A Hierarchical Classification of Landsat TM Imagery to Identify Natural Grassland Areas and Rare Species Habitat

Abstract

*A hierarchical classification of Landsat TM data was used to identify native grasslands with high levels of plant species diversity (i.e., high quality grasslands) in eastern Kansas. The initial stage of the hierarchical strategy involved the use of unsupervised clustering to generate a grassland map. Discriminant analysis of ground reference data was then employed to determine multidimensional combinations of original TM bands and selected band transforms that distinguished high quality grasslands from low quality (i.e., native grasslands with low levels of plant species diversity) sites. In the final stages, supervised classification and thresholding were applied to the optimal combination of TM2, TM4, TM5, TM7, and the normalized difference vegetation index (NDVI) to produce a realistic estimate of high quality grassland coverage. Spectral differences between high and low quality grassland training data were attributed to variation in living plant biomass, cover, and water status. Using map overlays of potential high quality sites, 77 previously unknown natural grassland areas were identified from ecological surveys of 135 sites. Additionally, nine of these high quality grasslands contained populations of the federally threatened Mead's milkweed (*Asclepias meadii*). The use of TM data in the hierarchical strategy resulted in improved efficiency of natural area identification in Kansas. The natural grasslands identified are critical components of a conservation database that is used for planning protection of biological diversity in Kansas and for facilitating the integration of development planning with conservation goals.*

Introduction

The recent land-use changes in the tropics, resulting in destruction of large areas of the rain forest system, have sparked worldwide interest in biological diversity and conservation of natural areas. Efforts to conserve biological diversity are shifting from protection of individual threatened species to integrated strategies that protect ecosystems or habitats. This general policy shift from crisis management to preventative conservation addresses the broader problems of ecosystem fragmentation and worldwide loss of biological diversity (Norton, 1986).

To integrate conservation with development goals, a current, easily-accessible database on the biological diversity of a given area is essential (Ahearn *et al.*, 1990). In Kansas, this information is gathered and maintained by the Kansas Natural Heritage Inventory of the Kansas Biological Survey (KBS). The Kansas Inventory is part of a larger effort by The Nature Conservancy to develop an international conservation database (Ahearn *et al.*, 1990). In addition to providing data on ecologically sensitive areas during environmental reviews of planned developments, the KBS administers the Kansas Natural and Scientific Areas Preservation Act which calls for the establishment of a system of preserves. Thus, developing a current database on natural areas and biological diversity is a critical component in achieving conservation goals in Kansas.

This paper presents the results of an analysis of a single Landsat Thematic Mapper (TM) scene to provide natural grassland data useful for the Kansas conservation database. The objectives of this study were to determine the original TM bands and selected band transforms that are significant in identifying native grasslands with high levels of plant species diversity and to assess the effectiveness of a hierarchical classification of TM data to locate unknown grasslands of similar quality in eastern Kansas.

Background

There is a pressing need to identify the remaining tracts of relatively undisturbed tallgrass prairie in the central United States. The original extent of the tallgrass prairie ecosystem in the U.S. was 57,351,100 ha (Sims, 1988), which has drastically declined during the last 100 years. Because their fertile soils were ideal for planting crops, this prairie ecosystem has been highly altered by agricultural development. Additionally, many native prairies have been fragmented and severely disturbed by land development and grazing. Although some estimates suggest that as little as 1 percent of the pre-European settlement acreage remains (Diamond and Smeins, 1988), accurate figures on the extent and condition of the remaining tallgrass prairie ecosystem in the central U.S. are lacking. Because relatively undisturbed prairies can provide critical habitat for several rare plant and animal species, systematic surveys for these native grasslands provide valuable information for use in conservation and development planning.

To provide data for a conservation database, a systematic inventory was initiated in northeastern Kansas in 1987 to identify and evaluate the remaining natural areas in this region. Natural areas are defined as relatively undisturbed tracts of land or water that reflect as nearly as possible the natural conditions prior to European settlement (Lauver, 1989). Criteria used to define a natural area are vegetation-based and include assessments of native and non-native plant species composition and abundance, and the degree of man-made alteration. The methods used to conduct the inventory in northeastern Kansas were adopted from the popular multistage approach developed by White (1978). These

Photogrammetric Engineering & Remote Sensing,
Vol. 59, No. 5, May 1993, pp. 627–634.

0099-1112/93/5905–627$03.00/0

Chris L. Lauver
Kansas Biological Survey, 2041 Constant Ave., Lawrence, KS 66047-2906.

Jerry L. Whistler
Kansas Applied Remote Sensing (KARS) Program, The University of Kansas, 2291 Irving Hill Dr., Lawrence, KS 66045

stages are (1) compile existing information, (2) examine maps and aerial photographs to select potential natural areas (PNAs), (3) conduct aerial reconnaissance of PNAs, and (4) conduct ground surveys (White, 1978).

Accuracy figures describing the effectiveness of White's (1978) methodology are sparse and variable. The inventory of Douglas County in northeast Kansas was completed in 1988, and 129 natural areas were identified. Because over 400 potential sites were field checked in this county (Lauver, 1989), the accuracy for identifying natural areas in this region using White's methodology was roughly 32 percent. A similar inventory in northwestern Missouri identified 113 natural areas in six counties from 312 potential sites (Currier and Smith, 1988) for an accuracy of 36 percent. In a seven-county study in Minnesota, 271 sites qualified as natural areas out of 1062 potential sites identified through photointerpretation of aerial photographs and satellite imagery (Converse, 1990), yielding an accuracy of 26 percent. Other methods using a variety of remotely sensed data are now under investigation in several states to improve the efficiency of natural area identification. This paper presents the results of one such study utilizing Landsat TM imagery to locate species-rich tallgrass prairies and rare species habitat in Kansas.

