Using Multisource Data in Global Land-Cover Characterization: Concepts, Requirements, and Methods

Abstract
Global land-cover data are needed as baseline information for global change research. Multisource data, both coarse-resolution satellite data and ancillary data, were used to produce a land-cover characteristics database for the conterqinous United States. Ancillary data, including elevation and ecological region data sets, were critical to the development, refinement, and information content of each class in the database. They contributed essential evidence for labeling and refining land-cover classes where differing types were represented by single spectral-temporal signatures. The characterization process can be expanded to a global effort depending on (1) the availability of global satellite coverage, (2) the quality and availability of ancillary data, and (3) the evolution of more sophisticated data visualization and analysis techniques.

Introduction
The global change research community has a critical need for current and comprehensive information on the Earth's land cover (Townshend *et al.*, 1991; IGBP, 1992). Both the National Research Council and the International Geosphere-Biosphere Programme (IGBP) have declared that global land-cover data are foremost priorities (U.S. National Academy of Science, 1990; IGBP, 1990). Traditionally, land-cover data have been used for extrapolation of site-specific field data into a global spatial context. More recently, land-cover data were used in the development of process-level models at individual sites and then were applied to the parameterization of algorithms for global analyses. Land cover is viewed as an essential element in biophysical remote sensing methods in which landscapes are stratified into enumeration units so that inversion techniques can be used to estimate regional biophysical parameters (i.e., leaf area index, evapotranspiration, and net primary production).

Investigators have explored methods for generating land-cover information from Advanced Very High Resolution Radiometer (AVHRR) data aboard the National Oceanic and Atmospheric Administration's (NOAA) TIROS series of satellites (Goward *et al.*, 1985; Townshend *et al.*, 1991; Loveland *et al.*, 1991; IGBP, 1992). Although the AVHRR sensor has restricted spectral and spatial resolution compared with the sensors on board the Landsat and SPOT satellites, it possesses temporal resolution advantageous for analyses of global land cover. A location on the Earth's surface is observed twice daily by the AVHRR sensor (once during daylight). The relatively coarse (1 km) spatial resolution of AVHRR data yields a smaller, manageable volume of data for global analyses. For land observations, AVHRR channels 1 and 2 are commonly used to compute a greenness index such as the normalized difference vegetation index (NDVI) (Goward, 1989). Composite images showing the maximum value of the NDVI for each pixel during a multiple-day period (usually 10 or 14 days) result in frequent clear observations of the Earth's surface (Holben, 1986; Eidenshink, 1992).

AVHRR "greenness data" such as the NDVI are useful for depicting change in vegetative activity over time (Goward *et al.*, 1985; Townshend *et al.*, 1987). Land-cover classification and characterization with AVHRR data is often based on associations between land-cover types and variations in periodic observations of greenness through one or more growing cycles (Lloyd, 1990). However, problems exist in discriminating between land-cover types exhibiting similar phenologies (Townshend *et al.*, 1991).

Research, conducted jointly by the U.S. Geological Survey's EROS Data Center and the Center for Advanced Land Management Information Technologies, recently demonstrated that multitemporal AVHRR data, supplemented by ancillary data and analyzed in a structured manner, can be used effectively to characterize land cover (Loveland *et al.*, 1991). A set of 28-day maximum NDVI composite images covering the conterminous United States were clustered into 70 spectral-temporal classes and subsequently stratified, refined, and labeled using ancillary data. The result was a 159-class land-cover characteristics database.

The U.S. study has illuminated many issues that must be addressed in large-area land-cover characterization founded on analysis of coarse-resolution satellite image data. Such efforts require the use of "multisource data." Multisource data include digital images obtained through satellite remote sensing and data from ancillary sources, including terrain, soils, ecoregion, meteorological, and climatic data sets. The U.S. prototype efforts define and illustrate concepts and methods central to land-cover classification at continental and global scales.

The principal objectives of this research are to

Jesslyn F. Brown
Center for Advanced Land Management Information Technologies, Conservation and Survey Division, University of Nebraska-Lincoln, Lincoln, NE 68588-0517.
Thomas R. Loveland
U. S. Geological Survey, EROS Data Center, Sioux Falls, SD 57198
James W. Merchant
Center for Advanced Land Management Information Technologies, Conservation and Survey Division, University of Nebraska-Lincoln, Lincoln, NE 68588-0517.
Bradley C. Reed
Donald O. Ohlen
Hughes STX Corporation, EROS Data Center, Sioux Falls, SD 57198.

Photogrammetric Engineering & Remote Sensing, Vol. 59, No. 6, June 1993, pp. 977–987.

0099-1112/93/5906–977$03.00/0

- establish the conceptual basis and requirements for using multisource data in global land-cover characterization,
- detail multisource data analysis methods used in the prototype effort to characterize land cover for the conterminous United States,
- evaluate the potential for employing such methods in global land-cover characterization, and
- identify needs for future research and data set development.

Background

Remote sensing specialists have long recognized the important role of ancillary data in image interpretation (Campbell, 1978; Townshend and Justice, 1981; Estes *et al.*, 1983). During the last two decades, significant advances have been made in developing techniques and strategies for using ancillary data to improve the results of satellite image analysis. Most efforts have focused on Landsat digital image classification for land-cover mapping (Hutchinson, 1982). The vast majority of research using ancillary data for improving digital image classification has dealt with applications of digital terrain data in classification refinement. Commonly, terrain data are used either to adjust image brightness values for changes in radiance due to rugged relief (Jones *et al.*, 1988), or to sort and (or) subdivide spectrally derived classes to reduce confusion and error induced by the spectral inseparability of certain land-cover types. The latter application is of more concern here.

In one of the first demonstrations of classification improvement with ancillary data, Fleming and Hoffer (1979) used models of observed relationships between land cover, slope, aspect, and elevation to significantly improve a forest-cover map of an area in the southern Rocky Mountains derived from Landsat multispectral scanner (MSS) data. Miller and Shasby (1982) tested both subjective and quantitative techniques for deriving terrain-related decision criteria to enhance Landsat MSS-based maps of vegetation and forest fuels in Arizona and Montana. They reported significant increases in classification precision and a 20 percent improvement in classification accuracy when terrain data were used in post-classification data analysis.

Investigators have made efforts to employ other types of ancillary data in satellite image analysis. Pettinger (1982) used agricultural, upland and lowland environmental strata, and decision rules based on field research to adjust a Landsat MSS land-cover classification in Idaho. He showed that the stratification improved both the detail and the accuracy of the map. Cibula and Nyquist (1987) used terrain data and climatological data (precipitation and climate regimes) in a Landsat MSS classification of Olympic National Park, Washington. The ancillary data and decision logic were used to increase the number of unique land-cover classes from 9 to 21, resulting in an overall accuracy of 91.7 percent. Franklin (1989) used Landsat thematic mapper (TM) data, aeromagnetic, and geologic data in developing maps of geologic structure in central Newfoundland. He also described the joint use of terrain data (elevation, slope, aspect, and upslope and downslope convexity) and land systems (ecoregion) data in a Landsat MSS classification of Gros Morne National Park, Canada. He reported that classification accuracy improved from 40 to 85 percent when ancillary data were used.

As geographic information system (GIS) technology has developed, and integrated GIS-image analysis software has become more common, opportunities for using ancillary data and remotely sensed data in concert have expanded (Trotter, 1991). Innovations in data analysis based on expert systems and related techniques facilitate implementation of complex multisource data analysis strategies (Mason *et al.*, 1988; Srinivasan and Richards, 1990; Wang and Civco, 1992). Bolstad and Lillesand (1992), for example, demonstrated that such approaches improved Landsat TM classification of northern Wisconsin land cover by about 15 percent. Although improved software and analysis methods are now available, progress in using multisource data for land-cover classification has been slow. In most respects, researchers have only begun to realize the potential for combining remotely sensed and ancillary data in such efforts (Trotter, 1991).

