4 A TWO-DATE REMOTE SENSING RELATION TO A SIMPLE MEASURE OF PLANT COMMUNITY COMPOSITION

4.0 Citation

Reprinted from Remote Sensing of Environment, Vol. 75 No. 1., Davidson A and Csillag F. *The influence of vegetation index and spatial resolution on a two-date remote sensing derived relation to C4 species coverage*, pp 138-151, Copyright (2001), with permission from Elsevier Science.

4.1 Preface

In the previous chapter, we showed that nested sampling and nested ANOVA provided stable and accurate estimates of spatial structure of patchy grassland landscapes under a limited sampling budget. We also noted that the ability of nested ANOVA to test for significant differences in variance between hierarchical levels makes it a useful tool for scaling studies, where the scale-dependence of variables (or the relationship between them) is of interest. In this chapter, we implement a nested sampling approach at three northern prairie sites to investigate the scale-dependence of the relationship between a two-date remotely-sensed productivity measure (early-to-late season aboveground live biomass) and a simple measure of plant community composition (C4 species cover).

4.2 Abstract

Changes in plant species composition 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, then 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 *Bearly/Blate* and *%C4* at multiple sample resolutions (0.5m, 2.5m, 10m and 50m), 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 10m and 50m. The strengths and forms of relationships were found to be partially vegetation index-dependent. Stronger relationships between variates 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 10m to 50m (e.g. Landsat TM) may offer the potential for estimating C4 species coverage, and that the choice of vegetation index used to estimate biomass is 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.

4.3 Introduction

In recent years, there has been a growing interest in the impacts that human activities are having 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 northerly 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 – which are less well adapted to the prevailing climatic conditions – are less common, and exist at their northern limit and under conditions marginal for growth. 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 CO_2 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). Remote sensing has already contributed significantly to the inventory and mapping of resources, the quantification of environmental characteristics, the description of matter and energy flow within ecosystems, and the evaluation of spatial change (Quattrochi and Pelletier, 1991). This has been particularly well demonstrated by those who have focused their attentions on grasslands.

Spectral observations over large tracts of grassland have typically been acquired using NOAA AVHRR data (e.g. Burke et al., 1991; Tucker et al., 1991; Paruelo and Lauenroth, 1995; Goetz, 1997; Tieszen et al., 1997). While these spatial resolutions (1-4km) 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 MSS (e.g. Pickup et al., 1993; Mino et al., 1998) and Landsat TM sensors (e.g. Henebry, 1993; Henebry and Su, 1993; Mino et al., 1998; Todd et al., 1998) at resolutions of 50m and 30m, respectively. At even finer resolutions, information has been provided by airborne (e.g. Walthall and Middleton, 1992) and ground-based observations (e.g. Tucker, 1979; Girard, 1982; Asrar et al., 1986; Weiser et al., 1986; Turner et al., 1992; Pickup et al., 1993; Goodin and Henebry, 1997; Goodin and Henebry, 1998), 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 multi-resolution 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 Ross, 1981; Goel, 1988; Myeni et al., 1989; Curran et al., 1992; Fourty et al., 1996; Asner, 1998; Asner et al., 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 in order 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. NDVI) or in linear combination (e.g. PVI), while others (e.g. 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 (Hurcom and Harrison, 1998; Purevdorj et al., 1998), aboveground biomass (Boutton and Tieszen, 1983; Weiser et al., 1986), green leaf area, (Asrar et al., 1986; Weiser et al., 1986; Baret and Guyot, 1991), photosynthetically active radiation (PAR) (Hatfield et al., 1984; Weiser et al., 1986; Baret and Guyot, 1991) and productivity (Box et al., 1989; Running et al., 1989), some – such as the NDVI – have received more attention than others. The popularity of the NDVI lies in its computational simplicity, its larger dynamic range and greater sensitivity to changes in vegetation cover than other commonly used indices (Jasinski, 1990), and its use in the readily available NOAA AVHRR-derived data sets offered through the U.S. Geological Survey EROS Data Center in Sioux Falls, South Dakota.

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, Figure 4.1), 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 10m (SPOT-P) to 4km (AVHRR), biological research has tended to focus on plot level (<1m) 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 spectral 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 that way in which pattern and process vary across scales are 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). In order 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 tthe prediction of C4 species coverage over a C3-dominated grassland. To achieve this, and to address the above sampling issues directly, we collected cover information for C4 species and spectrally-derived estimates of B_{early}/B_{late} over plots of comparable ground resolution (approx. 0.5m) from three field sites in the Canadian mixed grass prairie, then used a spatially nested sampling design to scale each property and their relationships to coarser observational scales (2.5m, 10m and 50m). Specifically, we seek to (1) compare the forms, strengths and similarities of 8 field-derived univariate models, each describing the statistical relationship between a spectral vegetation index and aboveground live biomass, and (2)