The Study Area

The grasslands of Anderson County in east-central Kansas (Figure 1) were selected for analysis. The landscape of this 149,344 ha county consists of gently rolling prairies with well defined drainage patterns. Land use in the county is mostly farming and ranching, with over 50 percent in cropland and 36 percent used for pasture and rangeland (USDA, Soil Conservation Service, 1984). Küchler (1974) classifies the potential natural vegetation of Anderson County as mostly tallgrass or bluestem prairie (*Andropogon-Panicum-Sorghastrum*), with oak-hickory forest (*Quercus-Carya*) in the northeast.

The prairie remnants of Anderson County contain populations of three rare species: the federally threatened Mead's milkweed (*Asclepias meadii*), the federally threatened Western prairie fringed orchid (*Platanthera praeclara*), and the federal candidate species, prairie mole cricket (*Gryllotalpa major*). Mead's milkweed and the prairie orchid historically had a wide distribution throughout the tallgrass region (Betz, 1989; Minnesota Dept. of Natural Resources, 1991), while the prairie mole cricket is largely limited to the southern extent of the tallgrass prairie region (Figg and Calvert, 1987). Mead's milkweed grows on prairie remnants that contain approximately 60 prairie plant associates (Betz, 1989), including many forbs (i.e., herbaceous non-grass species) that are commonly found only on ungrazed grasslands because livestock grazing reduces their abundance. The Western prairie fringed orchid is most often found in remnant native prairies but has been reported from disturbed sites. Its rarity is attributed to loss of habitat (Minnesota Dept. of Natural Resources, 1991). The habitat requirements for the prairie mole cricket are largely unknown because populations have been found on a variety of grasslands, but Figg and Calvert (1987) note that the presence of forbs are important components. Because this cricket is a soil-inhabiting arthropod, the compaction of soil by livestock grazing is viewed as a potential detrimental impact. The most attractive habitat to the species appears to be ungrazed prairie haymeadows (Busby, 1991).

Recent field surveys conducted in Anderson County generated data on the location and condition of many natural grassland sites and form the ground reference database for this study. The initial goal of the surveys was to locate high quality grasslands because the species described above appear to require relatively undisturbed grasslands and/or sites with high native plant species diversity. For the purposes of this paper, **high quality (HQ) grasslands** are defined as relatively undisturbed tallgrass prairies with moderate to high levels of native plant species diversity (Plate 1). HQ grasslands are primarily used for annual hay production (typically harvested during July), and have little to no grazing impacts from domestic livestock. In contrast, **low quality (LQ) grasslands** have been altered by domestic livestock grazing and agricultural practices (e.g., herbicidal treatment) and contain low levels of native plant species diversity (Plate 2). Both grassland types are dominated by native tallgrasses but HQ sites contain an abundance of forbs (Figure 2). Because many of these forbs are palatable to cattle, moderate to heavy grazing on these sites reduces their number and abundance. For sites with prolonged heavy grazing, plant biomass and cover are greatly reduced in some areas, often resulting in grass islands surrounded by bare ground (Hetzer and McGregor, 1951). Thus, HQ sites (haymeadows) are dominated by a relatively uniform cover of grasses and forbs whereas LQ sites (grazed prairies) are dominated by grasses of variable cover with scattered forbs. These differences in plant biomass, cover, and species composition were the biophysical charac-

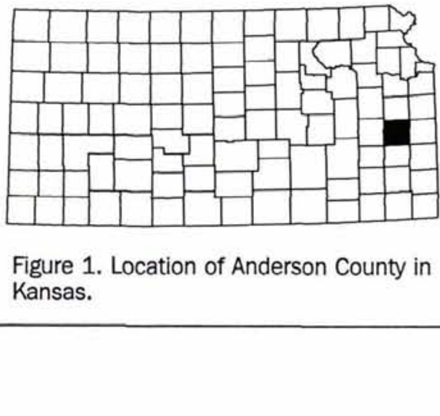

Figure 1. Location of Anderson County in Kansas.

Plate 1. An example of a high quality grassland identified by this study (photo date: 5 July 1990).

Plate 2. A low quality grassland in Anderson County that had recently been grazed (photo date: 21 June 1990).

Plate 3. Level I land-use/land-cover map of a portion of Anderson County.

teristics hypothesized to aid in distinguishing the spectral features of HQ from LQ grasslands.

Methods

Image Acquisition

To test the utility of digital satellite data to identify high quality grasslands, a single Landsat TM image of east-central Kansas (7 June 1988, Path 27, Row 33) was obtained that included nearly all of Anderson County. A multitemporal analysis of grasslands was prohibited due to limited project funds. Based on the biophysical characteristics of low and high quality grasslands during the growing season described above, a TM scene from June 1988 was selected for analysis. The traditional time to harvest HQ grasslands for hay is in early July; during June these sites typically contain healthy and vigorous plants that form dense canopies of grasses and forbs. Grazed low quality sites may appear fairly vigorous during June, but a history of grazing on these sites has resulted in a mosaic of grass patches of variable cover and biomass with scattered forbs.

Procedure

The hierarchical classification strategy that was developed to meet the study objectives is presented in Figure 2. The strategy employs unsupervised and supervised classification techniques sequentially to classify (and eliminate) non-grassland areas and to produce a map of high and low quality native grasslands. Discriminant analysis was used to determine the TM bands and selected band transforms that were significant in differentiating high from low quality grasslands. The microcomputer based SOLO statistical software package (Hintze, 1988) was used to conduct linear discriminant analyses of grassland data. The TM image was classified and raster map data were processed using the ERDAS image processing/geographic information system (GIS) software, and the vector ground reference data and map overlays were processed and produced using the pc ARC/INFO GIS software.