Rationale for Using Multisource Data

Our contention is that land cover characterization employing AVHRR data, because of their coarse spatial resolution and constrained spectral resolution, requires the use of ancillary data. Limitations of the sensor are compounded by the fact that AVHRR satellite data are usually employed for observing very large areas such as continents. These large land masses possess enormous variation in terrain, climate, land use, and ecosystems. Even classifications based on multitemporal satellite data are insufficient because phenologic similarities between disparate cover types are common at continental scales.

Virtually all previous research on the use of ancillary data in land-cover classification has focused on relatively small areal domains (e.g., one satellite scene or a relatively small geographic region). A new and different set of problems are encountered in land classification when the objective is to map a continent where environmental complexity and diversity can be much greater than that dealt with in typical satellite remote sensing studies. A multisource approach is essential for land-cover characterization over such vast areas.

In addition, ancillary data can contribute more stable information to characterization, in contrast to satellite data, which fluctuate with the changing seasonal patterns of vegetation. Although the multitemporal NDVI data provide a useful measure of photosynthetic activity through time, allowing aggregation and regionalization of the land surface, additional (nonsatellite) information is needed to describe and refine land-cover units. A multisource approach, including both satellite and ancillary data, is crucial to the development, refinement, explanation, and information content of each land-cover class under analysis.

Campbell (1978) classified image interpretation procedures into five broad strategies. Four of these, he asserted, depend on the use of ancillary data for success. Most analyses of AVHRR data fall within Campbell's class of "probabilistic interpretations." He noted that, "because of the importance of non-image information ... in the application of a probabilistic interpretation process, relationships defined ... are valid only for a particular region" (Campbell, 1978, p. 267). Research in the conterminous United States supports this contention. When dealing with areas encompassing large latitudinal, elevational, climatic, and anthropogenic variation, continual adjustments of models, rules, and assumptions about relationships among image, landscape, and environment must be made (or, in other words, between image and ancillary data).

Experiences from the Conterminous U. S. Database

Creation of the prototype conterminous U.S. land-cover database included (Figure 1)

- processing and classifying AVHRR data,
- labeling and postclassification stratification of pixel clusters using ancillary data, and
- compiling the final land-cover characteristics database. Al-

TABLE 1. ANCILLARY DATA SETS, SOURCES, UNITS OF MEASUREMENT, AND USE OF THE DATA.

Data Set	Source	Units	Use
Elevation	DMA[1]	feet (20-ft resolution)	L[2]/S[3]
Climate	NOAA[4]	frost-free days	L/S
Ecoregions	USEPA[5]	Ecoregions attributes	L/S
MLRA[6]	SCS[7]	MLRA attributes	L
LULC[8]	USGS[9]	LULC attributes	L

[1]Defense Mapping Agency (1986)
[2]Labeling
[3]Stratification
[4]National Oceanic and Atmospheric Service (1979)
[5]U.S. Environmental Protection Agency (Omernik, 1987; Omernik and Gallant, 1990)
[6]Major land resource areas
[7]Soil Conservation Service (USDA SCS, 1981)
[8]Land use and land cover
[9]U.S. Geological Survey (USGS, 1986; Anderson et al., 1976)

TABLE 2. ATTRIBUTES ACCOMPANYING ECOREGION AND MAJOR LAND RESOURCE AREA (MLRA) DATA SETS.

Data set	Attributes
Ecoregions	Name Landform Potential natural vegetation Land use Soils
MLRA	Name Land use Elevation Topography Average annual precipitation Average annual temperature Average frost-free period Potential natural vegetation

though this paper is mainly concerned with the second step, an overview of the first step is necessary. A more detailed summary of processing and classifying AVHRR data is found in Loveland *et al.* (1991).

Processing and Classifying AVHRR Data

Daily 1-km AVHRR data from NOAA 11 were calibrated to reflectance, scaled to byte range, and georeferenced to the Lambert Azimuthal Equal Area map projection. Eight 28-day composites acquired from March to October 1990, based on maximum NDVI decision rules, were used as input to clustering and classification. A masking procedure ensured that classes exhibiting high intraclass variance, such as water, bare soil, clouds, snow, and ice, did not dominate the clustering process. An unsupervised clustering algorithm (Isoclass) and minimum-distance-to-mean classification were used to define 70 spectral-temporal (seasonally distinct) clusters. A 20 percent systematic sample of each of the eight composites was employed to derive initial cluster statistics. In all but ten cases, the clusters were not associated with a single cover type. Confusion occurred where several different land-cover types (i.e., agriculture and deciduous forest) were classified into one cluster. Ancillary data were then used to identify within-cluster confusion and to stratify clusters into distinct land-cover types.

Introduction of Ancillary Data

Ancillary data in a variety of forms, including discrete, thematic, and continuous, were used to augment database development, labeling, and postclassification stratification. The ancillary data included elevation, frost-free period, ecoregions, major land resource areas (MLRA), and land-use and land-cover data. These data sets, described in more detail by Loveland *et al.* (1991), are summarized in Table 1. The ecoregions and MLRA data contain associated attributes (Table 2), which were especially useful in the database development.

Application of Ancillary Data

In the conterminous U.S. research, ancillary data were found to contribute to classification and characterization of land cover in two ways: (1) for iterative labeling and describing clusters and classes, and (2) for postclassification refinement of spectral-temporal classes by spatial subdivision and merging (see Figure 1).

Preliminary labeling involved comparison of regional patterns exhibited by individual spectral-temporal classes with a wide variety of maps, images, and published data. Both traditional visual-subjective methods and less traditional digital-objective methods were used. Postclassification refinement involved "deterministic modeling" (Franklin, 1989). Visualization software was used to develop decision rules utilizing a combination of field experience and analysis of digital ancillary data similar to that used by Miller and Shasby (1982).

LABELING WITH ANCILLARY DATA.

Each of the original 70 clusters was evaluated in terms of its spatial distribution, phenology (as portrayed graphically by plotting the NDVI against time), and association with the ancillary data. In this way, descriptions for the initial 70 clusters identified clusters with confusion which required subdivision and indicated which types of ancillary data might provide useful separation criteria.

Information was extracted from the ancillary data layers for labeling using GIS overlay procedures. Each spectral-temporal cluster was intersected with the ancillary data layers, and counts or proportions for each association were compiled. For example, Figure 2 shows the distribution of cluster 35. Descriptive (tabular) attributes of the ancillary data (i.e., elevation, MLRA, and ecoregion) are attached as characteristics of the spectral-temporal cluster (Tables 3 to 7). These data provided the basis for class labeling and description (Table 8).

Label refinement occurred at several stages following the clustering of the multitemporal NDVI data. The intersection of each class with each ancillary data set provided additional information about each land-cover region, contributing to a convergence-of-evidence methodology. Graphic visualization of the data contributed to improved understanding of the classes as they were labeled and stratified.

POSTCLASSIFICATION REFINEMENT WITH ANCILLARY DATA.

