Thematic Validation of High-Resolution Global Land-Cover Data Sets

Joseph Scepan

Abstract

This paper describes a procedure to validate the thematic accuracy of the International Geosphere-Biosphere Programme, Data and Information System (IGBP-DIS) DISCover (Version 1.0) 1-Kilometer Global Land-Cover Data Set. Issues of data set sampling design, image geometry and registration, and core sample interpretation procedures are addressed. Landsat Thematic Mapper and SPOT satellite image data were used to verify 379 primary core samples selected from DISCover 1.0 using a stratified random sampling procedure. The goal was to verify a minimum of 25 samples per DISCover class; this was accomplished for 13 of the 15 verified classes. Three regional Expert Image Interpreters independently verified each sample, and a majority decision rule was used to determine sample accuracy. For the 15 DISCover classes validated, the average class accuracy was 59.4 percent with accuracies for the 15 verified DISCover classes ranging between 40.0 percent and 100 percent. The overall area-weighted accuracy of the data set was determined to be 66.9 percent. When only samples which had a majority interpretation for errors as well as for correct classification were considered, the average class accuracy of the data set was calculated to be 73.5 percent.

Introduction

The focus of this research is the validation of the International Geosphere Biosphere Programme, Data and Information System (IGBP-DIS), DISCover (Version 1.0) 1-Kilometer Global Land-Cover Data Set. DISCover was developed in a cooperative effort by the U.S. Geological Survey EROS Data Center (USGS/EDC), Sioux Falls, South Dakota, and the European Commission's Joint Research Center, Ispra, Italy. Prior to this validation effort, there was no statistically based accuracy estimate for any global land-cover data set; DISCover 1.0 stands as the first such data set.

Global climate, ecological, and chemical cycle modeling are among the most important scientific techniques currently available to measure, monitor, and predict critical physical and biological processes and environmental changes in the Earth's environmental system. Global ecological models are developed and implemented to provide information about a variety of ecological and biogeochemical regimes. Important among these are carbon cycles, hydrologic cycles, and terrestrial energy balance (Tucker *et al.*, 1985). These processes are quite complex and must be addressed using numerical models.

Global ecological models require as inputs spatially referenced terrestrial vegetation and land-cover data sets. Requirements have been demonstrated for these data at spatial resolutions of 1 kilometer (and finer) with high temporal resolution (Lunetta *et al.*, 1991). To date, a number of global coverage land-cover data sets have been developed (Defries and Townshend, 1994). In the past, these data have been spatially and temporally incomplete, inadequate, and inaccurate (Townshend *et al.*, 1991). The DISCover land-cover data set and associated validation procedures were implemented, in part, because scientifically valid continental- or global-scale land-cover data sets of known accuracy did not exist (Estes and Mooneyhan, 1994; Townshend *et al.*, 1994).

The DISCover Data Set was assembled to meet data requirements for studies of climate, biogeochemical cycles, atmospheric chemistry, water, energy, vegetation, and ecosystems (Loveland and Belward, 1997). DISCover was compiled on an individual continental basis using data from the National Oceanographic and Atmospheric Administration (NOAA) Advanced Very High Resolution Radiometer (AVHRR) data collected daily through the efforts of a number of AVHRR receiving stations coordinated by USGS/EDC. A summary description of this data set, including processing and classification, is found in Eidenshink and Faundeen (1994) and Loveland *et al.* (1999, in this issue).

The material that follows provides a brief background on the DISCover data set and the data compiled during the validation effort. The techniques and methodologies employed in the core sample development are then presented. The procedures developed at a Verification Test Workshop and employed at the Global Validation Workshop (GVW) are also discussed. Results obtained from the conduct of the GVWare reviewed and analyzed and conclusions and recommendations are presented.

Background

The specific protocol for validation of DISCover 1.0 was developed by the IGBPValidation Working Group (VWG) and reviewed and approved by the Land Cover Working Group (LCWG). The protocol specified a stratified random sample design and a methodology that relies on testing the DISCover thematic classes against an independent data source, in this instance, higher spatial resolution satellite imagery. This general method of validating a data product by employing higher spatial resolution data is well established (Fitzpatrick-Lins, 1980; Rosenfield *et al.*, 1981) and has been employed extensively at local and regional scales (Borella *et al.*, 1982; Estes *et al.*, 1987). This study is the first application of such a technique at the global scale.

The core sample validation had the principal objective of providing relatively simple statistical statements of accuracy to (1) characterize the accuracy of the DISCover 1.0 product as a single data product and (2) estimate the error variance in areal totals of individual DISCover 1.0 land cover types.

