biomass and each spectral vegetation index were derived using a stratified random sampling approach. At each sample site, we established eight “calibration” plots (5 m by 5 m), in which four circular subplots (0.5 m diameter) were randomly located. Calibration plots were situated so that areas of low (5 g m$^{-2}$) to high (135 g m$^{-2}$) aboveground live biomass were adequately represented within the sampling framework. At various times during the 1995 growing season (early May to mid-August), simultaneous spectral and biophysical data (litter and live, dead, and total standing biomass) were collected from a randomly chosen subplot within each calibration plot. Once a subplot was sampled, it was precluded from being sampled again (i.e., sampling was random without replacement).

Spectrally reflected radiation in Landsat TM bands 3 (0.63–0.69 µm) and 4 (0.76–0.90 µm) was measured over each subplot using an Exotech Model 100BX radiometer, Exotech Incorporated, Gaithersburg, Maryland, USA. The radiometer has a 15° FOV and was mounted at a height of approximately 1.5 m, giving a spatial sample resolution of approximately 0.5 m (nadir view). All canopy measurements were taken on cloudless days within 2 hours of solar noon and frequently referenced to a barium sulfate panel (once every 7 to 12 minutes). Reflectance factors were then computed from a ratio of canopy radiance to that of the reference panel (after Weiser et al., 1986). Zenith angles ranged from 39.6° to 30.9° on May 15th and from 38.4° to 30.1° on July 28th. We used the mean of three separate reflectance measurements as a representative measure of canopy reflectance at each subplot. These means were then transformed into the suite of VIs described in Table 1 (Richardson and Everitt, 1992; Jordan, 1969; Rouse et al., 1973; Roujean and Breon, 1995; Huete, 1988; Qi et al., 1994; Crippen, 1990; Chen, 1996).

The harvesting of aboveground biomass from each subplot occurred immediately after the collection of spectral data. All vegetation within a subplot was clipped and separated into litter, live, and dead components. All samples were dried in an oven at 60°C for 48 hours and weighed. These weights (in g) were multiplied by a conversion factor (5.09) to provide biomass estimates in g m$^{-2}$. Log-linear regressions were then used to characterize the relationship between each VI and aboveground live biomass (see description later).

**Scaling Plot-Level Estimates of %C4 and Remotely Sensed $B_{\text{early}}/B_{\text{late}}$ to Coarser Resolutions**

Estimates of %C4 and $B_{\text{early}}/B_{\text{late}}$ were derived directly from individual plot observations ($n=72$ at each sample site). At this resolution, we used a square quadrat frame (0.5 m length, marked at 0.1 m intervals) to visually estimate the proportional areal coverage of individual species present in each plot. Following the sampling methodology of Michalsky and Ellis (1994), species cover was

<table>
<thead>
<tr>
<th>Index</th>
<th>Formula</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>DVI</td>
<td>$\frac{\text{NIR} - \text{R}}{\text{R}}$</td>
<td>Richardson and Everitt (1992)</td>
</tr>
<tr>
<td>RVI</td>
<td>$\frac{\text{NIR}}{\text{R}}$</td>
<td>Jordan (1969)</td>
</tr>
<tr>
<td>NDVI</td>
<td>$\frac{\text{NIR} - \text{R}}{\text{NIR} + \text{R}}$</td>
<td>Rousse et al. (1973)</td>
</tr>
<tr>
<td>RDVI</td>
<td>$\frac{\text{NIR} - \text{R}}{\text{NIR} + \text{R}}$</td>
<td>Roujean and Breon (1995)</td>
</tr>
<tr>
<td>SAVI</td>
<td>$\frac{(\text{NIR} - \text{R})(1+L)}{(\text{NIR} + \text{R} + L)}$, where $L=0.5$</td>
<td>Huete (1988)</td>
</tr>
<tr>
<td>MSAV1</td>
<td>$\frac{2\text{NIR}+1-\sqrt{(2\text{NIR}+1)^2-8(\text{NIR}-\text{R})}}{2}$</td>
<td>Qi et al. (1994)</td>
</tr>
<tr>
<td>IPVI</td>
<td>$\frac{\text{NIR}}{\text{NIR} + \text{R}}$</td>
<td>Crippen (1990)</td>
</tr>
<tr>
<td>MSR</td>
<td>$\frac{\text{NIR}}{\text{R}}$, $\frac{\sqrt{\text{NIR}}}{\text{R}}$</td>
<td>Chen (1996)</td>
</tr>
</tbody>
</table>