Unsupervised Classification

The first step in the hierarchical strategy was to produce a Level I land-use/land-cover classification (Anderson *et al.*, 1976) for Anderson County (Figure 4). An unsupervised classification was performed on the original Landsat TM data to help compose the Level I map. Six TM bands with a spatial resolution of 30 by 30 m were used during this step; their wavelength intervals in μm units are: band 1 (TM1), 0.45 to 0.52; TM2, 0.52 to 0.60; TM3, 0.63 to 0.69; TM4, 0.76 to 0.90; TM5, 1.55 to 1.75; and TM7, 2.08 to 2.35. The thermal band, TM6, was not employed because of its low spatial resolution.

An unsupervised clustering algorithm (STATCL) identified 78 unique spectral classes (from an initial total of 100 user-specified classes) that were inherent in the original TM data of Anderson County. A maximum-likelihood classifier assigned each image pixel to one of these spectral classes. Each spectral class then was analyzed visually with the aid of color infrared photographs and assigned to one of the following land-use/land-cover types: (1) Urban/Built-up, (2) Agricultural land, (3) Grassland, (4) Woodland, (5) Water, (6) Mixed-inert, and (7) Mixed-organic (Plate 3). Although this level of class detail was unnecessary for the present study, this stage served as a prototype for a planned statewide land-use/land-cover mapping project. Spectral classes were assigned to cover types based on the type that dominated the class. Mixed classes with apparently low proportions of grass were assigned to the grassland cover type to minimize potential omission errors of HQ grasslands. The non-grassland cover types were then eliminated to form an initial grassland map. Using this map as a mask, image pixel values coincident with the grassland cover type were extracted for further analysis.

Variable Selection and Discriminant Analysis

Using discriminant analysis, we investigated the utility of five raw data bands (TM2, TM3, TM4, TM5, TM7) and four band transforms (NDVI, brightness, greenness, wetness; see Table 1) to distinguish high from low quality grasslands. Data from TM1 was not used as an independent variable in discriminant analysis because haze was visibly evident over the study area when displaying this band. The normalized difference vegetation index (NDVI) (Rouse *et al.*, 1974) was used because it responds to changes in amount of green biomass (Tucker, 1979), and biomass differences between high and low quality grasslands were noted from field surveys. Other ratio-based indices using data from the red and near infrared regions have been reported in the literature, but many of these transforms are functionally equivalent (Perry and Lautenschlager, 1984). The other transforms investigated were the TM tasseled cap features of brightness, greenness, and wetness (Crist and Cicone, 1984). Because differences in plant biomass, cover, and species composition were assumed between LQ and HQ sites during the TM scene date (based on field surveys conducted in 1988-1989), the brightness and greenness features were analyzed for their ability to distinguish these vegetation differences. Additionally, the wetness feature was evaluated because the amount of plant water per unit area is correlated with biomass and this transform might detect variation in moisture conditions between HQ and LQ grasslands.

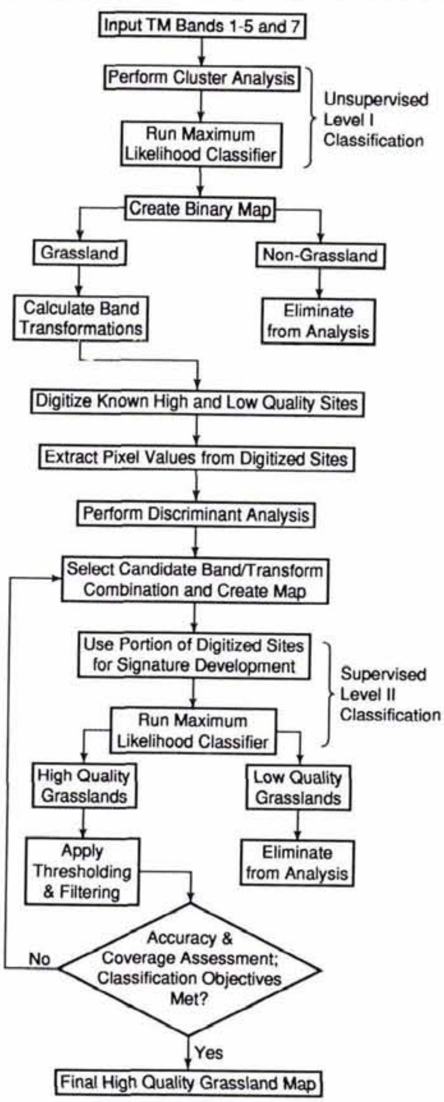

Figure 2. Schematic diagram of the hierarchical classification and methods used to produce a map of potential high quality grassland.

TABLE 1. THE NORMALIZED DIFFERENCE VEGETATION INDEX (NDVI) (ROUSE ET AL., 1974) AND THREE TM TASSELED CAP FEATURES (CRIST AND CICONE, 1984) USED IN DISCRIMINANT ANALYSES OF GRASSLANDS IN EASTERN KANSAS. COMPUTED VALUES WERE SCALED FROM 0 TO 255.

Normalized Difference Vegetation Index:

NDVI = (TM4 − TM3)/(TM4 + TM3)

TM Tasseled Cap Transformations:

Brightness = 0.3037 TM1 + 0.2793 TM2 + 0.4743 TM3 + 0.5585 TM4 + 0.5082 TM5 + 0.1863 TM7

Greenness = − 0.2848 TM1 − 0.2435 TM2 − 0.5436 TM3 + 0.7243 TM4 + 0.840 TM5 − 0.1800 TM7

Wetness = 0.1509 TM1 + 0.1973 TM2 + 0.3279 TM3 + 0.3406 TM4 − 0.7112 TM5 − 0.4572 TM7

To provide data for discriminant analyses, known sites of high and low quality grasslands were digitized and overlaid on the original and transformed images. Values for the five raw bands and the four transforms were then extracted for the pixels that were coincident with the digitized sites, forming nine data values per grassland pixel.