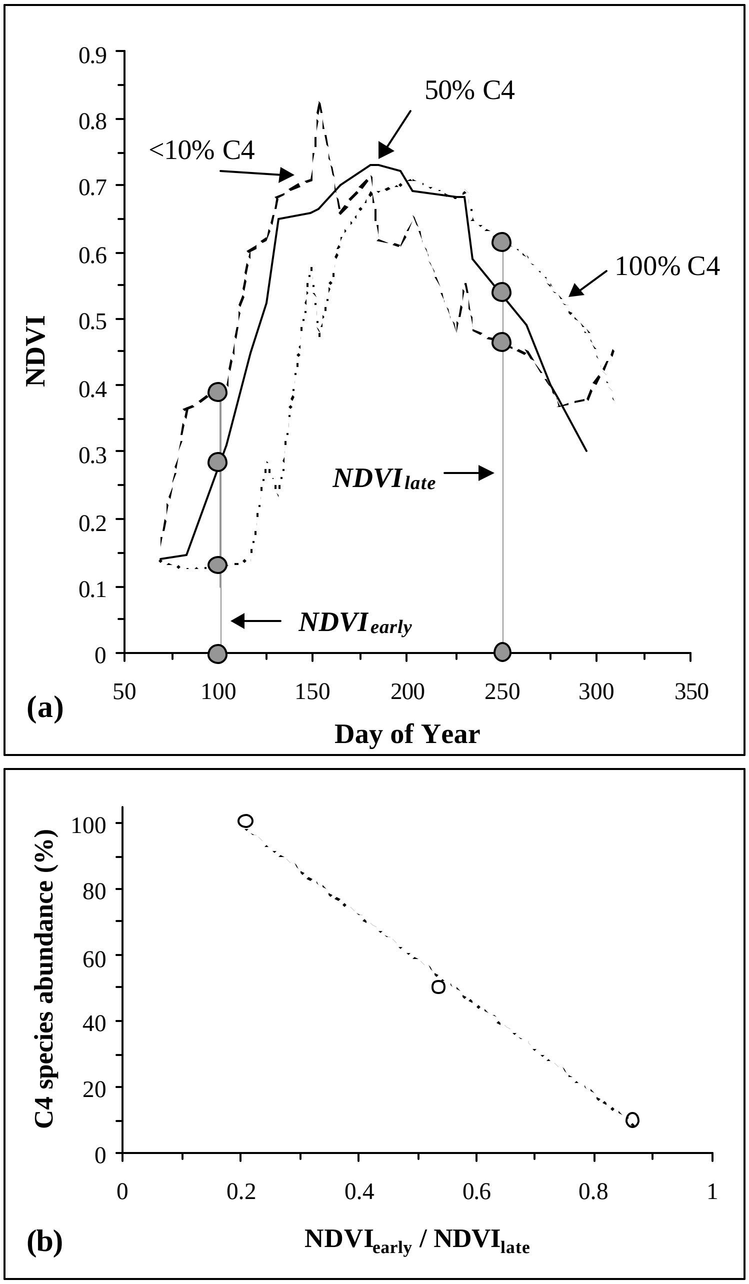


Figure 4.1. (a) Temporal NDVI trajectories derived from ground-based remote sensing information (1m resolution) for Tallgrass prairie plots containing differing contributions of C4 species (adapted from Goodin and Henebry, 1997). (b) Our further examination of these data supports our hypothesis that the ratio of early- to late-season biomass (as estimated using the NDVI) is negatively correlated with C4 species abundance for this particular prairie site.

examine the form of the functional relationship between C4 species coverage and remotely-sensed estimates of B_{early}/B_{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.

4.4 Methods

4.4.1 Location and description of sample sites

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) (see section 2.2.3 for a general description of the study site). Potential sampling locations 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 which 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, where there was sufficient soil moisture for plant growth, even during hot summer months (Davidson, Unpublished Data). In contrast, the “dry” site was located on an exposed plateau, where lower soil moisture content 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.

4.4.2 Sampling design

4.4.2.1 Estimating B_{early} and B_{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 8 “calibration” plots (5m by 5m), in which 4 circular sub-plots (0.5m diameter) were randomly located (Figure 4.2a). Calibration plots were situated so that areas of low (5 gm^{-2}) to high (135 gm^{-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; live, dead, and total standing biomass) were collected from a randomly chosen sub-plot within each calibration plot. Once a sub-plot 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 sub-plot using an Exotech Model 100BX radiometer (Exotech Inc, Gaithersburg, MD, USA). The radiometer has a 15° Field Of View (FOV) and was mounted at a height of approximately 1.5m, giving a spatial sample resolution of approximately 0.5m (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 to12 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 3 separate reflectance measurements as a representative measure of canopy reflectance at each sub-plot. These means were then transformed into the suite of VIs described in Table 4.1 .

The harvesting of aboveground biomass from each sub-plot occurred immediately after the collection of spectral data. All vegetation within a sub-plot was clipped and separated into litter, live and dead components. All samples were dried in a drying oven at 60°C for 48 hours and weighed. These

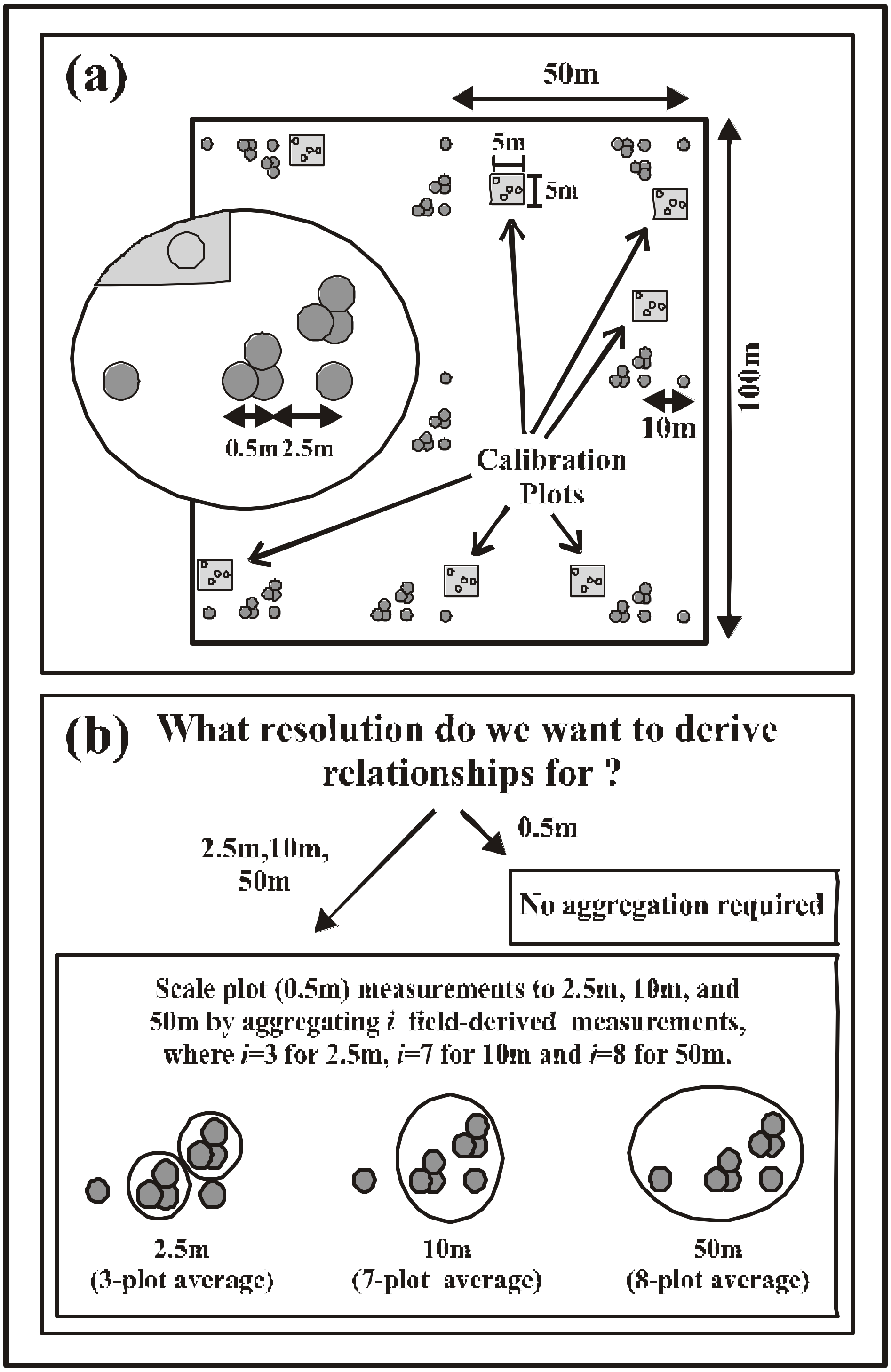


Figure 4.2. Field sampling scheme: (a) The location of calibration (*n*=8) and spatially nested (*n*=72) plots at each of our three sample sites. (b) The spatial averaging method through which plot-resolution (0.5m) information is scaled to coarser observational scales (2.5m, 10m, 50m).

weights (in g) were multiplied by a conversion factor (5.09) in order to provide biomass estimates in gm^{-2}. Log-linear regressions were then used to characterize the relationship between each VI and aboveground live biomass (see section 4.3.3).