After preliminary labeling, each cluster was evaluated visually to identify separation criteria. Using visualization techniques developed at the EROS Data Center, each cluster was observed with the associated distribution (histogram) of the elevation, ecoregions, and frost-free data. Subdivisions were made to avoid separating spatially contiguous pixels. Of the 59 spectral-temporal clusters that were subject to postclassification stratification, 27 were divided into two classes, 14

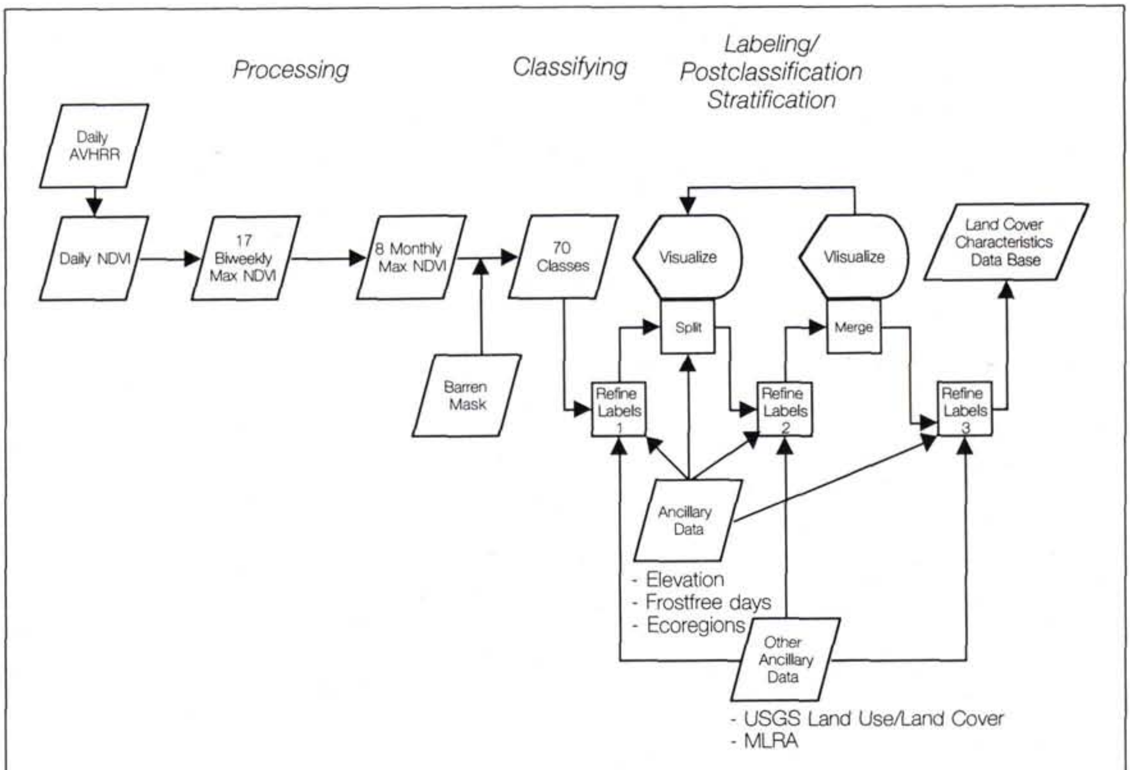

Figure 1. Processing flow for the conterminous U.S. land-cover characteristics database.

were split into three, and 18 were subdivided into four to six classes, resulting in 189 initially stratified classes.

The ecoregions data were used as the most frequently applied stratifier. These data provided separation rules for clusters 86 times as the sole separation criteria and 59 times in conjunction with either frost-free period or elevation criteria. The ecoregions data were especially useful in situations where a single spectral-temporal cluster represented several different land-cover types that were spatially separate. Elevation data were used 31 times as the sole separation criteria and 47 times in combination with the other two ancillary layers. Elevation data helped in situations where disparate land-cover types were in close proximity to each other. The continuous nature of the elevation data allowed interactive adjustment of the separation criteria. Altogether, elevation and ecoregions data were used, either alone or in combination, to solve 127 occurrences, or 67 percent, of cluster confusion during the conterminous U.S. data base construction.

Combining small (<1,000 pixels) land-cover classes was determined by spatial association and similar phenology as depicted in the NDVI multitemporal characteristics. Using these criteria, 70 cluster segments from different parent clusters were merged into 28 classes in the final database.

Land-Cover Issues and the Multisource Data Approach

It was apparent early in the U.S. land-cover work that common approaches for applying ancillary data in labeling and postclassification would not perform well. Within small geographic regions, elevation data can often be used to sort and (or) subdivide spectrally similar classes based on models of terrain-vegetation relationships. However, because of the large latitudinal range of the United States, a single elevation threshold could not be used to divide disparate land-cover types that had similar phenologies defined by their NDVI greenness curves and belonged to the same spectral-temporal class. As one moved northward, a given elevation threshold had to be adjusted "downslope" to compensate for the diminishing length of the growing season, making establishment of "global" models difficult and impractical. In many instances, the frost-free period was a more effective ancillary data source; however, it may be more difficult to obtain in global coverage than elevation data.

Experiences from the conterminous U.S. land characterization study suggest that classification confusion was most often related to climatic and anthropogenic factors. Climatic influences include the simple latitudinal and elevational limits on the growth of vegetation. Land-cover types that were often found to have similar spectral-temporal NDVI signatures included irrigated agriculture and forest, coastal wetlands and desert shrubland, and tundra and desert shrubland.

Four types of confusion were encountered: (1) those between natural landscapes, (2) those between natural and anthropogenic landscapes, (3) those between anthropogenic landscapes, and (4) undistinguished outliers. These types of problems probably will occur in any continental or global land-cover classification effort.

Natural Landscape Confusion.

Confusion between natural land-cover types is frequently influenced by climatic factors. Many of these problems are related to topography and are easily solved using digital elevation data, when available at the appropriate resolution. However, for mountain ranges that cover large latitudinal distances (e.g., the Rocky Mountains of North America or the Andes of South America), elevation influences vary. The