Remote Sensing Research Unit, Department of Geography, University of California, Santa Barbara, CA 93106-4060 (scepan@geog.ucsb.edu).

Photogrammetric Engineering & Remote Sensing
Vol. 65, No. 9, September 1999, pp. 1051–1060.

0099-1112/99/6509–1051$3.00/0
© 1999 American Society for Photogrammetry and Remote Sensing

In order to determine the most effective and practical approach to sampling the DISCover data set, four sampling methods were evaluated by the VWG. A random unstratified (simple random) sampling approach was considered along with three variations of a stratified random sample. The three stratified sampling techniques were (1) random sampling stratified by continent (based on the fraction of total land area in each continent), (2) random sampling stratified by land-cover class, and (3) random sampling stratified by both continent and class.

The simple random sampling approach insures that each location in the data set has an equal chance of being sampled. The random selection technique also results in a sample set with good statistical properties (Congalton and Green, 1999). In the context of the DISCover validation, a random sample would have the advantage of allowing the samples acquired for the DISCover data set to be employed for validation of the other land-cover data products. Of particular interest are those data sets utilizing different classification systems that were produced as a part of the IGBP-DIS global land-cover mapping effort. The simple random sampling technique is limited by an inherent area bias: land-cover classes that are small in areal extent will be sampled relatively less frequently and their classification accuracy will thus be less well known. Implementing this technique with the DISCover data product also would likely have resulted in certain land-cover classes not being sampled at all.

Using a stratified random approach insures that samples are acquired from all strata, no matter their size or distribution. Original plans called for the global DISCover 1.0 data set to be produced by a number of participating laboratories processing portions of the global data set on a continental basis. In an effort to document possible variations in classification accuracy resulting from this system of distributed processing, a random sample stratified by continent was considered. This sampling technique would yield such regional accuracy measures, but the problem of class area bias found in simple random sampling would not be overcome.

By specifying a minimum sample set for each class, a random sampling procedure stratified by class produces the desired unbiased class accuracy estimates of cover with roughly equal confidence ranges. This technique was chosen over a sampling procedure stratified by both class and region. Although sampling stratified by class and region would yield statistical accuracy estimates applicable independently over all processing regions, the costs would have been prohibitive because the specified number of samples per class would have to be located and validated for each continent.

Utilizing higher spatial resolution data to validate a product derived from lower spatial resolution data on a global scale requires several assumptions. Foremost is the assumption that the validation is being done on the thematically classified product of the processing of the lower resolution data set, not the specific component data that are the basis for the product. Second, it must be assumed that the registration accuracy of the data sets is sufficient to minimize the errors resulting from verifying a sample that has been incorrectly located on the validation imagery. Finally, it must also be assumed that a predictive relationship exists between the structural and phenologic characteristics of the land-cover classes represented across the Earth's surface in the DISCover data set and the spectral signatures of these classes recorded by the sensors acquiring the verification data: Landsat Thematic Mapper (TM) and Systeme Probatoire de la Observation de la Terre (SPOT) image data.

In order to address these assumptions, the IGBP Land Cover Working Group organized a Validation Working Group (VWG), which then developed a two-tiered DISCover data set validation strategy. This strategy specified a core sampling effort to produce statistical statements of DISCover thematic accuracy and a confidence site mapping exercise to provide additional detail and depth to the validation. The rationale behind this validation protocol and its development are described in Belward (1996).

The validation protocol also specified that teams of Expert Image Interpreters (EIIs) verify the high-resolution image data selected from around the world. The procedure calls for three interpreters to independently interpret each core sample. Samples were verified as correct using a majority decision rule: at least two of the three independent interpretations had to agree in order for a sample to be verified as correct. EIIs shared regional knowledge and expertise and discussed among themselves any aspect of the validation interpretation process. EIIs were explicitly aware, however, that the final validation of each individual core sample was to be determined individually. The three interpreters worked together to perform the confidence mapping, but their specific interpretations and confidence site mapping were performed individually.

The IGBP protocol specifies a target classification accuracy for DISCover 1.0 of 0.85. The methodology also calls for the accuracy analysis to be conducted with the goal of providing a thematic classification accuracy estimate at the 95 percent confidence level (a measure of the reliability of the accuracy measurement). This goal has an impact on the sample size that must be verified for each DISCover class. The Normal approximation to the binomial distribution suggests that a sample size of 25 ($n = 25$) for each class based upon an expected accuracy of 0.85 at the 95 percent confidence level would yield an interval with a range of ± 0.143.