4.4.2.2 Scaling plot-level estimates of %C4 and remotely-sensed B_{early}/B_{late} to coarser resolutions

Estimates of *%C4* and B_{early}/B_{late} were derived directly from individual plot observations (n=72 at each sample site). At this resolution, we used a square quadrat frame (0.5m length, marked at 0.1m 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 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 in section 4.3.2.1. Our *a priori* identification of sampling dates for remotely-sensed measurements of B_{early} and B_{late} was based on the knowledge of “typical” seasonal trends in plant production across the GNP area. In order to capture early season and late season aboveground live biomass, plot spectral measurements were taken on the 15th May 1995 and 28th July 1995, respectively. For both measurement dates, these reflectances were used to calculate the vegetation indices described in Table 4.1, from which various estimates of aboveground live biomass – and subsequently B_{early}/B_{late} – were derived using empirical relationships.

We used an unbalanced nested sampling design to scale these plot-level (0.5m resolution) measurements of *%C4* and remotely-sensed B_{early}/B_{late} “up” to coarser spatial resolutions of 2.5m, 10m and 50m (Figures 4.2a and 4.2b). Spatially nested sampling applies classical analysis of variance to a hierarchical design in order to partition the total variance of a study variable into varying components, each corresponding to a different spatial resolution (see Webster, 1979; Sokal and Rohlf, 1981; Ver Hoef et al., 1993; Belleheumer and Legendre, 1998). 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

INDEX	FORMULA	SOURCES
Difference Vegetation Index (DVI)	$NIR - R$	Richardson and Everitt (1992)
Ratio Vegetation Index (RVI)	$\frac{NIR}{R}$	Jordan (1969)
Normalized Difference Vegetation Index (NDVI)	$\frac{NIR - R}{NIR + R}$	Rouse et al. (1974)
Ratio Difference Vegetation Index (RDVI)	$\frac{NIR - R}{\sqrt{NIR + R}}$	Roujean and Breon (1995)
Soil-Adjusted Vegetation Index (SAVI)	$\frac{(NIR - R)(1 + L)}{(NIR + R + L)}$, where L=0.5	Huete (1988)
Second Modified Soil-Adjusted Vegetation Index (MSAVI2)	$\frac{2NIR + 1 - \sqrt{(2NIR + 1)^2 - 8(NIR - R)}}{2}$	Qi et al. (1994)
Infrared Percentage Vegetation Index (IPVI)	$\frac{NIR}{NIR + R}$	Crippen (1990)
Modified Simple Ratio (MSR)	$\frac{\frac{NIR}{R} - 1}{\sqrt{\frac{NIR}{R} + 1}}$	Chen (1996)

Table 4.1. Vegetation indices used to estimate early season (B_{early}) and late season biomass (B_{late}). *NIR* and *R* correspond to reflectances in the near-infrared and red wavelengths, respectively.

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 (Belleheumer and Legendre, 1998). Although our design contains only one sample difference between 10m and 50m resolutions, we use it because previous analyses (Davidson, Unpublished Data) indicate that it does not provide significantly different estimates of productivity to a nested sample design using 144 points.

At each sample site, individual plots were spatially nested as 4 levels inside a 100m by 100m grid (Figure 4.2a). Each level of nesting corresponded to a specific sampling resolution (level IV=0.5m; level III=2.5m; level II=10m; level I=50m). Plot-level measurements of *%C4*, B_{early} and B_{late} were then spatially averaged (Figure 4.2b) in order to provide estimates of *%C4* and B_{early}/B_{late} at “upper” levels (i.e. coarser resolutions) of the hierarchy (3-plot averages at level III; 7-plot averages at level II; 8-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 (2.5m), n=9 estimates at level II (10m) and n=9 estimates at level I (50m). 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_{early}/B_{late} for each vegetation index, and at each sampling resolution (see section 4.3.3).

4.4.3 Statistical methods

All statistical analyses were conducted using the S-PLUS statistical package (Version 4.5, MathSoft Inc., 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_{early}/B_{late}. 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 0.5m; n=52 at 2.5m; n=26 at 10m, n=26 at 50m). 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 $\log_e$-transformed. In analysis (2), estimates of B_{early}/B_{late} 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 provide a sufficiently large sample (especially at coarser resolutions for analysis (2)) to facilitate validation through data splitting methods (i.e. the sub-setting 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, 1000 new samples, each of the same size of the observed data (i.e. n=92 for analysis (1); n=208 at 0.5m, n=52 at 2.5m, n=26 at 10m and 50m 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_{early}, B_{late} and B_{early}/B_{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 homogeneous 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 pairwise comparisons were used to contrast mean values for significant differences.

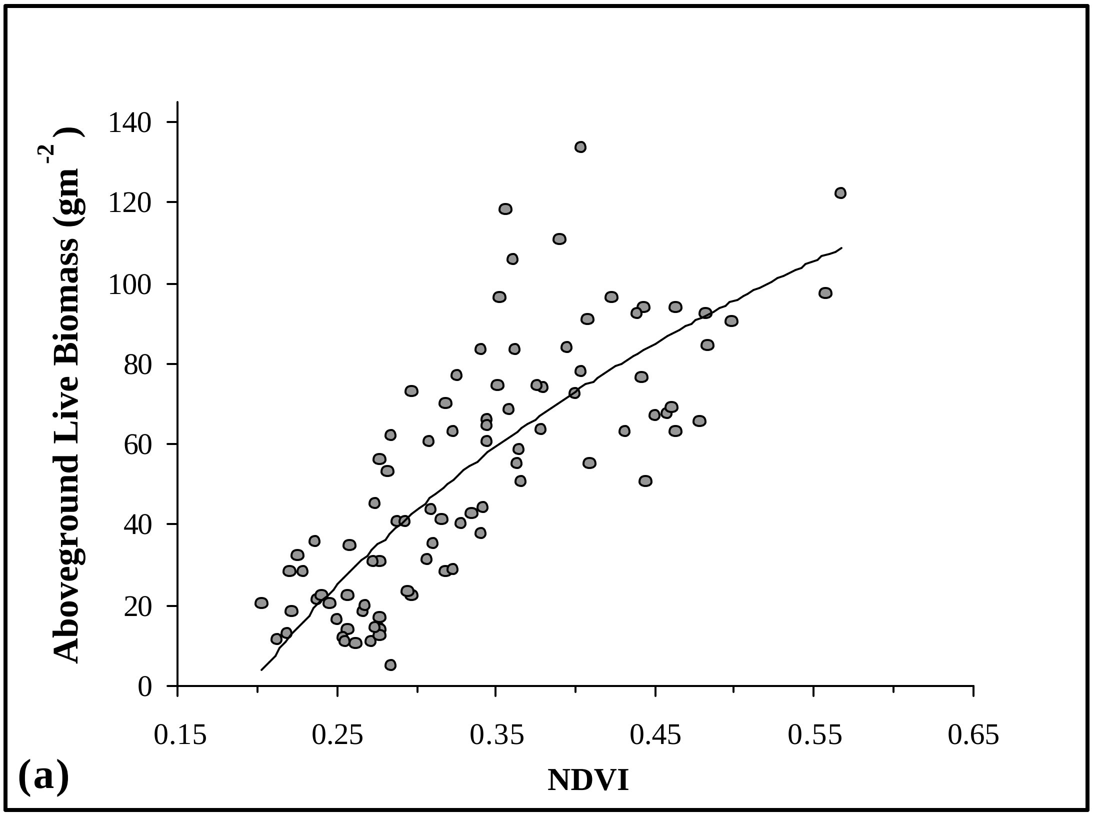
4.5 Results

4.5.1 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 (Figure 4.3a). These relationships were described statistically using log-linear regressions of each $\log_e$-transformed index (independent variable) on aboveground live biomass (dependent variable) (Figure 4.3b). 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, SAVI, MSAVI2 and MSR (r^2=0.64; *RSE*=18.9), and least-well explained by the 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 standard errors) 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.