Multidimensional raw band/transform combinations were analyzed to determine the optimal combination that satisfied the study objectives. Research in determining an optimal band combination for thematic classification has shown classification accuracies will obtain some maximum level and slowly decrease with increasing dimensionality, given a fixed number of training samples (Swain and Davis, 1978; Nelson *et al.*, 1984). Thus, an optimal subset of the nine independent variables (five bands, four transforms) was sought that provided a "best measurement complexity" (Swain and Davis, 1978). Several candidate combinations, each representing the best mix of significant variables ($p < 0.01$) of a specific dimension (Table 2), were chosen from the discriminant analysis results. To prepare the grassland image for supervised classification, new grassland images containing only those variables for each candidate combination were created.

Supervised Classification

A supervised approach was used to create a map of potential high and low quality grasslands. Using the grassland image composed of the variables of a candidate combination, groups of pixels from known sites were selected that best represented the two grassland types to form training samples. Contingency table analysis of the selected training samples was performed to eliminate samples that provided low classification accuracy. A total of 1896 pixels from eight sites were selected as the training set for high quality grasslands, while 2196 pixels from six sites formed the set for low quality grasslands. Utilizing the statistics provided by these training samples, a maximum-likelihood classifier was used to assign all grassland image pixels to one of the two quality types. The pixels classified as low quality were eliminated to

TABLE 2. RESULTS OF DISCRIMINANT ANALYSES OF DATA FROM HIGH AND LOW QUALITY GRASSLANDS SHOWING THE BEST RAW BAND/TRANSFORM COMBINATION ($P < 0.01$) FOR THREE THROUGH SIX DIMENSIONS. THE NINE INDEPENDENT VARIABLES USED WERE TM2, TM3, TM4, TM5, TM7, AND FOUR BAND TRANSFORMS (NDVI, BRIGHTNESS, GREENNESS, WETNESS).

Dimensions	Raw Band/Transform Combination
3	TM2, TM7, Brightness
4	TM2, TM5, TM7, Brightness
5	TM2, TM4, TM5, TM7, NDVI
6	TM2, TM3, TM4, TM5, TM7, Brightness

create a map of potential high quality grassland areas. Although identifying LQ grasslands may be important for restoration or preserve design purposes, the study objective was to evaluate the use of TM data to locate HQ sites.

Post-Classification Processing and Selection of an Optimal Combination

Post-classification processing of the high quality grassland map was conducted to provide accuracy and coverage assessment data for determination of an optimal raw band/transform combination. The selection of an optimal combination was based on two criteria: the combination yielded a comparatively high classification accuracy while minimizing dimensionality, and provided a realistic estimate of the total area of the HQ grassland class. To meet the area criterion, combinations that classified less than 10 percent of the land area in Anderson County as high quality grassland were sought. A previous natural resources inventory of Anderson County reported that 16 percent of the county was rangeland in good to excellent condition (USDA, Soil Conservation Service, 1984). Excellent and good condition rangelands are surrogate terms for high quality grasslands, but because plant diversity is not used as an explicit criterion for rangeland condition, these terms are not equivalent. Based on field observations and the rangeland condition figures, an estimated 8 percent of the county land area (i.e., 50 percent of the good to excellent rangeland) was thought to be high quality tallgrass prairie.

Post-classification processing of the HQ grassland map was also conducted to reduce potential errors of commission. Two related sources of error were the inclusion of mixed spectral classes containing some non-grass pixels to the grassland cover type during unsupervised clustering, and the lack of training data to eliminate all low quality grassland types during supervised classification. The technique of thresholding was applied to the map to identify and eliminate pixels that had a low probability of being assigned to the correct class. Although the application of thresholds results in lower classification accuracies, thresholds can be used to improve estimates of the total area covered by target classes within a study area (Schowengerdt, 1983).

A test area was used to evaluate the application of thresholds to the raw band/transform combinations derived from discriminant analyses. Areas from the ground reference data that were withheld during training sample selection were combined to form a relatively unbiased sample. The number of test pixels for the HQ grassland class was 2554. The estimated classification accuracy for each raw band/transform combination after thresholding was evaluated in relation to the total land area of the county that was classified as high quality grassland. After selecting an optimal combination based on the criteria described above, the image was classified and the output map filtered to portray a final map of potential high quality grasslands that were greater than 2 ha (or 23 pixels). This raster map was converted to vector format in order to plot overlays for ground surveys.

TABLE 3. FIFTEEN DECREASER AND INCREASER PLANT SPECIES OF TALLGRASS PRAIRIES IN EASTERN KANSAS. NOMENCLATURE FOLLOWS THE GREAT PLAINS FLORA ASSOCIATION (1986).