Examination of standard charts for binominal confidence intervals shows that $n = 25$ meets the basic LCWG requirements and also shows the impacts on accuracy confidence intervals when smaller sample sizes are used (Pearson and Hartley, 1966). A sample size of 25 for each class was also determined to be an affordable and practical sample set for the 15 classes to be validated.

Core Sampling Procedure

Sample points corresponding to 1-kilometer DISCover pixels were selected from the data set using a stratified random sample. The DISCover data set includes 17 land-cover classes (Table 1).

Fifteen of these classes were validated; samples in Snow and Ice (Class 15) and Water (Class 17) were not verified, principally due to issues of high-resolution data availability. The sampling routine was automated by producing a computer algorithm (written in C++) to segment DISCover into its constituent classes and sequentially cast 50 samples in each class (Plate 1). Universal Transverse Mercator (UTM) centroid coordinates, initially seeded by a random number generator, identified the location of each sample. This over sampling was performed in order to locate a sufficient number of samples to include at least 25 in each class for which TM or SPOT data coverage would be available.

An image search was then performed using the USGS Global Land Information System (GLIS) to identify those sample locations for which Thematic Mapper imagery was available. This search was performed for each sample in the order in which the samples appeared in the random selection process. The search included the following parameters:

- Images acquired within $\pm$ 1 year of the DISCover 1992-93 AVHRR time series data,
- Images contained < 0.40 to 0.50 cloud cover,
- Images were individually browsed to insure that they contained the DISCover core sample,
- Images contained maximum spectral signature variation, and
- Images were chosen with consideration of sample class phenology.

TABLE 1. IGBP DISCOVER DATA SET LAND-COVER CLASSIFICATION SYSTEM

Class	Class Name	Description
1	Evergreen Needleleaf Forests	Lands dominated by trees with a percent canopy cover >60% and height exceeding 2 meters. Almost all trees remain green all year. Canopy is never without green foliage.
2	Evergreen Broadleaf Forests	Lands dominated by trees with a percent canopy cover >60% and height exceeding 2 meters. Almost all trees remain green all year. Canopy is never without green foliage.
3	Deciduous Needleleaf Forests	Lands dominated by trees with a percent canopy cover >60% and height exceeding 2 meters. Consists of seasonal needleleaf tree communities with an annual cycle of leaf-on and leaf-off periods.
4	Deciduous Broadleaf Forests	Lands dominated by trees with a percent canopy cover >60% and height exceeding 2 meters. Consists of seasonal broadleaf tree communities with an annual cycle of leaf-on and leaf-off periods.
5	Mixed Forests	Lands dominated by trees with a percent canopy cover >60% and height exceeding 2 meters. Consists of tree communities wiit interspersed mixtures or mosaics of the other four forest cover types. None of the forest types exceeds 60% of landscape.
6	Closed Shrublands	Lands with woody vegetation less than 2 meters tall and with shrub canopy cover is >60%. The shrub foliage can be either evergreen or deciduous.
7	Open Shrublands	Lands with woody vegetation less than 2 meters tall and with shrub canopy cover is between 10–60%. The shrub foliage can be either evergreen or deciduous.
8	Woody Savannas	Lands with herbaceous and other understory systems, and with forest canopy cover between 30–60%. The forest cover height exceeds 2 meters.
9	Savannas	Lands with herbaceous and other understory systems, and with forest canopy cover between 10–30%. The forest cover height exceeds 2 meters.
10	Grasslands	Lands with herbaceous types of cover. Tree and shrub cover is less than 10%.
11	Permanent Wetlands	Lands with a permanent mixture of water and herbaceous or woody vegetation that cover extensive areas. The vegetation can be present in either salt, brackish, or fresh water.
12	Cropland	Lands covered with temporary crops followed by harvest and a bare soil period (e.g., single and multiple cropping systems. Note that perennial woody crops will be classified as the appropriate forest or shrub land cover type.
13	Urban and Built-up	Land covered by buildings and other man-made structures. Note that this class will not be mapped from the AVHRR imagery but will be developed from the populated places layer that is part of the Digital Chart of the World.
14	Cropland/Natural Vegetation Mosaics	Lands with a mosaic of croplands, forest, shrublands, and grasslands in which no one component comprises more than 60% of the landscape.
15	Snow and Ice	Lands under snow and/or ice cover throughout the year.
16	Barren	Lands exposed soil, sand, rocks, or snow and never has more than 10% vegetated cover during any time of the year.
17	Water Bodies	Oceans, seas, lakes, reservoirs, and rivers. Can be either fresh or salt water bodies