4.5.2 The effects of vegetation index on estimates of B_{early}, B_{late} and B_{early}/B_{late}

We used the previously described empirical relationships to estimate B_{early} and B_{late} at each scale of observation (Table 4.2), from which corresponding estimates of B_{early}/B_{late} were subsequently derived. Early-season biomass estimates showed a non-significant response to vegetation index at each sampling resolution (Table 4.3a). In comparison, significant responses were found for estimates of B_{late} at sampling resolutions of 0.5m and 2.5m, and for estimates of B_{early}/B_{late} at 0.5m. In all cases, Tukey's pairwise comparisons showed that these significant responses were due to lesser average biomass estimates derived using the DVI (Table 4.3b). This was especially evident for DVI-derived estimates of late-season biomass, which were statistically different to those generated from all but one (the RDVI) of the other indices at both 0.5m and 2.5m resolutions.



b	*x*	*c*	*r*^{2}	*RSE*
104.7	$\log_e$ (NDVI)	168.9	0.64	18.9
127.1	$\log_e$ (RVI)	-37.9	0.61	19.7
408.8	$\log_e$ (IPVI)	217.4	0.63	19.2
104.6	$\log_e$ (SAVI)	128.0	0.64	18.9
137.7	$\log_e$ (MSAVI2)	151.7	0.64	18.9
81.0	$\log_e$ (DVI)	-136.3	0.51	22.0
82.1	$\log_e$ (MSR)	98.0	0.64	18.9
95.8	$\log_e$ (RDVI)	-5.8	0.59	20.0

(b)

Figure 4.3. (a) A graphical illustration of the logarithmic relationship between NDVI and aboveground live biomass (gm^{-2}), as derived from plot-level (0.5m) observations. (b) Coefficients generated from linear least squares regressions of each spectral vegetation index listed in Table 4.1 ($\log_e$-transformed) on aboveground live biomass (gm^{-2}).

4.5.3 Relationships between remotely-sensed B_{early}/B_{late} and %C4

Slopes (*b*), intercepts (*c*), coefficients of determination (r^2) and the residual standard errors (*RSE*) were derived from the simple linear regressions of square-root transformed B_{early}/B_{late} on square-root transformed *%C4* at each scale of observation (Figures 4.4, 4.5 and 4.6 and Table 4.4). Note that a direct inter-scale comparison of correlations and regression coefficients by means of residual standard errors and *P*-values is difficult because sample size varies between resolutions (the exception being 10m and 50m). 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_{early}/B_{late} at each sampling resolution (Table 4.4, Figure 4.4). 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 (Figure 4.5). This indicates that linear regression model estimates of *%C4* are more sensitive to changes in B_{early}/B_{late} at coarser observational scales, and that inter-scale differences in model sensitivity are maximal between sampling resolutions of 0.5m and 2.5m, and minimal between 10m and 50m (also see Figure 4.6, where *%C4* (as opposed to its square-root transformation) is expressed graphically as a function of square-root transformed B_{early}/B_{late}). The coefficients of determination generated from these regressions indicate that the strength of the above relationships is also sample resolution-dependent (Table 4.4). 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 0.5m resolutions (r^2=0.08 to 0.14) become stronger as samples are aggregated to 2.5m (r^2=0.33 to 0.47) and 10m (r^2=0.58 to 0.68). However, further aggregation to 50m 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 (Figures 4.5 and 4.6, Table 4.4). At each sampling resolution, the effects of vegetation index are most evident at the lowest extreme of the observed range in B_{early}/B_{late} (Figure 4.6). Here, these differences correspond to estimates of *%C4* ranging from (a) 38% (MSAVI2) to 52% (RVI) at 0.5m, (b) 56% (MSAVI2)

		Estimated	**B_{early} (gm^{-2})**				**Estimated**	**B_{late} (gm^{-2})**	
Resolution	**0.5m**	**2.5m**	**10m**	**50m**		**0.5m**	**2.5m**	**10m**	**50m**
RVI	34.9 ± 13.0	34.8 ± 8.7	34.8 ± 7.4	34.9 ± 7.0		70.1 ± 23.8	72.1 ± 22.5	70.8 ± 21.7	70.1 ± 20.8
NDVI	33.7 ± 17.2	33.6 ± 11.4	33.6 ± 9.7	33.7 ± 9.1		70.3 ± 19.8	72.0 ± 18.1	70.8 ± 17.7	70.3 ± 17.1
IPVI	33.9 ± 14.9	33.8 ± 10.0	33.8 ± 8.5	33.9 ± 8.0		70.2 ± 22.2	72.1 ± 20.6	70.8 ± 20.1	70.2 ± 19.3
MSAVI	33.9 ± 17.2	33.8 ± 11.3	33.8 ± 9.7	33.9 ± 9.1		70.4 ± 19.8	72.2 ± 18.1	71.0 ± 17.7	70.4 ± 17.0
MSAVI2	33.7 ± 17.7	33.7 ± 11.7	33.7 ± 10.0	33.7 ± 9.4		70.2 ± 19.0	71.9 ± 17.2	70.7 ± 16.8	70.2 ± 16.2
DVI	33.0 ± 14.5	31.9 ± 8.5	32.3 ± 7.0	33.0 ± 5.9		59.3 ± 17.4	60.4 ± 14.7	60.3 ± 14.2	59.3 ± 13.6
MSR	33.8 ± 16.3	33.8 ± 10.9	33.8 ± 9.2	33.8 ± 8.7		70.4 ± 21.0	72.2 ± 19.4	70.1 ± 18.8	70.4 ± 18.2
RDVI	32.5 ± 15.8	31.8 ± 9.8	32.0 ± 8.3	32.5 ± 7.3		64.8 ± 18.8	66.2 ±16.8	65.6 ± 16.4	64.8 ± 15.7

Table 4.2. Spectrally-derived estimates of early-season (B_{early}) and late-season (B_{late}) aboveground live biomass (mean ± SD, gm^{-2}) at 0.5m, 2.5m, 10m and 50m sampling resolutions. Estimates at 0.5m were derived directly from the empirical relationships presented in Figure 4.4, and subsequently scaled to 2.5m, 10m and 50m resolutions using the nested sampling scheme illustrated in Figure 4.3.