NIR and R correspond to reflectances in the near-infrared and red wavelengths, respectively.
estimated to the nearest 5% for cover values ranging from 10% to 90% and to the nearest 1% for cover values less than 10% and greater than 90%. Individual C4 species cover measurements were then summed to provide an estimate of total plot C4 species coverage (i.e., %C4).

The collection of plot-resolution spectral data followed the sampling protocols outlined previously. Our a priori identification of sampling dates for remotely sensed measurements of \( B_{cab} \) and \( B_{lab} \) was based on the knowledge of “typical” seasonal trends in plant production across the GNP area. To capture early season and late season aboveground live biomass, plot spectral measurements were taken on 15 May 1995 and 28 July 1995, respectively. For both measurement dates, these reflectances were used to calculate the vegetation indices described in Table 1, from which various estimates of aboveground live biomass (and subsequently \( B_{cab}/B_{lab} \)) were derived using empirical relationships.

We used an unbalanced nested sampling design to scale these plot-level (0.5 m resolution) measurements of %C4 and remotely sensed \( B_{cab}/B_{lab} \) “up” to coarser spatial resolutions of 2.5 m, 10 m, and 50 m (Figs. 1a and 1b). Spatially nested sampling applies classical analysis of variance to a hierarchical design to partition the total variance of a study variable into varying components, each corresponding to a different spatial resolution (see Webster, 1979). In effect, nested sampling provides a good measure of landscape “patchiness” at varying spatial scales. This approach is based on the conditions that the spatial components of the population can be divided into distinct levels and that observations are viewed as the result of the nested contributions of these levels (Belle-heumer and Legendre, 1998).

At each sample site, individual plots were spatially nested as four levels inside a 100 m by 100 m grid (Fig. 1a). Each level of nesting corresponded to a specific sampling resolution (level IV=0.5 m; level III=2.5 m; level II=10 m; level I=50 m). Plot-level measurements of %C4, \( B_{cab} \), and \( B_{lab} \) were then spatially averaged (Fig. 1b) to provide estimates of %C4 and \( B_{cab}/B_{lab} \) at “upper” levels (i.e., coarser resolutions) of the hierarchy (three-plot averages at level III; seven-plot averages at level II; eight-plot averages at level I). Thus, during the aggregation process, the original data were reduced to a smaller number of data points, which in turn represented a greater area than the original units. The aggregation of plots, resulted in \( n=18 \) estimates of each parameter at level III (2.5 m), \( n=9 \) estimates at level II (10 m), and \( n=9 \) estimates at level I (50 m). At the wettest site, however, the presence of cloud during the collection of spectral data meant that only 64 plot measurements were usable at level IV (\( n=64 \)). This resulted in \( n=16 \) estimates of each parameter at level III, \( n=8 \) estimates at level II, and \( n=8 \) estimates at level I for this site. Linear least squares regressions were used to examine the forms and strengths of the relationship between %C4 and \( B_{cab}/B_{lab} \) for each vegetation index and at each sampling resolution.

**Statistical Methods**

All statistical analyses were conducted using the S-PLUS statistical package (Version 4.5, MathSoft Incorporated, Seattle, Washington, USA, 1998), and \( p<0.05 \) was used to determine significance in all tests. Simple linear regression models were used (1) to characterize the relationship between each spectral vegetation index and aboveground live biomass, and (2) to assess the scale-dependency of the functional relationship between %C4 and remotely sensed estimates of \( B_{cab}/B_{lab} \). Regression statistics for analysis (1) were generated using \( n=92 \) pairs of observations. Because we wished to investigate general trends and because sample sizes are relatively small at coarser resolutions, regression statistics for analysis (2) were generated for all sites combined (\( n=208 \) at 0.5 m; \( n=52 \) at 2.5 m; \( n=26 \) at 10 m; \( n=26 \) at 50 m). We tested all variates for normality prior to their use. Much of the data were heteroscedastic and so were transformed to meet the distributional assumptions of the regression models. In analysis (1), each vegetation index was log transformed. In analysis (2), estimates of \( B_{cab}/B_{lab} \) and %C4 were square root transformed. In all
cases, proper diagnostic checks were performed on model residuals. All assumptions regarding residuals were met, including those concerning spatial autocorrelation.