Decreaser Species	Increaser Species
Amorpha canescens	Achillea millefolium
Andropogon gerardii	Antennaria neglecta
Andropogon scoparius	Aster ericoides
*Ceanothus herbaceous	Baptisia bracteata
*Dalea candida	Bouteloua curtipendula
*Dalea purpurea	Bouteloua gracilis
Helianthus rigidus	Dichanthelium oligosanthes
Hieracium longipilum	Erigeron strigosus
Lespedeza capitata	Linum sulcatum
Liatris pycnostachya	Physalis pumila
Phlox pilosa	Plantago patagonica
Sorghastrum nutans	Solidago canadensis
Sporobolus heterolepis	Solidago missouriensis
Stipa spartea	Verbena stricta
Viola pedatifida	Vernonia baldwinii

* key indicator species

Ground Surveys

Using map overlays of potential high quality sites registered to 1:24,000-scale USGS topographic maps, 135 sites ranging from 2 to over 120 ha were field checked during May-October of 1990. The natural quality of each site was determined by collecting data on plant species composition. Natural quality or condition refers to the similarity of the plant community in relation to its potential state under undisturbed conditions (Carneggie *et al.*, 1983). The condition or quality of each site was based on the relative proportions of increaser, decreaser, and invader plant species (Dyksterhuis, 1949). Sites in good to excellent condition (i.e., HQ sites) were dominated by decreaser species (Table 3), whereas sites in fair to poor condition (LQ sites) were dominated by increaser species and invaders. Additionally, the presence of key indicator species (Table 3), including the federally listed Mead's milkweed, facilitated the determination of natural quality. These indicator species are particularly sensitive to disturbance and their presence represents a relatively undisturbed site.

Results and Discussion

Four raw band/transform combinations derived from discriminant analyses were evaluated for classification accuracy and predicted land coverage of the target class (Table 2). The variables that formed these combinations were TM2, TM3, TM4, TM5, TM7, NDVI, and the brightness feature. Spectral differences between high and low quality grasslands were revealed in training data of TM5, TM7, and the two transforms; training data from the other raw data bands showed only minor spectral differences (Table 4). Discriminant analysis results showed no significant differences between HQ and LQ sites for the greenness and wetness features. Other variables (TM5, TM7, NDVI, and brightness) were apparently more important than these two features in distinguishing vegetation differences between HQ and LQ grasslands.

An examination of the training data in Table 4 indicates

TABLE 4. AVERAGE SPECTRAL VALUES AND STANDARD ERRORS (S.E.) OF TRAINING DATA OF THE SEVEN VARIABLES THAT WERE SIGNIFICANT IN DISCRIMINANT ANALYSES OF HIGH AND LOW QUALITY GRASSLAND DATA.

	Low Quality Grasslands		High Quality Grasslands	
Variable	Mean	S.E.	Mean	S.E.
TM2	43.8	2.6	41.9	1.7
TM3	46.4	4.1	42.1	3.2
TM4	104.3	7.3	105.5	6.0
TM5	123.1	10.1	109.7	7.5
TM7	45.7	6.3	36.8	4.6
NDVI	97.2	12.8	108.9	10.9
Brightness	119.5	13.1	103.3	10.0
Number of pixels	8945		5655	

that the spectral differences between high and low quality grasslands appear to be related to the biomass and water status of living plants on these sites at the time of TM data acquisition. Low quality grasslands were spectrally brighter than HQ sites in the mid-infrared region represented by TM5 and TM7 (Table 4). This region, specifically the 1.55- to 1.75-µm interval (TM5), can provide an accurate indication of leaf water content or plant canopy water status (Tucker, 1980). Because reflectance of TM5 from green leaves is reduced because of water absorption (Knipling, 1970; Tucker, 1980), the evidence suggests that, per unit area, HQ grasslands contained greater water content than LQ sites. In a recent study of native tallgrasses subjected to various management treatments, Price *et al.* (1992) found that spectral data from the infrared region were highly correlated with total leaf moisture. The moisture content variation between HQ and LQ sites may also be associated with biomass differences between the two grassland types. Using the vegetation index NDVI as a biomass indicator, the ungrazed HQ grasslands contained more plant biomass than the grazed LQ sites (Table 4). Previous studies of the tallgrass prairie in Kansas have shown the utility of the NDVI to assess vegetation characteristics. Tappan (1980) found this index to be strongly correlated with green vegetation cover and biomass, and Turner (1990) reported decreased NDVI values following foliage removal by mowing. Price *et al.* (1992) reported strong relationships between spectral reflectance data and the biomass and leaf area index (LAI) of several native tallgrasses. Thus, the greater biomass (per pixel or per unit area) on the ungrazed HQ grasslands contributed to a higher water content per unit area compared to the grazed low quality sites. Although grazing had influenced the species composition of the LQ sites, living plant biomass and water status appear to be two parameters that facilitated the identification of grassland quality with TM data.

Other biophysical characteristics that may have contributed variation to grassland spectral response can be related to the effects of grazing on the low quality training sites. For example, physical characteristics that increase total reflectance, such as decreased canopy cover and increased soil exposure, will be expressed in the brightness feature (Crist and Cicone, 1984). Field observations revealed that the variable grazing intensity on the LQ sites created a mosaic of vegetation patches that were lightly to heavily grazed (Figure 3), with many areas having decreased vegetation cover and increased soil exposure. As expected, training data show higher brightness values associated with the grazed LQ sites (Table 4) due to the effects of grazing on the soil-plant interface. Thus, living plant cover appears to be another parameter that facilitated the discrimination of grassland quality. The use of the soil-adjusted vegetation index (SAVI) developed by Huete (1988) might have assisted in discerning these grazing effects. However, the SAVI was not investigated in this study because its utility is limited to intermediate levels (50 percent) of green plant cover (Huete, 1988); the dense green vegetation of the HQ grasslands (often near 100 percent; see Figure 2) minimizes the soil signal emerging from their canopies.