(a)		**Sampling resolution**		
	0.5m	**2.5m**	**10m**	**50m**
B_{early}	NS	NS	NS	NS
B_{late}	***	**	NS	NS
B_{early}/B_{late}	*	NS	NS	NS

NS $P>0.05$; * $P<0.05$; ** $P<0.01$; *** $P<0.001$

(b)	B_{early}	B_{late}	B_{early}/B_{late}
0.5m	-	dvi-rvi dvi-ndvi dvi-ipvi dvi-msavi dvi-msavi2 dvi-msr	dvi-msavi2
2.5m	-	dvi-rvi dvi-ndvi dvi-ipvi dvi-msavi dvi-msavi2 dvi-msr	-
10m	-	-	-
50m	-	-	-

Table 4.3. (a) Summary of results of single-factor analysis of variance (ANOVA) for the effects of vegetation index on estimates of B_{early}, B_{late} and B_{early}/B_{late} at sampling resolutions of 0.5m, 2.5m, 10m and 50m. (b) Statistically significant comparisons that can be declared different by Tukey's method.

Res. (m)	Reg. Params.	Sqrt (B_{early} / B_{late})							
		DVI	RVI	NDVI	RDVI	IPVI	MSR	SAVI	MSAVI2
0.5	*b* , *c*	-2.3, 6.9	-3.9, 8.0	-2.2, 6.7	-2.2, 6.7	-3.0, 7.3	-2.4, 6.9	-2.2, 6.7	-2.0, 6.6
	r^2	0.09	0.14	0.08	0.08	0.11	0.09	0.08	0.07
	RSE	1.96	1.90	1.97	1.96	1.93	1.96	1.97	1.97
2.5	*b* , *c*	-6.5, 10.2	-7.4, 10.7	-5.6, 9.2	-5.9, 9.5	-6.5, 9.9	-5.6, 9.4	-5.7, 9.3	-5.4, 9.1
	r^2	0.44	0.47	0.35	0.40	0.41	0.37	0.35	0.33
	RSE	1.19	1.16	1.29	1.24	1.22	1.27	1.28	1.30
10	*b* , *c*	-8.4, 11.6	-9.5, 12.3	-8.5, 11.3	-8.4, 11.3	-9.1, 11.8	-8.8, 11.5	-8.5, 11.3	-8.4, 11.2
	r^2	0.63	0.68	0.60	0.62	0.65	0.62	0.60	0.58
	RSE	0.91	0.85	0.94	0.92	0.88	0.92	0.94	0.97
50	*b* , *c*	-9.7, 12.8	-10.2, 12.8	-9.3, 11.9	-9.7, 12.4	-9.9, 12.5	-9.6, 12.2	-9.4, 12.0	-9.2, 11.9
	r^2	0.63	0.65	0.58	0.64	0.63	0.60	0.59	0.57
	RSE	0.87	0.84	0.92	0.86	0.87	0.89	0.91	0.93

Table 4.4. Estimated slopes (*b*), intercepts (*c*), coefficients of determination (r^2) and the residual standard errors (*RSE*) for simple linear regressions of remotely-sensed B_{early}/B_{late} (square-root transformed) on *%C4* (square-root transformed) at 0.5m (*n*=206), 2.5m (*n*=52), 10m (*n*=26) and 50m (*n*=26) sample resolutions. All slope and intercept estimates are highly significant (*P* = 0.0000).

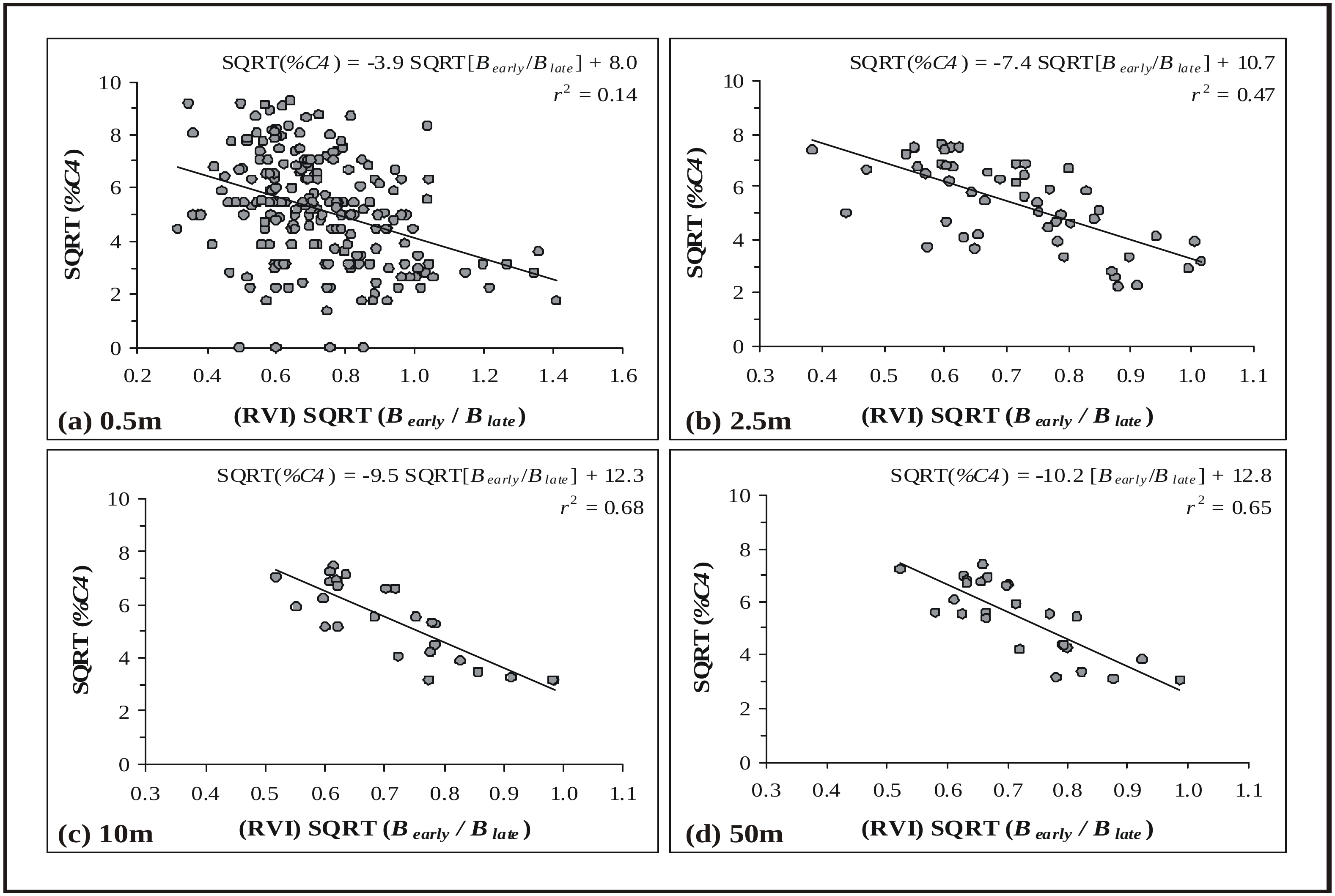


Figure 4.4. Linear least squares regressions of RVI-derived B_{early}/B_{late} (square-root transformed) on *%C4* (square-root transformed) at sampling resolutions of 0.5m (a), 2.5m (b), 10m (c) and 50m (d). Point data illustrate the scatter (*RSE*) around the regression line.