Because of sampling limitations, we were unable to collect additional independent data for the validation of regression-derived relationships, nor to provide a sufficiently large sample [especially at coarser resolutions for analysis (2)] to facilitate validation through data splitting methods (i.e., the subsetting of original data into independent training and validation data sets). For this reason, cross-validations of the relationships derived from analyses (1) and (2) were performed using Monte Carlo bootstrap resampling. Bootstrap resampling is a computer-based method for assigning measures of accuracy to statistical estimates (see Efron and Tibshirani, 1993). Using the bootstrap approach, 1,000 new samples, each of the same size of the observed data [i.e., n = 92 for analysis (1); n = 208 at 0.5 m; n = 52 at 2.5 m; n = 26 at 10 m and 50 m for analysis (2)] were drawn with replacement from the observed data. Regression coefficients (slopes, intercepts) were first derived using the observed data and then recalculated for each new set of observations, giving a bootstrap distribution for each statistic. These resampled estimates were then used to estimate the bias of the statistic derived from the original observations.

We tested the effects of vegetation index on estimates of $B_{\text{early}}, B_{\text{late}}, $ and $B_{\text{early}}/B_{\text{late}}$ at each sampling resolution using single-factor analysis of variance (ANOVA). Vegetation index was treated as a fixed factor. Analysis of variance is robust to violations of the assumptions of normality and homogeneity of variance as long as sample sizes are nearly equal (Zar, 1984), a condition that held true for our study. Where a significant effect of vegetation index was detected, Tukey’s pair-wise comparisons were used to contrast mean values for significant differences.

RESULTS

Relationships between Vegetation Indices and Aboveground Live Biomass

The relationship between each vegetation index and aboveground live biomass was found to be logarithmic in nature (Fig. 2a). These relationships were described statistically using log-linear regressions of each log-transformed index (independent variable) on aboveground live biomass (dependent variable) (Fig. 2b). In all cases, estimates of slope ($b$) and intercept ($c$) were highly significant ($p = 0.0000$). The variation in aboveground live biomass was best explained using the NDVI, Soil-Adjusted Vegetation Index (SAVI), Second Modified Soil-Adjusted Vegetation Index (MSAVI2), and Modified Simple Ratio (MSR) ($r^2 = 0.64$; residual standard error (RSE) = 18.9) and least well explained by the Difference Vegetation Index (DVI) ($r^2 = 0.51$; RSE = 22.0). Monte Carlo bootstrap resampling verifies that linear least squares regression provides unbiased estimates of slope and intercept for each univariate relationship. In all cases, bootstrapped estimates (and their associated RSEs) do not differ significantly from least squares regressions of the original data. Such results indicate that these statistical models are appropriate descriptors for the majority of our data.

Effects of Spectral Vegetation Index on Estimates of $B_{\text{early}}, B_{\text{late}}, $ and $B_{\text{early}}/B_{\text{late}}$

We used the previously described empirical relationships to estimate $B_{\text{early}}$ and $B_{\text{late}}$ at each scale of observation (Table 2) from which corresponding estimates of $B_{\text{early}}/B_{\text{late}}$ were subsequently derived. Early-season biomass estimates showed a nonsignificant response to vegetation index at each sampling resolution. In comparison, significant responses were found for estimates of $B_{\text{late}}$ at sampling resolutions of 0.5 m and 2.5 m and for estimates of $B_{\text{early}}/B_{\text{late}}$ at 0.5 m. In all cases, Tukey’s pair-wise comparisons showed that these significant responses were due to lesser average biomass estimates derived using the DVI (Table 2). This was especially evident for DVI-derived estimates of late-season biomass, which were statistically different to those generated from all but one [the Ratio Difference Vegetation Index (RDVI)] of the other indices at both 0.5-m and 2.5-m resolutions.
Table 2. Spectrally Derived Estimates of Early-Season ($B_{\text{early}}$) and Late-Season ($B_{\text{late}}$) Aboveground Live Biomass (mean±SD, g m⁻²) at 0.5-m, 2.5-m, 10-m, and 50-m Sampling Resolutions