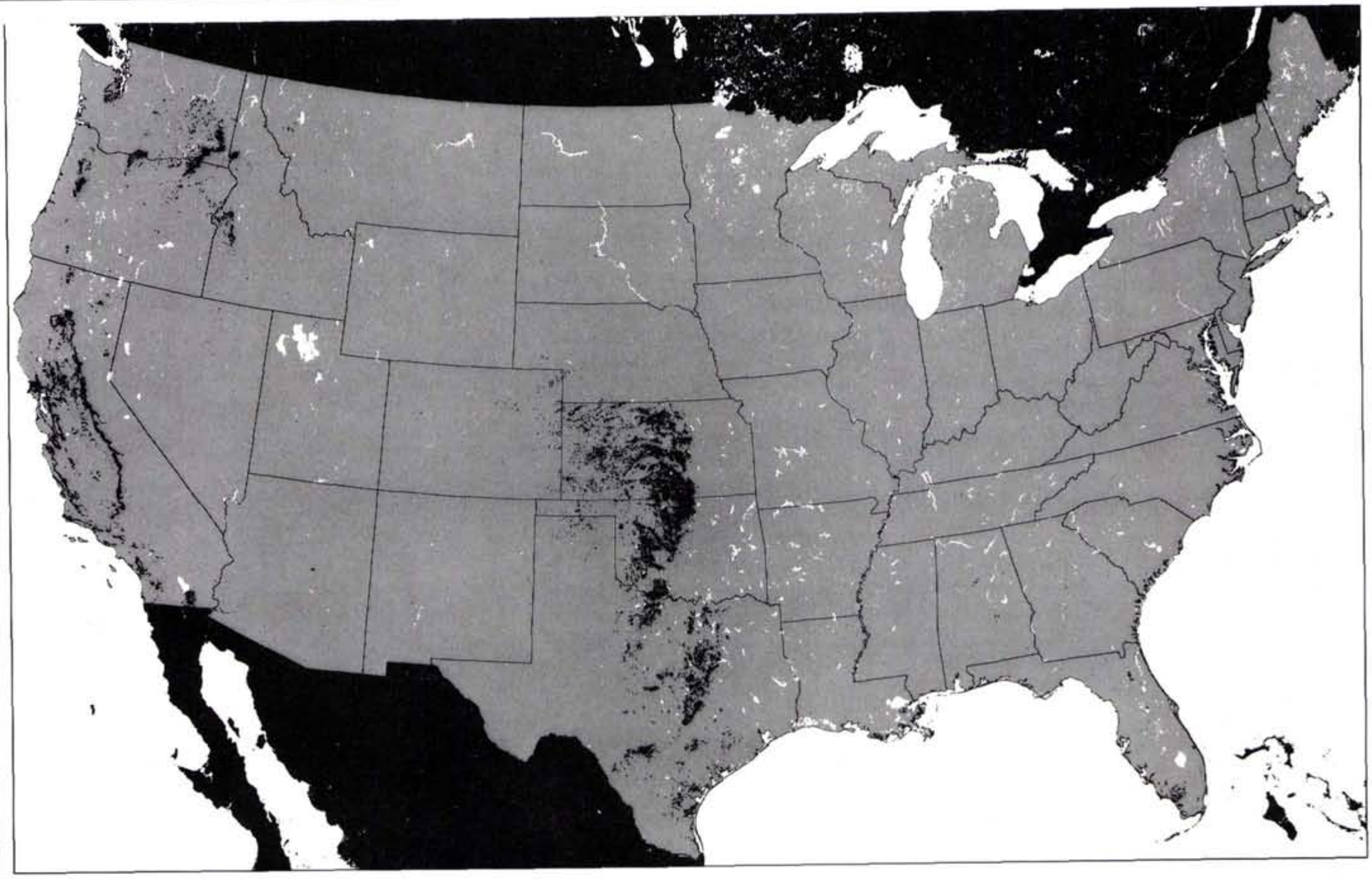

Figure 2. Spatial distribution of cluster 35.

Table 3. Percent Co-Occurrence of Cluster 35 within Selected Ecoregions.

Attribute	Attribute Value
Percent	46.6
Name	Central Great Plains
Landform	Irregular plains
PNV*	Bluestem/grama prairie, bluestem prairie
Land Use	Cropland, cropland with grazing
Soils	Dry mollisols
Percent	11.5
Name	Southern and central California plains and hills
Landform	Irregular plains, tablelands, low mountains
PNV*	California oakwoods, chaparral, California steppe
Land Use	Open woodland, grazing
Soils	Light colored soils of subhumid regions
Percent	7.5
Name	Texas Blackland Prairies
Landform	Irregular plains
PNV*	Bluestem, needlegrass, buffalo grass
Land Use	Cropland
Soils	Vertisols

* potential natural vegetation

frost-free period was found to represent adequately the influence of latitude on elevation zonation of vegetation in the conterminous U.S. study, and is likely to be useful in other regions of the world. Elevation data, used in combination with ecoregions, are a suitable substitute for the frost-free period. Common natural landscape confusion encountered included areas with low seasonal NDVI values such as tundra, desert shrubland, and coastal wetlands.

Natural-Anthropogenic Landscape Confusion.
Similarity between natural and human-influenced landscapes (i.e., wildland forests and grasslands versus agricultural land) was the most common landscape confusion in the U.S. database. It takes many forms and can be addressed using several ancillary variables. A common problem is agricultural vegetation sharing the same spectral-temporal profiles as natural vegetation. In some cases, altitude is a factor, while other cases are influenced by the spatial patterns of rainfall or soil temperatures. Altitude-related problems can be solved with digital elevation data. The others may be corrected with either climate variables or ecoregions.

Anthropogenic Landscape Confusion.
Land management and settlement practices cause complex problems in landscape confusion. If the confusion is interspersed, correcting it is nearly impossible. When the confusion is spatially separated, ecoregions are the most logical ancillary variable for postclassification stratification.

Undistinguished Outlier Confusion.
A certain amount of confusion is caused by image artifacts, mixed pixels, and atmospheric contamination. There are rarely good strategies to solve these types of problems.

Table 4. Percent Co-Occurrence of Cluster 35 with Land-Use and Land-Cover (LULC) Data by Anderson Level II Categories.

Percent	Category
78.00	Cropland
5.42	Evergreen forest land
5.18	Shrub and brush rangeland

Methods for Solving Land-Cover Confusion in the Conterminous United States

The following three examples illustrate working solutions for landscape confusion encountered in characterization of the conterminous United States. Original cluster 35, an example of natural-anthropogenic landscape confusion, was described as mostly winter wheat in the central United States and Pacific Northwest, but also included a significant area of cool season grasslands in the California hills. Labeling was based on ancillary data (see Tables 3 to 7) and the class multitemporal NDVI curve (Figure 3). The distribution of the cluster within several ecoregions with similar attributes indicated a suitable separation criteria. Pixels within five ecoregions (Plate 1), all located along the Pacific coast and containing open woodland or chaparral land cover, were separated by an overlay procedure to form new class 150, and the remaining pixels became new class 13. Class 150 was subsequently labeled as containing annual grasses, manzanita, oak, and white pine; class 13 was described as winter wheat.

In another example of natural-anthropogenic confusion, cluster 57 was identified initially as agricultural land cover

Table 5. Percent Co-Occurrence of Cluster 35 within Selected Major Land Resource Areas.

Attribute	Attribute Value
Percent	17.8
Name	Central Rolling Red Prairies
Land Use	Range/grazing (40%); crop/w.wheat (20%); woodland/urban/pasture (20%)
Elevation	300–500m
Topography	Dissected plain, undulating to gently rolling hills
AAP[1]	625–900mm; maximum in spring
AAT[2]	14–18°C
AFFP[3]	190–230 days
PNV[4]	Mixed prairie (indiangrass, bluestem); trees/shrubs
Percent	9.9
Name	Central Rolling Red Plains
Land Use	Range/grazing (60%); crop/w.wheat/sorg. (35%); irrig. (5%)
Elevation	500–900m
Topography	Dissected plains
AAP[1]	500–750mm; maximum in spring
AAT[2]	14–18°C
AFFP[3]	185–230 days
PNV[4]	Mid/tall grasses (bluestem, sand sagebrush, gramas)
Percent	6.7
Name	Texas Blackland Prairie
Land use	Crops/cotton/sorghum (40%); pasture (45%)
Elevation	100–200m
Topography	Level to gently rolling dissected plain
AAP[1]	750–1150mm; maximum in spring and fall
AAT[2]	17–21°C
AFFP[3]	230–280 days
PNV[4]	Prairie (bluestem) with oak/elm savanna along river
Percent	5.2
Name	Sierra Nevada foothills
Land use	Grassland (75%); dryland ag. (5%); brush/open forest (20%)
Elevation	200–500m with peaks to 1200m
Topography	Rolling to steep dissected hills and low mts.
AAP[1]	350–900mm; dry hot summers, cool moist winters
AAT[2]	13–18°C
AFFP[3]	200–320 days
PNV[4]	Annual grasses, shrubs (chamise, manzanita), trees (oak/pine)

[1]average annual precipitation
[2]average annual temperature
[3]average frost-free period
[4]potential natural vegetation

Table 6. Frost-Free Period Statistics for Cluster 35.

Statistic	Days
Mean	209
Std. Dev.	44
Maximum	338
Minimum	48
Mode	180
Median	202

Table 7. Elevation Statistics for Cluster 35.

Statistic	Feet
Mean	1,469
Std. Dev.	897
Maximum	7,068
Minimum	1
Mode	200
Median	1,361

Table 8. Description of Original Cluster 35.

35.0 Class 35
- 35.1 General—main area is in Kansas and Oklahoma. Some in Washington, Oregon, and California.
- 35.2 Groups
  - 35.2.1 Kansas/Oklahoma—winter wheat with some rangeland.
  - 35.2.2 Oregon/Washington—eastern parts of States near Columbia River. Winter wheat in east. Also found in Willamette Valley.
  - 35.2.3 California—cool season grasses.