The lower biomass on the grazed LQ sites, as indicated by the NDVI values, also may have contributed to greater reflectance in the red spectral region (TM3) and lower near infrared (TM4) reflectance values (Table 4). Asrar *et al.* (1989) found that reduced levels of green leaf area on tallgrass prairie resulted in an increase of red reflectance and a decrease of near infrared reflectance. Other studies by Asrar and associates of burned and unburned prairies in Kansas have shown that the layer of senescent vegetation on unburned prairies significantly increases red and infrared reflectances (Asrar *et al.*, 1986; Asrar *et al.*, 1987). Although no data are available concerning the burning of the prairies in the present study, the variable grazing effects on low quality grasslands may have resulted in patches of senescent vegetation (i.e., unharvested growth from previous years) on these sites. In contrast, the potential presence of senescent vegetation on high quality grasslands was low because the green biomass was harvested yearly on these sites for hay. The presence of senescent vegetation, in addition to the other effects discussed above, may have contributed to greater reflectance of TM3, TM5, and TM7 on the low quality grasslands (Table 4).

Accuracy Assessment of Known Sites and Estimated Coverage

Using ground reference data withheld from training, classification accuracies were obtained for the four band-transform combinations (Table 5). Because the combinations were composed of significant variables, their accuracy figures were similar, ranging from 76.5 percent to 82.3 percent. However, each combination classified at least 16 percent of the study area as HQ grassland. This was thought to be an overestimate of HQ grassland coverage based on recent field surveys and rangeland condition coverages reported for Anderson County. Different threshold levels were then applied to the candidate combinations to reduce the estimated coverage of the HQ class. These applications tended to eliminate small groups of isolated pixels along drainage ways that had a low probability of being correctly classified.

The band-index combination of TM2, TM4, TM5, TM7, and NDVI subjected to thresholding at the 10 percent significance level was selected to produce the final HQ grassland map. This combination-threshold version yielded a comparatively high classification accuracy and provided a realistic estimate of the total area of the HQ grassland class (Table 6). Application of the threshold reduced the accuracy from 76.7 percent to 63.4 percent but significantly decreased the estimated total land area from 16.6 percent to the targeted 8 percent of the

Table 5. Accuracy Figures and Estimated Land Coverage of High Quality (HQ) Grasslands.

Raw Band/Transform Combination	*Classification Accuracy (%)	Percent of Study Area as HQ grassland
TM2, TM7, Brightness	82.3	20.7
TM2, TM5, TM7, Brightness	76.5	16.4
TM2, TM4, TM5, TM7, NDVI	76.7	16.6
TM2, TM3, TM4, TM5, TM7, Brightness	79.1	18.3

*Accuracy figures based on 2554 HQ grass pixels withheld from training

Table 6. Accuracy Figures and Estimated Land Coverage of High Quality (HQ) Grasslands Following Application of Thresholds.

Raw Band/Transform Combination	Threshold Level (%)	*Classification Accuracy (%)	Percent of Study Area as HQ grassland
TM2, TM7, Brightness	20	58.5	10.1
TM2, TM5, TM7, Brightness	10	64.1	9.5
TM2, TM5, TM7, Brightness	15	54.8	6.9
TM2, TM4, TM5, TM7, NDVI	10	63.4	8.0
TM2, TM3, TM4, TM5, TM7, Brightness	5	63.0	8.3

* Accuracy figures based on 2554 HQ grass pixels withheld from training

Table 7. Error Matrix of Supervised Classification (TM2, TM4, TM5, TM7, and NDVI) Versus Grassland Reference File Withheld from Training.

Classified Data	Reference Data: Low Quality	Reference Data: High Quality	Totals	Commission Error (%)
Low Quality	3885	595	4480	13.3
High Quality	1319	1959	3278	40.2
Totals (Pixels)	5204	2554	7758	
Omission Error (%)	25.4	23.3		

study area (Tables 5 and 6). Although the four-dimension combination (TM2, TM5, TM7, and brightness) thresholded at the 10 percent level approximated the results of the five-dimension combination (Table 6), the lower predicted coverage and the use of the biomass indicator NDVI favored the selection of the five-dimension combination. The increased dimensionality represented by the six-band combination of TM2, TM3, TM4, TM5, TM7, and brightness produced results similar to the five-band combination (Table 6). Thus, the five-dimension combination represented the mix of the fewest number of variables that met the classification objectives.

The errors of commission and omission associated with the selected combination (TM2, TM4, TM5, TM7, and NDVI) are shown in Table 7 using grassland reference pixels withheld from training. The omission errors for both grassland types are approximately 24 percent (Table 7). However, the commission error for high quality grasslands (40.2 percent) is much greater than the error for LQ sites (13.3 percent). This result was anticipated because of the assignment of mixed spectral classes to the grassland cover class during unsupervised classification.

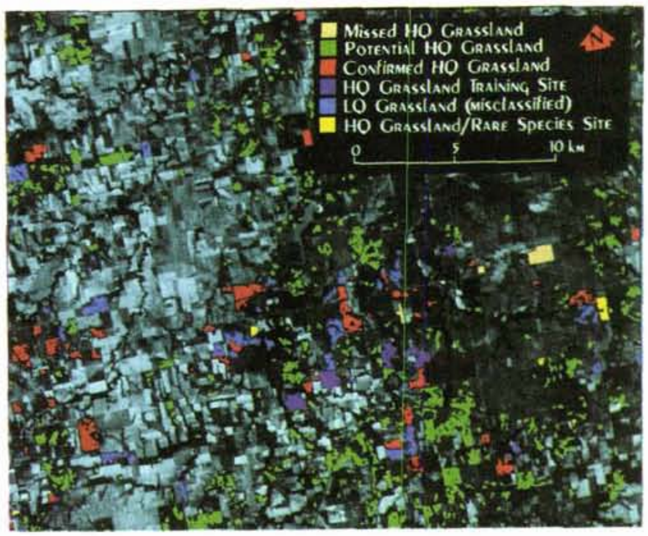

Plate 4. Map of potential, misclassified, omitted and confirmed high quality grasslands that occur in the same area as Plate 1. Areas in yellow contained populations of Mead's milkweed (*Asclepias meadii*) during 1990.