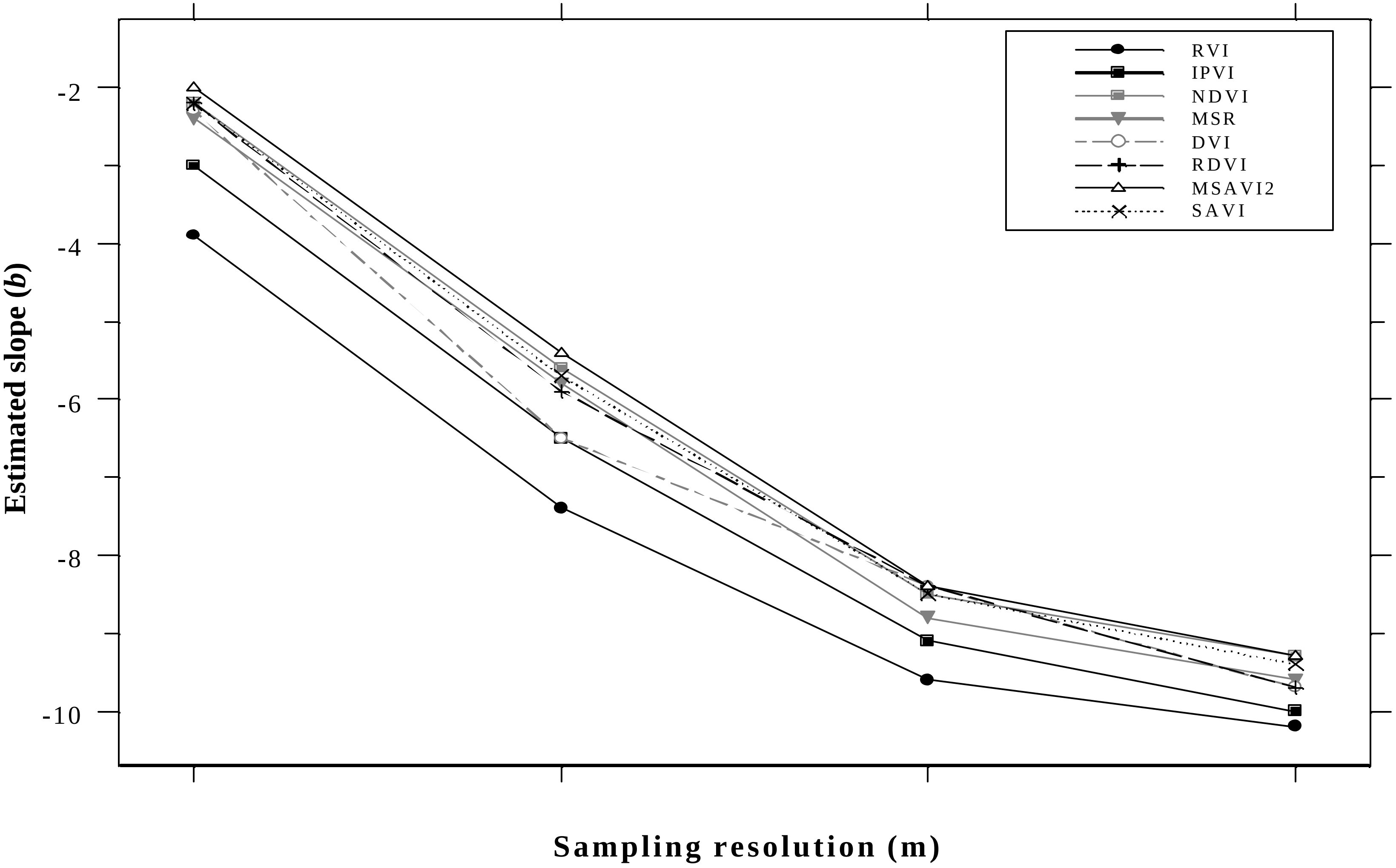


Figure 4.5. Sample resolution- and vegetation index-dependence of slope estimates derived from the simple linear regressions of remotely-sensed estimates of B_{early}/B_{late} (square-root transformed) on *%C4* (square-root transformed).