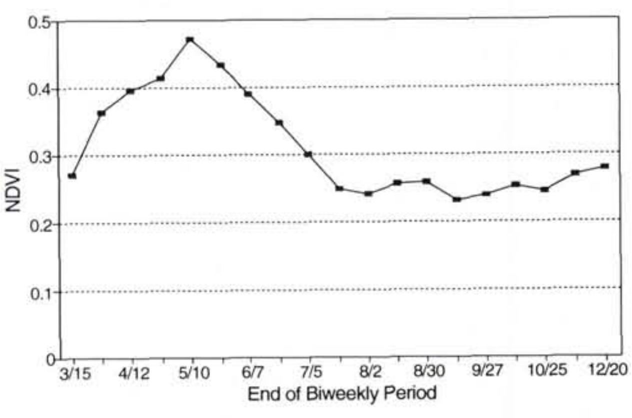

Figure 3. Multitemporal normalized difference vegetation index plot for cluster 35.

in the Midwest, but mostly coniferous forest elsewhere. The distribution of elevation for original cluster 57 (Plate 2) was positively skewed toward low elevations (less than 1,500 ft). Interactive visualization of the data permitted examination of possible subdivision thresholds and their resulting spatial distributions. Accordingly, spatial context was considered along with the ancillary data for each stratification decision. Cluster 57 was subsequently subdivided using a 3,000 ft elevation threshold and divided into two new classes: class 156, which contains lodgepole pine, Douglas fir, and aspen, and class 34, containing wheat, soybeans, corn, and pasture (see Plate 2).

Figure 4 shows an example of multistage stratification using ancillary data applied to a more complicated form of cluster confusion. Original cluster 58 presented a challenge to label and stratify successfully because it contained several kinds of landscape confusion. The cluster had portions of natural, anthropogenic, and mosaic landscapes, as well as undistinguished outliers. Using a three-step process, cluster 58 was subdivided into six classes in the final data base (Table 9). First, class 55 was subdivided using the ecoregions data (see Figure 4, e); pixels within four ecoregions located in the southeastern plains and mid-Atlantic and Gulf coastal plains were included in this class. Class 55 was labeled a cropland/woodland mosaic, with predominant vegetation including oak, pine, soybeans, corn, cotton, and peanuts (see Table 9). The rest of the pixels of cluster 58 were stratified in two stages (see Figure 4), first into two areas and second into smaller regions, some of which were merged with parts of other clusters to form the final class structure.

Classes 47, 117, and 93 were stratified using both ecoregions and elevation (see Figure 4, de). As a group, they were either within the ecoregions of the Northwest coastal mountains or between 5,500 and 12,000 ft in elevation in remaining ecoregions (except within those defining class 55). Class 47 was further stratified by the ecoregions of the central and northeastern plains. This cropland/woodland class was dominated by corn, soybeans, sorghum, and mixed woodlots (see Table 9). Class 93 was further divided by the ecoregions of the Appalachian Mountains and northeastern upland regions. This class was labeled southeast deciduous forest, consisting mostly of oak, hickory, and some mixed cropland. Class 93 was small and was subsequently merged with portions of two other parent clusters. Class 117, western coniferous forest, consisted of all the pixels from cluster 58 within the Northwest coastal range ecoregions above 5,500 ft in elevation, mostly in the Rocky Mountains. This coniferous forest class was dominated by ponderosa pine, lodgepole pine, western white pine, and Douglas fir (see Table 9).

The remaining two classes, 36 and 136, were initially stratified together using elevation (less than 5,500 ft) and ecoregions other than the Northwest coast and the southeastern plains (see Figure 4, de). Class 136 was subdivided by the ecoregions in the northwest Great Lakes and northeast highlands regions. This class merged with another cluster segment and is a mixed forest composed mostly of oak, maple, ash, jack pine, and red pine. The final cluster segment, class 36 (cropland and pasture), consisted mainly of corn, soybeans, and hay pasture (see Table 9) and was contained within the northwest Great Lakes ecoregions.

The above procedures demonstrate flexible and interactive approaches to land-cover regionalization combining clustering procedures with postclassification stratification and merging. The process developed for the conterminous U.S. land database provides a methodology that can be applied to global land characterization, given the availability of appropriate satellite and ancillary data.

Data Requirements and Sources for Global Land-Cover Characterization

The global change research community is increasingly organized in their call for the development of global land-cover databases. The IGBP, for example, has identified land-cover data as a top research priority (IGBP, 1990). The IGBP has further clarified this requirement by specifying that 1-km AVHRR data are the logical basis for global land-cover data sets (IGBP, 1992). An initiative is underway to organize the multitemporal 1-km AVHRR data needed for a global land-cover mapping effort. However, a parallel effort is needed to organize the multisource data required to permit accurate global land-cover characterization.

Class labeling requires data that support the identification of land cover, vegetation, and other environmental characteristics of individual land-cover regions. For this purpose, consistent, detailed, and spatially comprehensive vegetation, land-cover, and soils maps are especially useful. However, such data with continental coverage are rare. Although not ideal, local or regional maps are valuable, particularly if applied to visual interpretation rather than automated class labeling. Because methods of visual interpretation for class labeling are more flexible and adaptive, data requirements are less rigid. However, additional data may be needed for the more rigorous postclassification refinement application.

Postclassification refinement requires more control and consistency in ancillary data. These data must (1) span the entire continent; (2) serve as surrogates for the climatic, ecological, or anthropogenic factors that create spectral or multitemporal class confusion; (3) contain the qualitative or quantitative information that correlates to the sources of confusion; and (4) have compatible geographic scale and locational accuracy.

Global Data Requirements

The minimum set of ancillary variables needed for postclassification refinement contains digital elevation data and ecoregions. Both data types are likely to be available for continental mapping. Climate variables, particularly the frost-free period and monthly precipitation, are also desirable but are less likely to be available in a suitable form.

Digital Elevation Data.

A global digital elevation data set, ETOPO5, is currently available from NOAA's National Geophysical Data Center. However, because of its approximate resolution of 10 km (based on a cell size of 5 arc minutes), it is inadequate for representing vegetation patterns related to elevation zonation in irregular terrain. Several organizations have programs underway to develop 1-km global digital elevation models (DEM). The U.S. Geological Survey, with the National Aeronautic and Space Administration, is investigating several options for generating a global DEM. The Committee on Earth Observation Satellites is also promoting an multiagency effort to prepare a global DEM. It is likely that appropriate elevation data soon will be available for a global land-cover mapping effort.

Ecoregions.

There are several global ecoregion frameworks that can be used for land-cover characterization. The U.S. Forest Service has developed ecoregions based on climate, climax vegetation, and soils (Bailey, 1989; Bailey and Cushwa, 1981). Researchers at Moscow State University have mapped two