Accuracy Assessment of New Sites and Rare Species Habitat

Of the 135 previously unknown sites surveyed in the field, 77 were determined to be high quality grasslands and 58 sites were low quality grasslands. The distribution and coverage of many of the sites surveyed are presented in Plate 4. Potential high quality sites were concluded to be correctly classified if at least 51 percent of the polygon area on the HQ map corresponded with a high quality grassland. Thus, the site-specific accuracy (Mead and Szajgin, 1982) of this method was 57 percent. However, because 58 of the 135 potential HQ sites were assessed as low quality grasslands, commission errors were substantial (43 percent). This finding was anticipated because of the error matrix results of the training data (Table 7). Although no quantitative data were obtained from these 58 sites, field observations suggest that grazing impacts were relatively low on these sites and did not produce significant changes in plant biomass, cover, or water status compared to the ungrazed HQ sites.

One reason accuracy figures are generally low for natural area inventories is their stated objective to locate all of the remaining natural areas within a specific region. Previous work using aerial photointerpretation to identify natural areas has accuracy figures ranging from 26 percent (Converse, 1990) to 36 percent (Currier and Smith, 1988). Because of their objective, inventory methods are designed to minimize errors of omission, and results often include high commission errors. Although the findings of this study contain notable commission errors (Table 7), the hierarchical classification of TM data presented here represents nearly a two-fold increase in the accuracy of natural area identification.

The map of potential high quality sites contained few omission errors. Of the 135 sites field checked, three additional HQ sites were located during field surveys (Plate 2); these sites were classified as LQ grasslands from supervised classification. The low commission error (13.3 percent) for low quality sites shown in Table 7 provides evidence suggesting the potential number of other LQ grassland sites that may have been misclassified is low.

A significant additional benefit to the identification of these natural areas in eastern Kansas was the discovery of several new populations of the federally threatened Mead's milkweed (Plate 2). Of the 77 new sites identified, nine contained small populations of this species. These data support previous work indicating that this species only occurs on unplowed native tallgrass prairies with high species diversity (Morgan, 1980; U.S. Fish and Wildlife Service, 1988; Betz, 1989). Although systematic surveys for rare species were not conducted in this study, these results indicate great potential for utilizing Landsat TM data to facilitate the collection of data on rare species that appear to require high quality native grassland habitat.

Conclusions

Application of a digital image classification strategy improved the accuracy of natural grassland identification compared to methods using aerial photography and aerial surveys. Whereas the latter method uses ground reference data to develop qualitative "search images" to locate potential natural areas, these data (i.e., training sites) are used statistically in the hierarchical classification to develop quantitative spectral signatures to identify potential high quality sites. The signatures discriminated high from low quality sites based on raw and transformed bands which have been shown previously to be effective surrogates for measuring living plant biomass, cover, and water status.

Although we do not recommend the hierarchical strategy for detailed mapping of natural areas, the improved accuracy of natural area identification over other methods is significant. Given the relative short time required for image processing as compared to our experience using aerial photographs and aerial surveys to generate maps of potential natural areas, this method is faster and more accurate. A disadvantage of our approach is the requirement of training data to use during discriminant analysis and supervised classification. However, if the locations of a few exemplary sites are known, the strategy presented here can be applied to large study areas to rapidly develop maps of sites with high natural area potential. The efficient identification of natural areas promotes an efficient assessment of biological diversity. Once potential areas have been identified, biological field surveys of these sites will generate data on the best remaining natural areas and rare species habitat. These data are essential for conservation planning of biological diversity, and for facilitating the integration of development planning with conservation goals.

Acknowledgments

This investigation was supported by University of Kansas General Research allocation #3171-X0-0038. We thank Drs. Brad Reed and Mary Dillworth for assistance in the development of the study methodology, and Frank Norman for contributing field survey data. We also thank Drs. B. Reed, Susan Glenn, and Kevin Price for reviewing earlier versions of the manuscript. This paper was significantly improved by the constructive criticisms and recommendations of two anonymous reviewers.

References

Ahearn, S.C., J.L.D. Smith, and C. Wee, 1990. Framework for a geographically referenced conservation database: case study Nepal. *Photogrammetric Engineering & Remote Sensing*, 56(11):1477-1481.

Anderson, J.R., E.E. Hardy, J.T. Roach, and R.E. Witmer, 1976. *A Land Use and Land Cover Classification System for Use with Remote Sensor Data*, United States Government Printing Office, Washington, D.C.

Asrar, G., R.L. Weiser, D.E. Johnson, E.T. Kanemasu, and J.M. Killeen, 1986. Distinguishing among tallgrass prairie cover types from measurements of multispectral reflectance. *Remote Sensing of Environment*, 19:159-169.

Asrar, G., R.B. Myneni, Y. Li, and E.T. Kanemasu, 1987. Diurnal and seasonal characteristics of a tallgrass prairie. *Proceedings of IGARSS Symposium*. University of Michigan, Ann Arbor, pp. 199-201.

———, 1989. Measuring and modeling spectral characteristics of a tallgrass prairie. *Remote Sensing of Environment*, 27:143-155.

Betz, R.F., 1989. Ecology of Mead's milkweed (*Asclepias meadii* Torrey). *Proceedings, 11th North American Prairie Conference*, Lincoln, Nebraska, pp. 187-191.

Busby, W.H., 1991. *Prairie Mole Cricket Report for Kansas, 1991*, Report No. 47, Kansas Biological Survey, Lawrence, Kansas.

Carneggie, D.M., B.J. Schrumpf, and D.A. Mouat, 1983. Rangeland applications. *Manual of Remote Sensing*. American Society of Photogrammetry, Falls Church, Virginia, pp. 2325-2374.