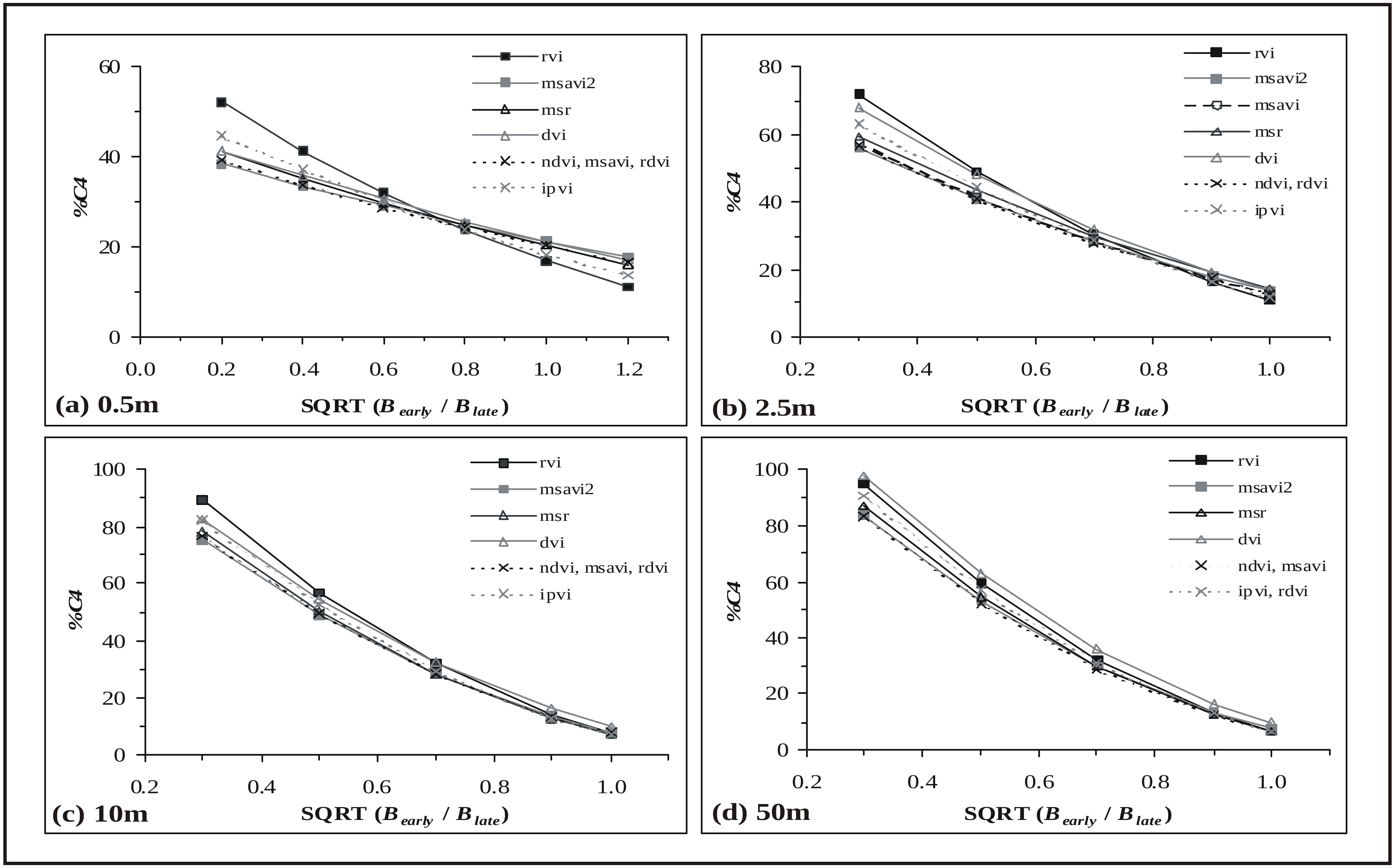


Figure 4.6. Regression lines showing the influence of sampling resolution and vegetation index on the predictability of *%C4* (non-transformed). For the sake of clarity, where the responses of two or more vegetation indices are similar, trends are presented as a single line.

to 72% (RVI) at 2.5m, (c) 75% (MSAVI2) to 89% (RVI) at 10m, and (d) 84% (MSAVI2) to 96% (DVI) at 50m. At resolutions coarser than 2.5m, the effects of vegetation index decrease to a minimum as observed Bearly/Blate increases to a maximum. Here, the inter-index ranges in estimates of %C4 are approximately 3%. Figures 4.5 and 4.6 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 IPVI which, in turn, are more negative than those derived using the MSR, SAVI, NDVI and MSAVI2 (i.e. bRVI < bIPVI < bMSR < bSAVI , bNDVI < bMSAVI2). 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 Bearly/Blate 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 10m to 50m. 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 r2, lowest RSE) at each resolution was produced using the RVI, while the weakest relationship was produced using the MSAVI2 (Table 4.3).

Monte Carlo bootstrap resampling indicates that least squares regression provides unbiased estimates of slope (Figure 4.7) and intercept for each of the previously described relationships. For all indices, and at all sample resolutions, bootstrapped estimates (and their associated standard errors) 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.

4.6 Discussion

4.6.1 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 ($\log_e$) curve. Our results differ from other grassland studies, such as those of Weiser et al. (1986) and Roy et al. (1991) (Figure 4.8). 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 < 95g 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, 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-$\log_e$-transformed) data (Figure 4.8). A comparison of each logarithmic relationship with its corresponding linear fit reveals that both statistical models provide statistically similar predictions of B_{early} and B_{late} from spectral information. As a result, we speculate that the logarithmic nature of our observed relationships may simply be an artifact of sampling from within a relatively narrow biomass gradient, and that the "true" biomass-spectral index relationship is actually linear in nature.

Our results also indicate that the strength of each logarithmic relationship is spectral index-dependent. As previously noted, variations in aboveground live biomass are best explained by the NDVI, SAVI, MSAVI2 and MSR, and least-well explained by the DVI. The inability of the "soil-adjusted" indices (SAVI, MSAVI2) to explain more of the variability in aboveground live biomass than some of the "non-soil-adjusted" indices (e.g. NDVI, MSR) is possibly due to the minimal soil background contributions of each sample plot. Only a few plots contained bare patches of soil and, in most cases, these patches comprised less than 10% of the total plot area. This was even true of the most arid plots that were sparsely populated with grass, shrubs or forbs because these sites were almost completely covered by *Selaginella densa* (Dense Club Moss). The similarities between the non soil-adjusted and soil-adjusted indices are consistent with the results of Huete (1988), who found very similar responses between NDVI and SAVI under conditions typical of our grassland sites (i.e. light-colored (dry) soil and low LAI (LAI<1.5)). We are

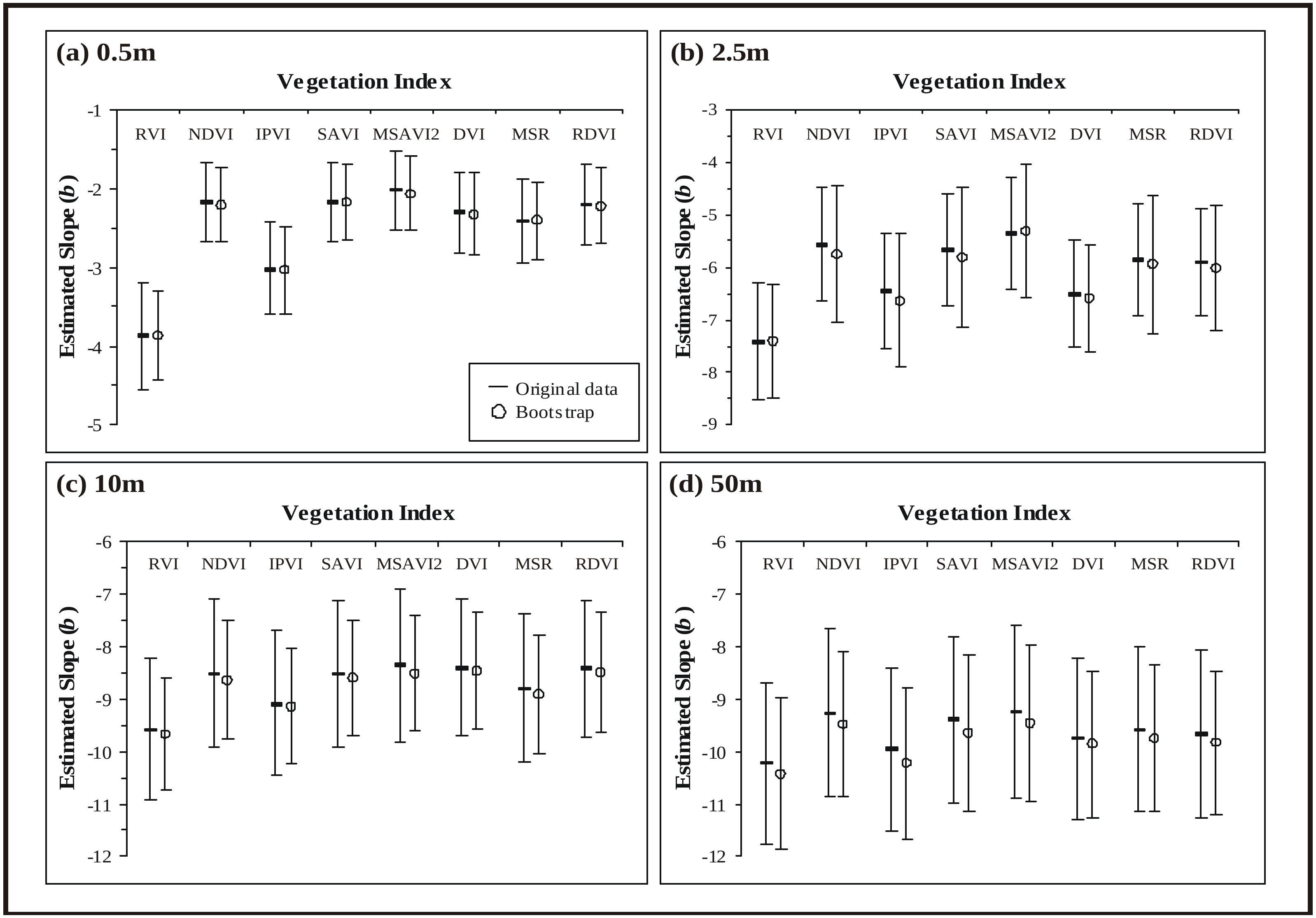


Figure 4.7. Comparisons of the regression-derived slope estimates (and their associated *RSE*) for our original data and those derived from Monte Carlo bootstrap resampling at each sampling resolution.