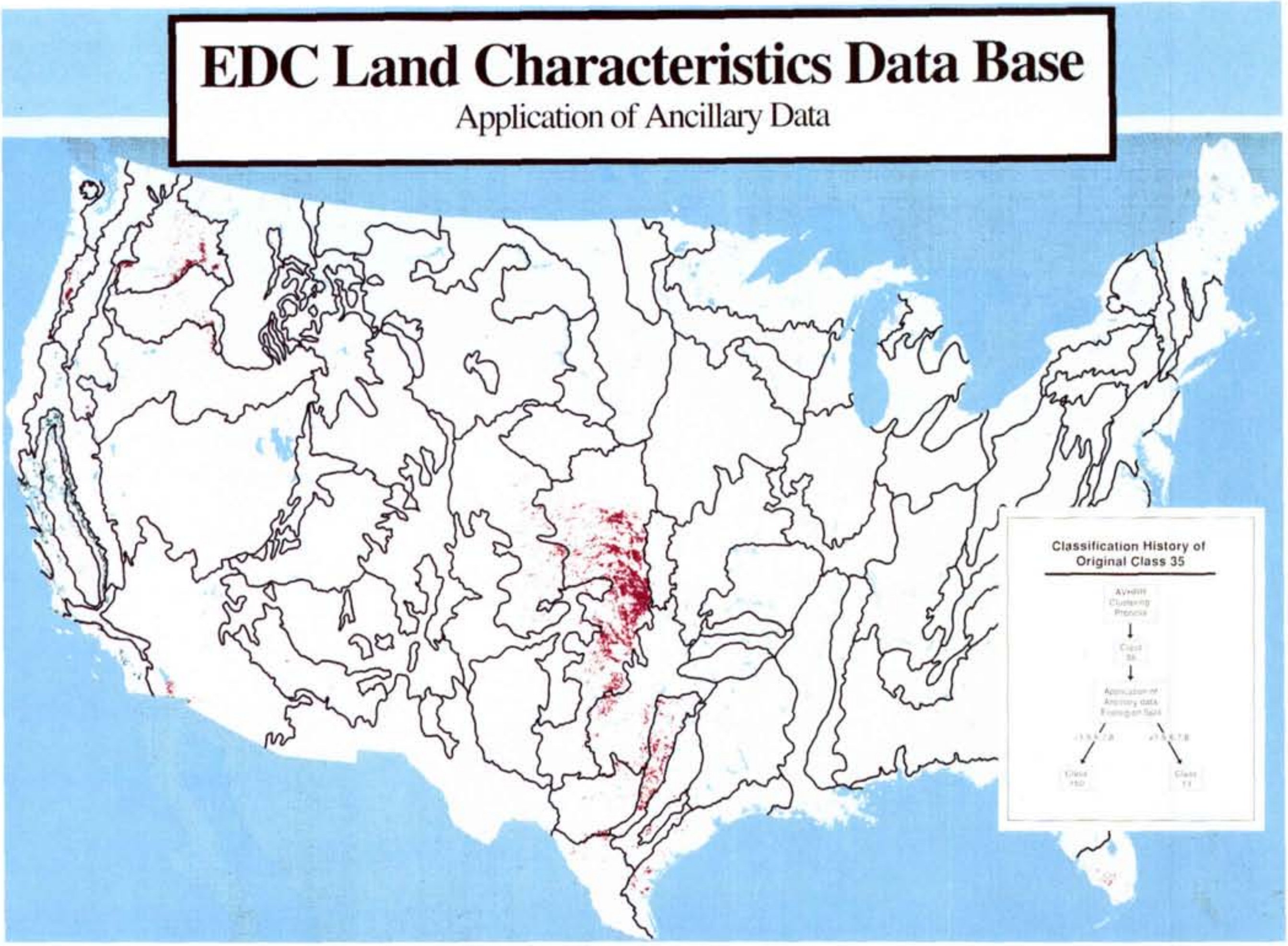

Plate 1. Postclassification stratification of cluster 35 into class 13 (green) and class 150 (red).

ecoregion frameworks for the globe. One is based on climate, soils, and potential natural vegetation, and the other includes those plus current land use (Alekseyev *et al.*, 1988). A drawback of these products is that their minimum mapping unit (map scale 1:15,000,000) is more coarse than 1-km AVHRR data. Small-scale data, however, may be the best available sources. Although there are excellent national ecoregion interpretations, it is unlikely that they can be combined to form consistent continental coverage. However, these data can contribute effective labeling information.

Climate Variables.

To use climate data for postclassification stratification, it is necessary to spatially interpolate surfaces for each variable. Adequate representation of precipitation, temperature, or frost-free period require enough stations so that major regional patterns are depicted. Global monthly temperature, precipitation, and atmospheric pressure climatic data are available (Vose *et al.*, 1992); however, the spatial distribution of available stations may not be adequate for postclassification analyses at a 1-km resolution.

Global Data Sources

The United Nations Environment Programme Global Resource Information Database is one source for global and national data sets (UNEP, 1990). Many countries—including Australia, Canada, and the former Soviet Union—have thematic data sets equivalent to the ecoregions data used in this research (Omernik, 1987). Ancillary data sets are available, both in global coverages (Matthews, 1985) and national coverages (for example, Canada (Wiken, 1986; Ecoregions Working Group, Canada Committee on Land Classification, 1989) and Australia (Walker *et al.*, 1985)).

Summary and Conclusions

Ancillary data provide information crucial to the success of large-area land-cover characterization based on NOAA AVHRR data. They contribute essential evidence for labeling and refining land-cover classes where differing types are represented by a single spectral-temporal signature. The attributes of ancillary data, used in conjunction with the spatial distribution of a spectral-temporal region, are the basis

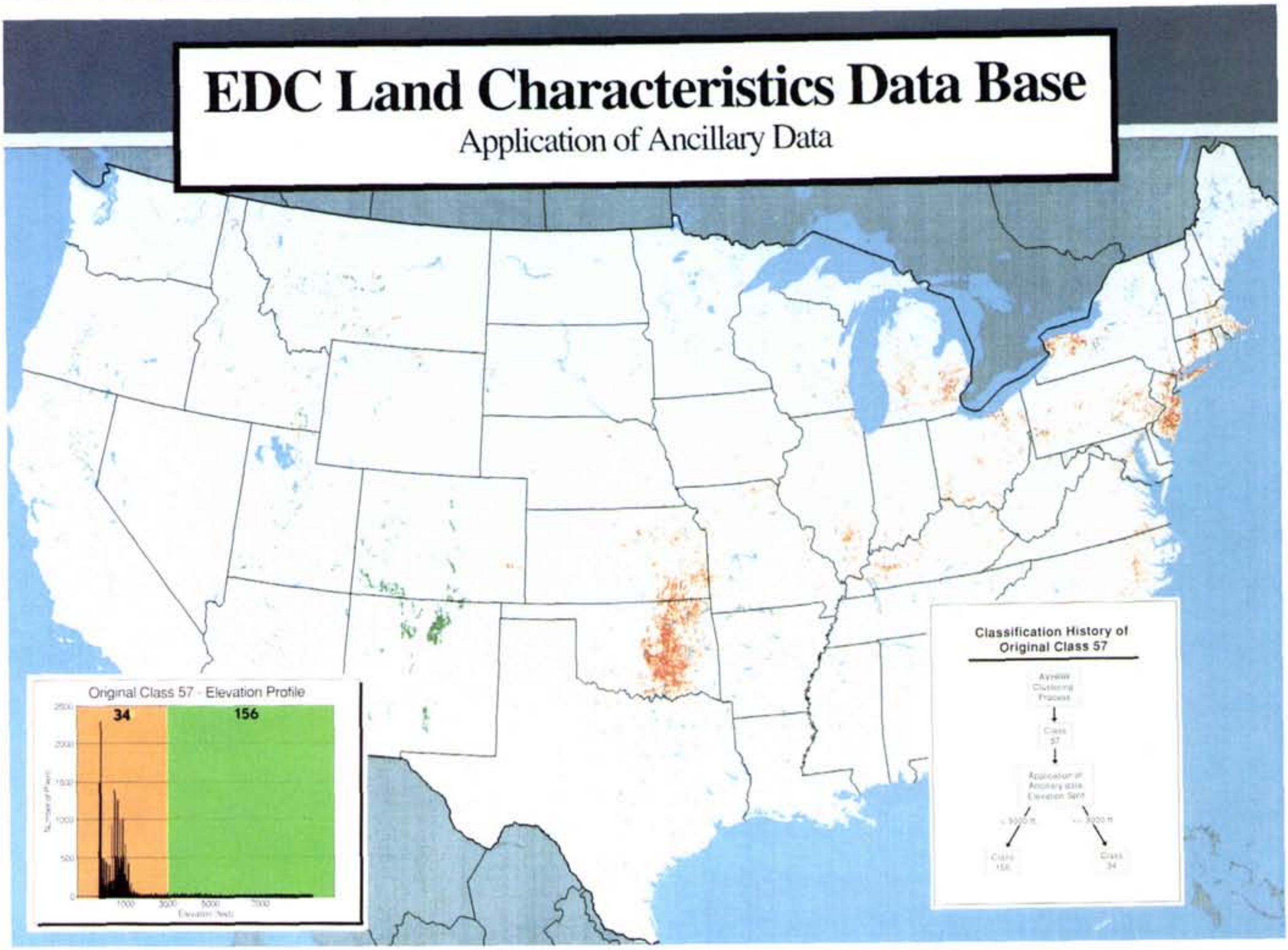

Plate 2. Postclassification stratification of cluster 57 into class 34 (orange) and class 156 (green).