Converse, C.K., 1990. *Interim Report 1990*, Minnesota County Biological Survey. Minnesota Department of Natural Resource, Biological Report No. 31, St. Paul, Minnesota.

Crist, E.P., and R.C. Cicone, 1984. A physically-based transformation of thematic mapper data – the TM tasseled cap. *IEEE Trans. Geosc. Rem. Sens.*, GE. 22(3), 256-263.

Currier, M., and T.E. Smith, 1988. *Final Report on the Missouri Natural Features Inventory – Buchanan, Cass, Clay, Clinton, Jackson, and Platte Counties*. Missouri Department of Conservation, Jefferson City, Missouri.

Diamond, D.D., and F.E. Smeins, 1988. Gradient analysis of remnant true and upper coastal prairie grasslands of North America. *Canadian J. Botany* 66:2152-2161.

Dyksterhuis, E.J., 1949. Condition and management of range land based on quantitative ecology. *J. Range Management* 2:104-115.

Figg, D.E., and P.D. Calvert, 1987. *Status, Distribution and Life History of the Prairie Mole Cricket*, Gryllotalpa major Saussure, Missouri Dept. of Conservation, Jefferson City, Missouri.

Great Plains Flora Association, 1986. *Flora of the Great Plains*. Univ. Press of Kansas, Lawrence, Kansas, 1392 p.

Hetzer, W.A., and R.L. McGregor, 1951. An ecological study of the prairie and pasture lands in Douglas and Franklin Counties, Kansas. *Trans. Kansas Acad. Sci.* 54(3):356-369.

Hintze, J.L., 1988. *SOLO Statistical System, Version 2.0*. BMDP Statistical Software, Inc., Los Angeles, California.

Huete, A.R., 1988. A soil-adjusted vegetation index. *Remote Sensing of Environment*, 25:295-309.

Knipling, E.B., 1970. Physical and physiological basis for the reflectance of visible and near-infrared radiation. *Remote Sensing of Environment*, 1:155-159.

Küchler, A.W., 1974. A new vegetation map of Kansas. *Ecology*, 55:586-604.

Lauver, C.L., 1989. *Final Report on the Natural Areas Inventory of Douglas County, Kansas with a Progress Report on the Inventory of other Kansas Counties*, Report No. 38, Kansas Biological Survey, Lawrence, Kansas.

Mead, R.A., and J. Szajgin, 1982. Landsat classification accuracy assessment procedures. *Photogrammetric Engineering & Remote Sensing*, 48(1):139-141.

Minnesota Dept. of Natural Resources, 1991. *Western Prairie Fringed Orchid: A Threatened Midwestern Prairie Plant*, Biological Report #36, St. Paul, Minnesota.

Morgan, S.W., 1980. *Status Report on* Asclepias meadii Torr., Missouri Department of Conservation, Jefferson City, Missouri.

Nelson, R.F., R.S. Latty, and G. Mott, 1984. Classifying northern forests using thematic mapper simulator data. *Photogrammetric Engineering & Remote Sensing*, 50(5):607-617.

Norton, B.G. (editor), 1986. *The Preservation of Species – The Value of Biological Diversity*. Princeton University Press, Princeton, New Jersey, 283 p.

Perry, C.R., Jr., and L.F. Lautenschlager, 1984. Functional equivalence of spectral vegetation indices. *Remote Sensing of Environment*, 14:169-182.

Price, K.P., V.C. Varner, E.A. Martinko, and D.C. Rundquist, 1992. Analysis of multitemporal narrow-band spectroradiometer measurements from six prairie treatments in Kansas. *ASPRS Technical Papers, 1992 ASPRS-ACSM Annual Convention*, Albuquerque, New Mexico, pp. 372-385.

Rouse, J.W., R.H. Haas, J.A. Schell, and D.W. Deering, 1974. Monitoring vegetation systems in the Great Plains with ERTS. *Proc., Third Earth Resources Technology Satellite-1 Symp.*, Goddard Space Flight Center, NASA SP-351, Science and Technical Information Office, NASA, Washington, D.C., pp. 309-317.

Schowengerdt, R.A., 1983. *Techniques for Image Processing and Classification in Remote Sensing*. Academic Press, New York, N.Y., 249 p.

Sims, P.L., 1988. Grasslands. *North American Terrestrial Vegetation*. Cambridge University Press, New York, N.Y., pp. 266-286.

Swain, P.H., and S.M. Davis, 1978. *Remote Sensing: The Quantitative Approach*. McGraw-Hill International Book Co., New York, N.Y.

Tappan, G., 1980. *The Monitoring of Rangeland Vegetation Cover in the Kansas Flint Hills from Landsat Data*. M.S. Thesis. Univ. of Kansas, Lawrence, Kansas, 105 p.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8:127-150.

———, 1980. Remote sensing of leaf water content in the near infrared. *Remote Sensing of Environment*, 10:23-32.

Turner, C.L., 1990. *The Influence of Grazing on Plant Productivity and Canopy Spectral Reflectance Characteristics in Tallgrass Prairie*. Dissertation. Kansas State University, Manhattan, Kansas.

USDA, Soil Conservation Service, 1984. *Kansas Resource Inventory, Data for Anderson County*. Salina, Kansas.

U.S. Fish and Wildlife Service, 1988. Endangered and threatened wildlife and plants; determination of threatened status for *Asclepias meadii* (Mead's milkweed). *Federal Register* 53(170):33992-33995.

White, J., 1978. *Illinois Natural Areas Inventory Technical Report, Volume I: Survey Methods and Results*. Illinois Natural Areas Inventory, Urbana, Illinois, 426 p.

(Received 9 September 1991; revised and accepted 8 September 1992)