<table>
<thead>
<tr>
<th>Resolution</th>
<th>Estimated $B_{\text{early}}$ (g m⁻²)</th>
<th>Estimated $B_{\text{late}}$ (g m⁻²)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.5 m</td>
<td>2.5 m</td>
</tr>
<tr>
<td>RVI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34.9±13.0</td>
<td>34.8±8.7</td>
<td>34.8±7.4</td>
</tr>
<tr>
<td>NDVI</td>
<td>33.7±17.2</td>
<td>33.6±11.4</td>
</tr>
<tr>
<td>IPVI</td>
<td>33.9±14.9</td>
<td>33.8±10.0</td>
</tr>
<tr>
<td>MSAVI</td>
<td>33.9±17.2</td>
<td>33.8±11.3</td>
</tr>
<tr>
<td>MSAVI2</td>
<td>33.7±17.7</td>
<td>33.7±11.7</td>
</tr>
<tr>
<td>DVI</td>
<td>33.0±14.5</td>
<td>31.9±8.9</td>
</tr>
<tr>
<td>MSR</td>
<td>33.8±16.3</td>
<td>33.8±10.9</td>
</tr>
<tr>
<td>RDI</td>
<td>32.5±15.8</td>
<td>31.8±9.8</td>
</tr>
</tbody>
</table>

Estimates at 0.5 m were derived directly from the empirical relationships presented in Fig. 2 and subsequently scaled to 2.5-m, 10-m, and 50-m resolutions using the nested sampling scheme illustrated in Fig. 1.

Relationships between Remotely Sensed $B_{\text{early}}$/$B_{\text{late}}$ and %C4

Slopes ($b$), intercepts ($c$), coefficients of determination ($r^2$), and RSEs were derived from the simple linear regressions of square root-transformed $B_{\text{early}}$/$B_{\text{late}}$ on square root-transformed %C4 at each scale of observation (Figs. 3, 4, and 5 and Table 3). Note that a direct interscale comparison of correlations and regression coefficients by means of RSEs and p values is difficult because sample size varies between resolutions (the exception being 10 m and 50 m). We still present these statistics, but they should be viewed as general indicators.

We found negative and statistically significant ($p=0.0000$) linear relationships between square root-transformed %C4 and all VI-derived measures of square root-transformed $B_{\text{early}}$/$B_{\text{late}}$ at each sampling resolution (Fig. 3, Table 3). The slopes and strengths of these relationships are strongly sample resolution-dependent. For each vegetation index, estimated slope coefficients decrease nonlinearly as sample resolution becomes coarser (Fig. 4). This indicates that linear regression model estimates of %C4 are more sensitive to changes in $B_{\text{early}}$/$B_{\text{late}}$ at coarser observational scales and that interscale differences in model sensitivity are maximal between sampling...
Figure 4. Sample resolution and vegetation index dependence of slope estimates derived from the simple linear regressions of remotely sensed estimates of $B_{\text{early}}/B_{\text{late}}$ (square root transformed) on %C4 (square root transformed).

resolutions of 0.5 m and 2.5 m and minimal between 10 m and 50 m [also see Fig. 5, where %C4 (as opposed to its square root transformation) is expressed graphically as a function of square root-transformed $B_{\text{early}}/B_{\text{late}}$]. The coefficients of determination generated from these regressions indicate that the strength of the above relationships is also sample resolution-dependent (Table 3). These coefficients show that the level of explanation of all statistical models increases nonlinearly as sample plots are aggregated to coarser resolutions. Weak linear relationships at 0.5-m resolutions ($r^2=0.08$ to 0.14) become stronger as samples are aggregated to 2.5 m ($r^2=0.33$ to 0.47) and 10 m ($r^2=0.58$ to 0.68). However, further aggregation to 50 m does not increase the strength of the relationship between variates ($r^2=0.57$ to 0.65).