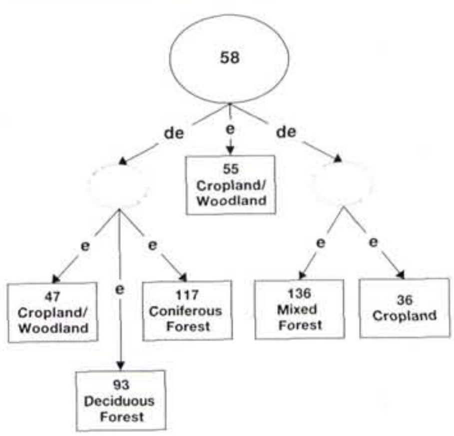

Figure 4. Postclassification processing of cluster 58.

for selecting the appropriate ancillary variable(s), choosing the most reasonable threshold(s), and making a division.

The challenge of characterizing the land surface at global scales is to analyze larger areas while extracting essential information from multisource data. Processing of global data sets presents new computational and interpretational challenges. Large areas possess great variation in climate, terrain,

Table 9. Final Class Descriptions for Cluster 58 Segments.

Class No.	Cover type	Primary vegetation types
36	Cropland/pasture	Corn, soybeans, pasture/hay
47	Cropland/woodlots	Corn, soybeans, sorghum, mixed-woodlots
55	Woodland/cropland	Mixed-oak, pine, soybeans, corn, cotton, peanuts
93	Mixed-forest/crop	Oak, hickory, mixed-pine, mixed-crops
117	Western-conifer	Western white, ponderosa, lodgepole, douglas fir
136	Northern-forest	Oak, maple, ash, jack-pine, red-pine

and vegetation, compounding spectral-temporal confusion among disparate land-cover types. Because a global effort to characterize land cover will probably incorporate a continental strategy, continents crossing the equator (i.e., Africa and South America) may pose more complex problems. Similar vegetation mosaics affected by different growing seasons in the Southern and Northern Hemispheres may exhibit significantly different spectral-temporal signatures; however, the basic types of landscape confusion discussed in this paper are expected.

As larger areas are studied, more advanced methods will be necessary to understand satellite and ancillary data relationships. New techniques in data exploration and visualization will permit analysis of ancillary data and may reveal complex relationships in the data that are vital in explaining land-cover characteristics. Although automated techniques may afford simpler, more efficient methods for doing such work, the interaction of the analyst remains the key to the process of postclassification stratification of large areas because of the complexity of the decision process and the requirement for expert knowledge and reasoning. Therefore, methods should be developed that assist rather than replace the analyst.

The success of global land-cover analysis, central to future global change research, will depend on (1) the availability of global satellite coverage, (2) the quality and level of information within available ancillary data to solve areas of confusion, and (3) the evolution of improved techniques for extracting information from both types of data.

Acknowledgments

The first author, Jesslyn F. Brown, performed this work under U. S. Environmental Protection Agency grant (X0075626-01) while a Visiting Scientist at EROS Data Center. The Hughes STX Corporation performed this work under U. S. Geological Survey contract 1434-92-C-40004.

References

Alekseyev, B., A. Vasilyeva, Y. Golubchikov, L. Kurakova, Y. Lukashova, Y. Milanova, L. Mikhailova, E. Romanova, A. Ryabchikov, N. Sen'kovshaya, V. Uledov, and L. Frolova, 1988. *Geographic Belts and Zonal Types of Landscapes of the World* (translated by A. Kushlin): U.S.S.R. Main Administration for Geodesy and Cartography and Moscow State University, Moscow, map scale 1:15,000,000.

Anderson, J.F., 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:* U.S. Geological Survey Professional Paper 964, 28 p.

Bailey, R.G., 1989. Explanatory Supplement to Ecoregions Map of the Continents: *Environmental Conservation*, Vol. 16, No. 4, pp. 302–309.

Bailey, R.G., and C.T. Cushwa, 1981. *Ecoregions of North America:* Map. U.S. Fish and Wildlife Service, Office of Biological Services, Eastern Energy and Land Use Team, map scale 1:12,000,000.

Bolstad, P.V., and T. M. Lillesand, 1992. Rule-Based Classification Models: Flexible Integration of Satellite Imagery and Thematic Spatial Data: *Photogrammetric Engineering & Remote Sensing*, Vol. 58, No. 7, pp. 965–971.

Campbell, J.B., 1978. A Geographical Analysis of Image Interpretation Methods: *The Professional Geographer*, Vol. 30, No. 3, pp. 264–269.

Cibula, W.G., and M.O. Nyquist, 1987. Use of Topographic and Climatological Models in a Geographical Data Base to Improve Landsat MSS Classification for Olympic National Park: *Photogrammetric Engineering & Remote Sensing*, Vol. 53, No. 1, pp. 67–75.

Defense Mapping Agency, 1986. *Defense Mapping Agency Product Specifications for Digital Terrain Elevation Data (DTED)*, (2d ed.): April, 1986.

Ecoregions Working Group, Canada Committee on Land Classification, 1989. *Ecoclimatic Regions of Canada:* Ecological Land Classification Series No. 23, Environment Canada, Ottawa, Ontario.

Eidenshink, J., 1992. The 1990 Conterminous U.S. AVHRR Data Set: *Photogrammetric Engineering & Remote Sensing*, Vol. 58, pp. 809–813.

Estes, J.E., E.J. Hajic, and L.R. Tinney, 1983. Fundamentals of Image Analysis - Analysis of Visible and Thermal Infrared Data, *The Manual of Remote Sensing* (R.N. Colwell, editor), Falls Church, Virginia, American Society of Photogrammetry and Remote Sensing, Vol. 1, pp. 987–1124.

Fleming, M.D., and R.M. Hoffer, 1979. Machine Processing of Landsat MSS Data and DMA Topographic Data for Forest Cover Type Mapping: *Proceedings of the 1979 Machine Processing of Remotely Sensed Data Symposium*, West Lafayette, Indiana, Purdue University/LARS, pp. 377–390.

Franklin, S.E., 1989. Ancillary Data Input to Satellite Remote Sensing of Complex Terrain Phenomena: *Computers and Geosciences*, Vol. 15, No. 5, pp. 799–808.

Goward, S.N., 1989. Satellite Bioclimatology: *Journal of Climate*, Vol. 2, pp. 710–720.

Goward, S.N., C.J. Tucker, and D.G. Dye, 1985. North American Vegetation Patterns Observed with the NOAA-7 Advanced Very High Resolution Radiometer: *Vegetatio*, Vol 64, pp. 3–14.

Holben, B., 1986. Characteristics of Maximum Value Composite Images from Temporal AVHRR Data: *International Journal of Remote Sensing*, Vol. 7, pp. 1417–1434.