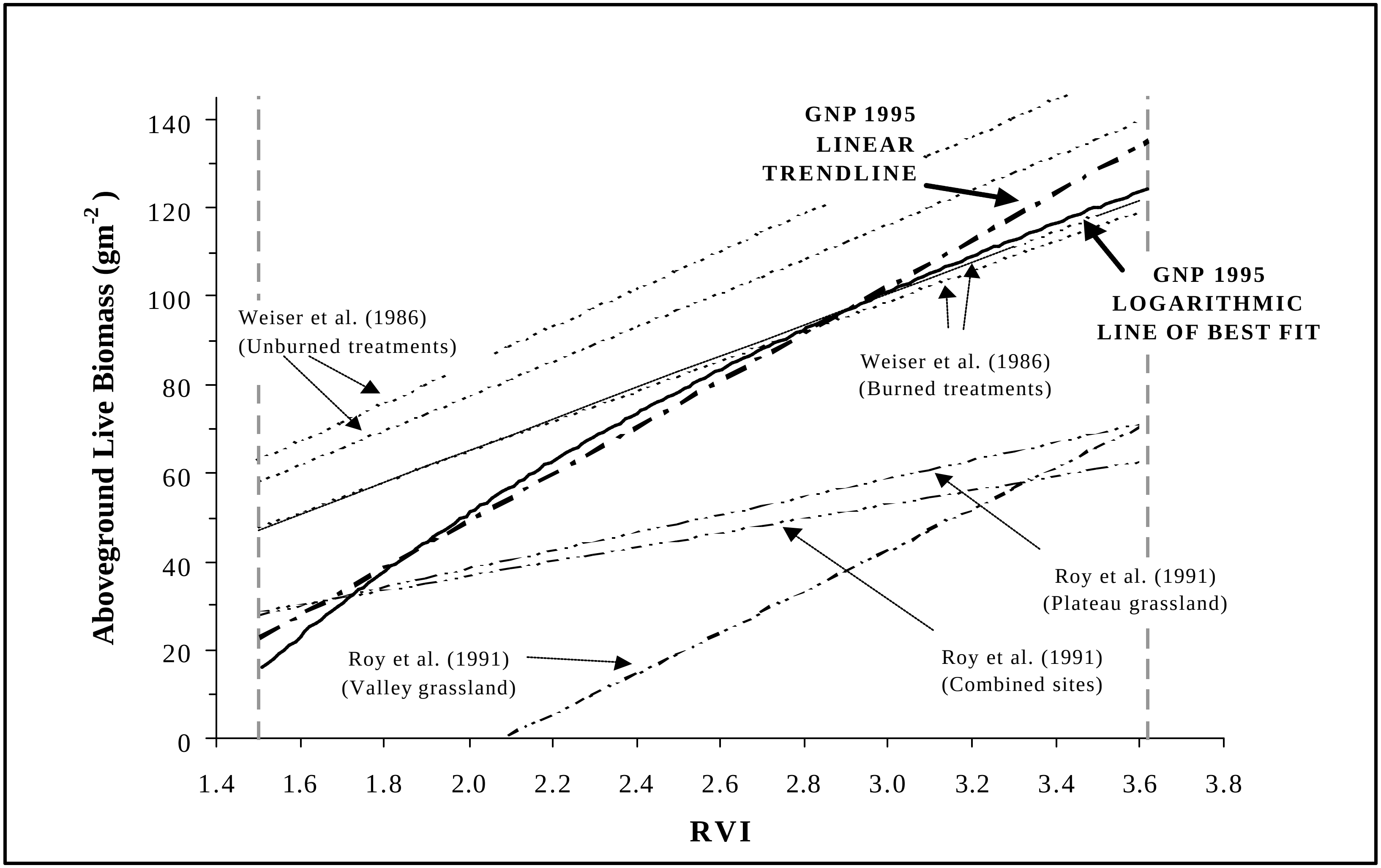


Figure 4.8. A comparison of various field-derived relationships between RVI and aboveground live biomass (gm^{-2}). We present our results as both logarithmic and linear trendlines. When both trendlines were used to predict aboveground live biomass from spectral data, statistically similar results were obtained.

unable to explain the comparatively poor performance of the DVI.

4.6.2 Relationships between remotely-sensed B_{early}/B_{late} and %C4

4.6.2.1 The effects of sample resolution

The evidence presented indicates that the form and strength of the relationship between remotely-sensed B_{early}/B_{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. >2200m for NDVI and elevation (Bian, 1997)).

The weak linear relationships between variates indicate that factors other than C4 species cover influence B_{early}/B_{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 upon the above relationships. The ability of pattern to constrain ecological processes has been well discussed (Wu and Levin, 1994), and its influence on plant productivity at multiple scales deserves further attention (Czárán and Bartha, 1989).

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

4.6.2.2 The effects of vegetation index

Although we found the relationship between B_{early}/B_{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_{early}/B_{late} have the potential to provide information on the relative C4 species coverages of various sample plots within a given dataset.

4.6.3 Implications of results, limitations of approach and future directions

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_{early}/B_{late}) will be negatively correlated with the areal coverage of C4 species (*%C4)*. This multi-resolution 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_{early}/B_{late} and *%C4*, especially at coarser sample resolutions (i.e. 10m to 50m) 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 (10m) and Landsat TM (30m), operate at comparable spatial resolutions to those where the relationships between B_{early}/B_{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. NOAA 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, 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 we are, at present, 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 which 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 4.1. We have not attempted to assess the relative performances of indices which require the derivation of a soil line (e.g. PVI (Richardson and Wiegand, 1977), TSAVI (Baret and Guyot, 1991), MSAVI (Qi et al., 1994)) or those which utilize additional wavebands (e.g. GVI (Crist and Cicone, 1984)).

However, our approach also has its limitations. Although, in theory, the calculation of B_{early}/B_{late} requires spectral measurements from only two dates, the identification of optimal sampling dates requires, in reality, an entire seasonal time series. This is because criteria for the *a priori* identification of such dates do not exist at present. Thus, the real novelty of this approach is that it not only allows the calculation of seasonal NPP from full seasonal time-series data, but that it also allows the estimation of C3/C4 abundances for locations where ground data does not exist.

As a result, even though a strong linear relationship between remotely-sensed estimates of B_{early}/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_{early}/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 greenup). 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_{early}/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_{early} 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.

4.7 Conclusions

In conclusion, although there are several unresolved issues, our results do not contradict our original expectation, and 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.