Where TM imagery was not available for the minimum sample set in a class, SPOTdata archives were queried using the same sample search parameters. From the two archives, imagery was available for at least 25 samples in 13 of the 15 validated classes. Deciduous Needleleaf Forests (Class 3) and Permanent Wetlands (Class 11) were underrepresented. Imagery was available for only 11 samples in Class 3, primarily due to the distribution of this class within the data set. Deciduous Needleleaf Forests are primarily found in central Siberian Russia. This region is not covered by TM or SPOT ground receiving stations, and there is seldom a priority to collect imagery of these areas via data recorders. Image data covering the core samples located for this class could not be found. Imagery was obtained for 17 samples in Class 11. Permanent Wetlands are typically small in areal extent in relation to the 1-km DISCover minimum mapping unit. In addition, most of the Permanent Wetlands mapped in DISCover 1.0 are found at high latitudes, and the same problem of image availability found in the Deciduous Needleleaf class occurs in this class.

Once identified, the high-resolution image data were then acquired for use in the validation. The distribution of core samples that were identified (and for which high-resolution imagery was obtained) is shown in Plate 2. Landsat TM and SPOT image data were employed as validation interpretation source data. Two-hundred eighty-nine Landsat TM scenes were contributed to this study by the USGS EROS Data Center. The European Commission Joint Research Center, the Centre Nationale Etudes Spatial, and the IGBP Secretariat contributed 143 SPOTscenes for use at the Global Validation Workshop (GVW).

Extensive image processing was required in order to compile, format, organize, and prepare the high-resolution validation data for analysis by the EII team. The process of reprojecting the validation imagery and registering it with DISCover 1.0 was a critical, and particularly complex, task. This task was made more efficient by developing a computer program to extract and re-project imagery from the original CD-ROM source media and to generate manual interpretation products. Transformation of image corner coordinates to the Interrupted Goode's Homolosine projection was accomplished by compiling portions of the USGS/EDC General Coordinate Transformation Program (GCTP) into the program. The reprojection task was complicated by the fact that none of the available image processing or GISsoftware packages supported the Goode's projection. A complete description of this procedure may be found in Husak *et al.* (1999, in this issue).

Once reprojection and registration was completed for each scene, a validation subscene was extracted for each core sample/confidence site. Each core sample was included as a Primary Sample Point centered in a 40- by 40-kilometer image subscene. The 40- by 40-km size was selected in order to provide the interpreter withsufficient regional land-cover context for each sample.

An important subset of the core sample development was the processing of data for the Confidence Site effort. Data and processing requirements for the Confidence Site mapping were the primary factors in many of the decisions that were made regarding acquisition and image processing of the high-resolution core validation data. Subscene size and geometry as well as selection and orientation of secondary sample points were determined based on parameters specified for Confidence Site mapping.