Hutchinson, C., 1982. Techniques for Combining Landsat and Ancillary Data for Digital Classification Improvement: *Photogrammetric Engineering & Remote Sensing*, Vol. 48, No. 1, pp. 123–130.

International Geosphere-Biosphere Programme, 1990. *Global Change:* Report No. 12. Stockholm, Sweden: IGBP Secretariat.

———, 1992. *Improved Global Data for Land Applications:* IGBP Report No. 20. Stockholm, Sweden: IGBP Secretariat.

Jones, A.R., J.J. Settle, and B.K. Wyatt, 1988. Use of Digital Terrain Data in the Interpretation of SPOT-1 HRV Multispectral Imagery: *International Journal of Remote Sensing*, Vol. 9, No. 4, pp. 669–682.

Lloyd, D., 1990. A Phenological Classification of Terrestrial Vegetation Cover using Shortwave Vegetation Index Imagery: *International Journal of Remote Sensing*, Vol. 11, No. 12, pp. 2269–2279.

Loveland, T.R., J.W. Merchant, D.O. Ohlen, J.F. Brown, 1991. Development of a Land-Cover Characteristics Database for the Conterminous U.S.: *Photogrammetric Engineering & Remote Sensing*, Vol. 57, pp. 1453–1463.

Mason, D.C., D.G. Corr, A. Cross, D.C. Hogg, D.H. Lawrence, M. Petrou, and A.M. Tailor, 1988. The Use of Digital Map Data in the Segmentation and Classification of Remotely-Sensed Images: *International Journal of Geographical Information Systems*, Vol. 2, No. 3, pp. 195–215.

Matthews, E., 1985. *Atlas of Archived Vegetation, Land Use and Seasonal Albedo Data Sets:* NASA Technical Memorandum 86199, Washington, D.C.

Miller, W.A., and M.B. Shasby, 1982. Refining Landsat Classification Results Using Digital Terrain Data: *Journal of Applied Photographic Engineering*, Vol. 8, No. 1, pp. 35–40.

National Oceanic and Atmospheric Service, 1979. *Climatic Atlas of the United States:* U.S. Department of Commerce, National Oceanic and Atmospheric Administration, Environmental Data Services, Asheville, North Carolina, 80 p.

Omernik, J.M., 1987. Ecoregions of the Conterminous United States:

Annals of the Association of American Geographers, Vol. 77, No. 1, pp. 118–125.

Omernik, J.M., and A.L. Gallant, 1990. Defining Regions for Evaluating Environmental Resources: *Proceedings of the Global Natural Resources Monitoring and Assessment Symposium*, Bethesda, Maryland, pp. 936–947.

Pettinger, L.R., 1982. *Digital Classification of Landsat Data for Vegetation and Land-Cover Mapping in the Blackfoot River Watershed, Southeastern Idaho:* U.S. Geological Survey Professional Paper 1219, 33 p.

Srinivasan, A., and J.A. Richards, 1990. Knowledge-Based Techniques for Multi-Source Classification: *International Journal of Remote Sensing*, Vol. 11, No. 3, pp. 505–525.

Townshend, J.R.G., and C.O. Justice, 1981. Information Extraction from Remotely Sensed Data - A User View: International Journal of Remote Sensing, Vol. 2, No. 4, pp. 313–329.

Townshend, J.R.G., C.O. Justice, and V. Kalb, 1987. Characterization and Classification of South American Land Cover Types: *International Journal of Remote Sensing*, Vol. 8. No. 8, pp. 1189–1207.

Townshend, J.R.G., C.O. Justice, W. Li, C. Gurney, and J. McManus, 1991. Global Land Cover Classification by Remote Sensing: Present Capabilities and Future Possibilities: *Remote Sensing of Environment*, Vol. 35, pp. 243–255.

Trotter, C.M., 1991. Remotely-sensed Data as an Information Source for Geographical Information Systems in Natural Resources Management - A Review: *International Journal of Geographical Information Systems*, Vol. 5, No. 2, pp. 225–239.

United Nations Environment Programme, 1990. *GRID: Global Resource Information Database:* Nairobi, Kenya, 16 p.

U.S. Department of Agriculture, Soil Conservation Service, 1981. *Land Resource Regions and Major Land Resource Areas of the United States:* Agriculture Handbook 296, Washington, D.C.

U.S. Geological Survey, 1986. *Land Use and Land Cover Digital Data from 1:250,000- and 1,100,000-Scale Maps:* U.S. Geological Survey Data Users Guide No. 4, Reston, Virginia.

U.S. National Academy of Sciences, 1990. *Research Strategies for the U.S. Global Change Research Program:* National Academy of Sciences National Research Council, National Academy Press, Washington, D.C., 219 p.

Vose, R. S., R. Heim, R.L. Schmoyer, T.R. Karl, P.M. Steurer, J.K. Eischeid, and T.C. Peterson, 1992. *The Global Historical Climatology Network: Long-Term Monthly Temperature, Precipitation, Sea Level Pressure, and Station Pressure Data:* Oak Ridge National Laboratory, Environmental Sciences Division, Publication No. 3912, p. 315.

Walker, P., K. Cocks, and M. Young, 1985. Regionalizing Continental Data Sets: *Cartography*, Vol. 14, No. 1, pp. 66–73.

Wang, Y., and D.L. Civco, 1992. Post-Classification of Misclassified Pixels by Evidential Reasoning - A GIS Approach for Improving Classification Accuracy of Remote Sensing Data: *Technical Papers of ASPRS/ACSM/RT 92*, Vol. 4, Bethesda, Maryland, pp. 160–170.

Wiken, E., 1986. *Terrestrial Ecozones of Canada:* Ecological Land Classification Series No. 19. Environment Canada, Ottawa, Ontario.

(Received 4 December 1992; accepted 21 January 1993)

Jesslyn F. Brown

Jesslyn F. Brown is a Research Associate in the Conservation and Survey Division, Institute of Agriculture and Natural Resources, University of Nebraska-Lincoln. Since completing her Masters program in Geography in 1990, she has been a Visiting Scientist at the EROS Data Center in South Dakota. Her current research includes the development and validation of large-area land-cover characteristics data sets using satellite and ancillary data.

Thomas R. Loveland

Tom Loveland heads the Land Sciences research program of the U.S. Geological Survey's EROS Data Center. He is responsible for research projects dealing with the relationship of satellite imagery to land surface properties. His personal research deals with the use of coarse resolution satellite imagery for regional land characterization for land process modeling. Prior to joining the USGS, he worked in land-cover mapping and GIS programs with state governments in South Dakota and Arizona. Mr. Loveland received a B.S. and M.S. in Geography from South Dakota State University, Brookings, South Dakota.

James W. Merchant

James W. Merchant is Associate Professor in the Conservation and Survey Division, Institute of Agriculture and Natural Resources, University of Nebraska-Lincoln where he is Associate Director of the Center for Advanced Land Management Information Technologies (CALMIT). His current research interests include (1) land-cover characterization and landscape regionalization using AVHRR and ancillary data, (2) landscape edge representation in digital satellite imagery, and (3) applications of visualization in GIS and digital image analysis.

Bradley C. Reed

Bradley C. Reed received his Ph.D. in Geography from the University of Kansas in 1990. He was Assistant Professor of Geography at New Mexico State University before taking his present position as Senior Scientist for Hughes STX Corporation at EROS Data Center. His current research involves analyzing interannual variability of vegetation and developing tools for data exploration and visualization.

Donald O. Ohlen

Donald O. Ohlen received a B.S. in Geography from South Dakota State University in 1976. He is currently a Senior Scientist with Hughes STX Corporation at the EROS Data Center, where he has been for 15 years. His research interests include image analysis and database development for spatial environmental analysis of global land processes.