The slopes and strengths of these relationships are also vegetation index-dependent (Figs. 4 and 5, Table 3). At each sampling resolution, the effects of vegetation index are most evident at the lowest extreme of the observed range in $B_{\text{early}}/B_{\text{late}}$ (Fig. 5). Here, these differences correspond to estimates of %C4 ranging from (a) 38% (MSAVI2) to 52% (RVI) at 0.5 m, (b) 56% (MSAVI2) to 72% (RVI) at 2.5 m, (c) 75% (MSAVI2) to 89% (RVI) at 10 m, and (d) 84% (MSAVI2) to 96% (DVI) at 50 m. At resolutions coarser than 2.5 m, the effects of vegetation index decrease to a minimum as observed $B_{\text{early}}/B_{\text{late}}$ increases to a maximum. Here, the interindex ranges in estimates of %C4 are approximately 3%. Figures 4 and 5 demonstrate that relative to each other, most indices “perform” similarly across observational scales. For example, at each sampling resolution, slope coefficients generated by regressions using the RVI are more negative than those generated using the Infrared Percentage Vegetation Index (IPVI), which in turn are more negative than those derived using the MSR, SAVI, NDVI, and MSAVI2 (i.e., $b_{\text{RVI}}<b_{\text{IPVI}}<b_{\text{MSR}}<b_{\text{SAVI}}, b_{\text{NDVI}}<b_{\text{MSAVI2}}$).

In contrast, slope estimates derived using the DVI and RDVI do not show consistent relative trends across resolutions. These results indicate that of all the indices analyzed, RVI-derived estimates of %C4 are the most sensitive to changes in $B_{\text{early}}/B_{\text{late}}$, while in comparison, MSAVI2-derived estimates are generally the least sensitive to such changes. Slope coefficients derived using the RDVI display the largest overall “jump” across sample resolutions, especially from 10 m to 50 m. In comparison, coefficients derived from the RVI display the least change across resolutions. The strength of these relationships is also vegetation index-dependent. The strongest relationship (i.e., highest $r^2$, lowest RSE) at each resolution was produced using the RVI, while the weakest relationship was produced using the MSAVI2 (Table 3).

Monte Carlo bootstrap resampling indicates that least squares regression provides unbiased estimates of slope (Fig. 6) and intercept for each of the previously
Figure 5. Regression lines showing the influence of sampling resolution and vegetation index on the predictability of %C4 (nontransformed). For the sake of clarity, where the responses of two or more vegetation indices are similar, trends are presented as a single line.

Described relationships. For all indices, and at all sample resolutions, bootstrapped estimates (and their associated SEs) do not differ significantly from least squares regressions of the original data. Again, such results illustrate that these regression models are appropriate descriptors of the bulk of our data.

DISCUSSION

Relationships between Vegetation Indices and Aboveground Live Biomass

Of the many studies that have attempted to estimate biophysical characteristics from remote sensing data, surprisingly few have concentrated their efforts on the prediction of aboveground live biomass. Those that have attempted to do so (e.g., Weiser et al., 1986) have demonstrated that the form of the relationship between parameters is highly site- and time-dependent and that a large number of samples is needed to adequately characterize the relationship (Friedl et al., 1994). We have shown that the relationship between each spectral vegetation index and aboveground live biomass is best described by a logarithmic (log.) curve. Our results differ from other grassland studies, such as those of Weiser et al. (1986) and Roy et al. (1991) (Fig. 7). In contrast to these studies, our results show a greater sensitivity of RVI to changes in aboveground live biomass for much of the observed range in both parameters (i.e., for sites where live biomass <95 g m⁻² and RVI<2.9). Above this threshold, our observed relationships are similar to those derived by Weiser et al. (1986) for sites where the previous year’s senescent material had been removed by burning.

We are unable to provide a physically based explanation for a logarithmic relationship between variates, and indeed we know of no other studies where similar results have been reported. For this reason, we also use linear least squares regression models to describe our original (i.e., non-log-transformed) data (Fig. 7). A comparison
relationships between Remotely Sensed $B_{\text{early}}/B_{\text{late}}$ and %C4

The Effects of Sample Resolution

The evidence presented indicates that the form and strength of the relationship between remotely sensed $B_{\text{early}}/B_{\text{late}}$ and %C4 is highly sample resolution-dependent, and that this relationship scales nonlinearly. Similar nonlinear scaling relationships have also been reported between NDVI and other terrain characteristics, such as elevation (Walsh et al., 1997; Bian, 1997) and soil depth (Lobo et al., 1998). However, these relationships were found to be strongest at much coarser resolutions than those identified in our study (e.g., >2,200 m for NDVI and elevation) (Bian, 1997).

The weak linear relationships between variates indicate that factors other than C4 species cover influence $B_{\text{early}}/B_{\text{late}}$ at the plot level. These factors likely include the diversity of plant species and plant functional types (i.e., grasses, shrubs, and forbs) present and the varying competitive abilities of these plant forms. The strong nonlinear relationship between plant functional diversity and productivity at the plot level has been well described (Tilman et al., 1997), as have the differing abilities of these functional types to compete for limiting resources.
such as water or nitrogen (Golluscio and Sala, 1993). Less well studied, however, are the constraints that spatial structure (pattern) imposes on the above relationships. The ability of pattern to constrain ecological processes has been discussed (Wu and Levin, 1994), and its influence on plant productivity at multiple scales deserves further attention (Czarán and Bartha, 1989).

At coarser resolutions (2.5 m, 10 m, 50 m), remotely sensed estimates of $B_{\text{early}}/B_{\text{late}}$ explain variations in %C4 more fully. $r^2$ values reach their maxima at a sample resolution of 10 m, indicating that any further aggregation does not “improve” the performance of the statistical model. We explain this trend as follows. Variations in $B_{\text{early}}/B_{\text{late}}$ and %C4 corresponding to processes operating at finer scales are filtered out, and like the previously cited studies of Bian (1997), Walsh et al. (1997), and Lobo et al. (1998), the aggregation of sample plots serves to reduce the variance of both variables while increasing the covariation between the two.

Effects of Vegetation Index

Although we found the relationship between $B_{\text{early}}/B_{\text{late}}$ and %C4 to be partially vegetation index-dependent, the lack of an independent data set limits our ability to identify the indices that provide the most accurate absolute estimates of %C4 at each sampling resolution. We acknowledge this limitation and the need for further detailed field investigation. However, we believe that our approach may still provide valuable information regarding C4 species distributions, even in the absence of further field validation. Our results clearly show that irrespective of the vegetation index used, remotely sensed estimates of $B_{\text{early}}/B_{\text{late}}$ have the potential to provide information on the relative C4 species coverages of various sample plots within a given dataset.

Figure 6. Comparisons of the regression-derived slope estimates (and their associated RSE) for our original data and those derived from Monte Carlo bootstrap resampling at each sampling resolution.
IMPLICATIONS OF RESULTS, LIMITATIONS OF APPROACH, FUTURE DIRECTIONS, AND CONCLUSIONS

The goal of this study has been to assess the feasibility of estimating the percentage of ground surface covered by C4 species using two-date remotely sensed data of varying spatial resolution. Our approach is based on the conditions that various remotely sensed indices can be used to estimate aboveground live biomass, and that the ratio of early season aboveground live biomass to late season aboveground live biomass \( (B_{\text{early}}/B_{\text{late}}) \) will be negatively correlated with the areal coverage of C4 species \( (%C4) \). This multiresolution approach distinguishes our study from others that have attempted to estimate the contributions of C3 and C4 species from remotely sensed signals (e.g., Goodin and Henebry, 1997) and allows us to better assess the operational usefulness of our technique.

Our results support the existence of a negative linear relationship between remotely sensed estimates of \( B_{\text{early}}/B_{\text{late}} \) and \%C4, especially at coarser sample resolutions (i.e., 10 m to 50 m) where these relationships are strongest. However, we have used fine-resolution field radiometry to derive these relationships, and for this technique to be operationally useful, satellite observations must be used. A number of existing satellite systems, such as SPOT-P (10 m) and Landsat TM (30 m), operate at comparable spatial resolutions to those where the relationships between \( B_{\text{early}}/B_{\text{late}} \) and \%C4 were found to be strongest. Although our nested field sampling scheme did not allow the scaling of these relations to the spatial resolution of sensors that are usually utilized in large-scale assessments (e.g., National Oceanic and Atmospheric Administration (NOAA) Advanced Very High Resolution Radiometer (AVHRR)), it is possible that the heterogeneity of such large pixels will mask the spectral-temporal features of the seasonal aboveground biomass curves that are crucial to the diagnosis of C4 abundance (Goodin and Henebry, 1997). Thus, of the commercial satellite information that is readily available to environmental scientists, Systeme Pour l’Observation de la Terre-High Resolution Visible (SPOT-HRV) and Landsat TM may be the most suitable data sources for this monitoring tool.

It is also clear from our results that this monitoring technique produces relatively similar results for all of the vegetation indices used in this study, especially at coarser sample resolutions. While at present we are unable to identify which indices best relate to \%C4 at each sampling resolution, our results suggest that computationally...
faster indices (e.g., RVI, NDVI) may be utilized in preference to those that are more cumbersome (e.g., MSAVI2), without significant losses or changes in statistical model performance. It is important to note that these conclusions only apply to the indices listed in Table 1. We have not attempted to assess the relative performances of indices that require the derivation of a soil line [e.g., PVI (Richardson and Wiegand, 1977), Transformed Soil-Adjusted Vegetation Index (TSAVI) (Baret and Guyot, 1991), MSAVI (Qi et al., 1994)] or those that utilize additional wavebands [e.g., GVI (Crist and Ci-cone, 1984)].

One of the most attractive features of our two-date approach is its minimal data requirements in comparison to other techniques. For example, while the discriminant mixture modeling approach of Goodin and Henebry (1997) required intensive field sampling (every 10 days during the 1995 growing season), our approach relies on the acquisition of aboveground biomass information from only two sampling dates. These requirements significantly reduce the amount of data that needs to be collected, processed, and analyzed.

However, our approach also has its limitations. Even though a strong linear relationship between remotely sensed estimates of $B_{esy}$/$B_{late}$ and $\%C4$ has been identified at coarser spatial resolutions, statistical model performance must be improved for this technique to be developed into a reliable and robust predictive tool that is equally applicable through time and space (i.e., across seasons and other community types or grassland regions). Temporal generality might be achieved through the identification of sampling dates where the relationship between $B_{esy}$/$B_{late}$ and $\%C4$ is strongest and the characterization of these dates in terms of objective and phenologically meaningful criteria (e.g., degree day; time after green up). This technique could then be tested using data from subsequent years and further adapted if necessary. Although the tallgrass prairie data (Konza, Kansas, USA) of Goodin and Henebry (1997) suggest that strong relationships between $B_{esy}$/$B_{late}$ and $\%C4$ may also exist in other grassland environments, total geographical generality will be difficult to achieve because seasonal trends in productivity (i.e., the timing of $B_{esy}$ and $B_{late}$) are often spatially variable. At the macro-scale, such variability is largely a result of factors relating to latitude (such as day length and temperature), while at local scales, variability is a result of local micro-climatic and topographic conditions. If the metrics used in this study can be linked to phenological stage (see previous discussion on temporal generality), it may be possible to apply our approach to other communities and, with a greater understanding of latitudinal grassland productivity patterns, other grassland biomes. However, additional independent field data collection from other plant communities at GNP as well as other geographical regions is needed to address these issues directly.

In conclusion, although there are several unresolved issues, our results do not contradict our original expectation, and they are encouraging for the prospect of monitoring C4 species coverage using commercially available two-date remotely sensed data (such as Landsat TM). While the methods presented here have the advantage of simplicity over other techniques, the operational applicability of our approach depends largely on the improvement of statistical model performance and the ability to provide a statistical model that is temporally and spatially robust. We are currently refining our approach in response to these issues.

We thank Grasslands National Park for their help and cooperation with this project. We thank H. Peat, L. Risling, and M. Wynn for their field assistance. We gratefully acknowledge the comments and suggestions made by R. Fernandez, F. Gadallah, D. Jackson, and S. Mitchell, and those of the anonymous reviewers. This research was supported by a Natural Sciences and Engineering Research Council (NSERC) of Canada grant to F. Csillag.

REFERENCES