In order for the validation to be applicable to an updated

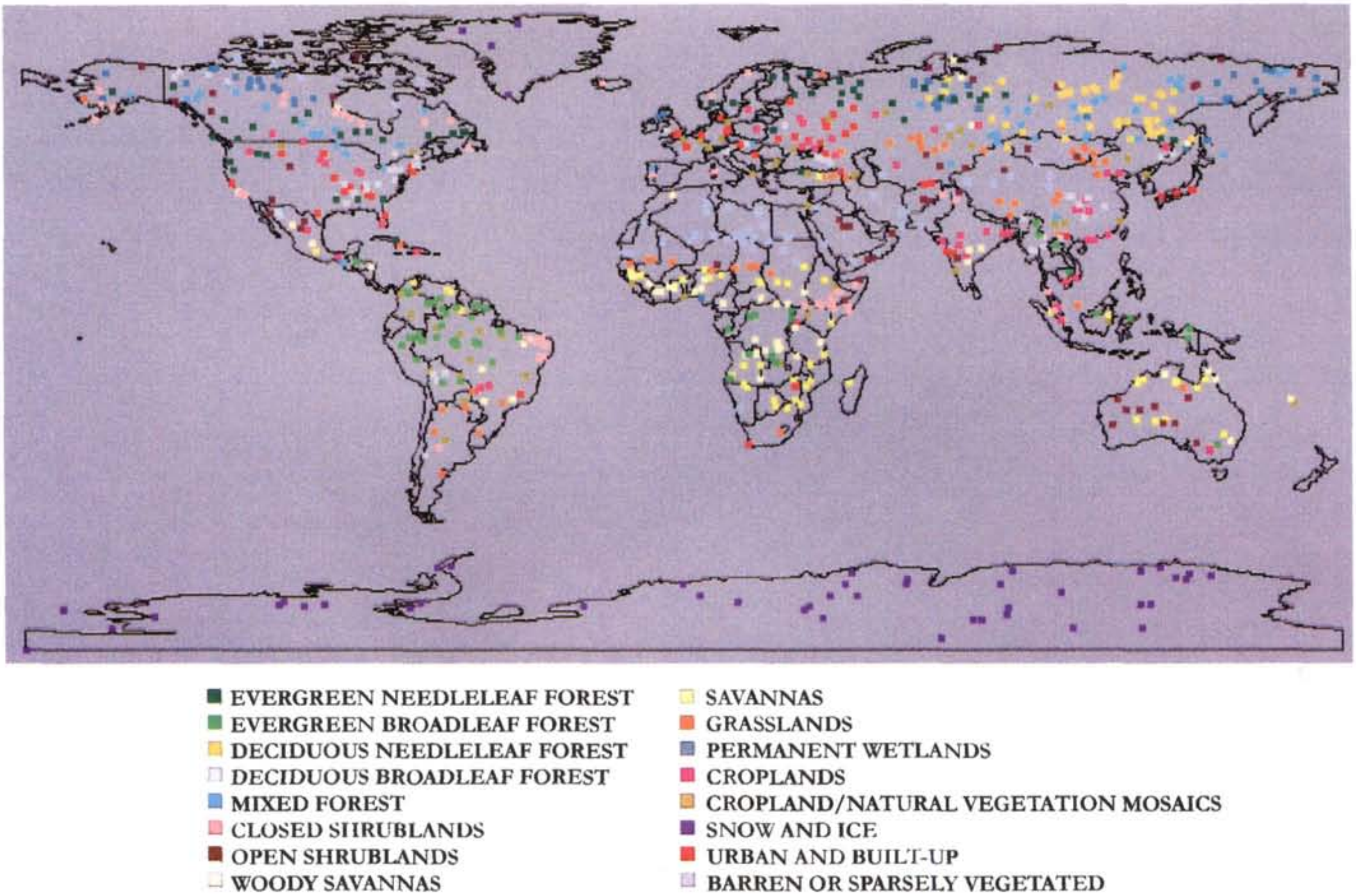

Plate 1. IGBP DISCover by class validation sample distribution.

global land-cover product, as well as to improve the core validation procedure, the idea of Confidence Site analysis was adopted by the IGBP VWG (Belward, 1996). The concept of Confidence Sites became the second important component of the overall validation methodology. As envisioned by the VWG, the Confidence Sites would comprise a set of locations at which a variety of fine-resolution land-cover descriptors and data sets have been acquired for a broad region and will be made readily available for future research.

The VWG felt that work conducted at Confidence Sites could serve specifically to improve our overall understanding of data accuracy and validation issues for a number of global land-cover research applications. These applications include

- Improved land-cover classification schemes,
- Improve the methods used in developing global land-cover databases,
- As a test bed for improved core sampling procedures,
- To test advanced thematic accuracy assessment procedures, and
- To assess improved locational accuracy assessments.

The process of Confidence Site development and mapping is described in Muchoney*et al.* (1999, in this issue).

Another important step in developing the validation procedure was the segmentation of the global land surface into 13 separate validation regions based on the IGBPGlobal Change System for Analysis, Research and Training (START) framework: i.e.,

- Region 1—North America: Canada and Alaska
- Region 2—North America: U.S. and N. Mexico
- Region 3—Central America and Northern South America
- Region 4—Amazonia and Brazil
- Region 5—South America and the Andes
- Region 6— Western Europe
- Region 7—Saharan and Subsarahan Africa
- Region 8—Central Africa
- Region 9—Southern Africa and Madagascar
- Region 10—Russia and Northern Scandinavia
- Region 11—China, India, Japan, and Central Asia
- Region 12—Southeast Asia
- Region 13—Australia and New Zealand

To address the complex technical and logistical problems associated with this global scale endeavor, the IGBP LCWG had previously developed a draft collaborative framework structured around three tiers of international cooperation. Based upon a regional approach, an informal group of DISCover Validation Regional Advisors was designated who could assist in the identification and recruitment of Cooperating Laboratories and EIIs. Cooperating Laboratories served as regional centers to support the collection of image data and ancillary materials for each region. EIIs were the regional experts who actually performed the interpretation of high-resolution imagery for DISCover 1.0 validation. A complete listing of the individuals who participated in this study as EIIs is included in Scepan *et al.* (1999, in this issue).

Validation samples were also aggregated into the 13 regions, primarily for convenience because the regional approach made compilation of materials and scheduling of validation activities easier.

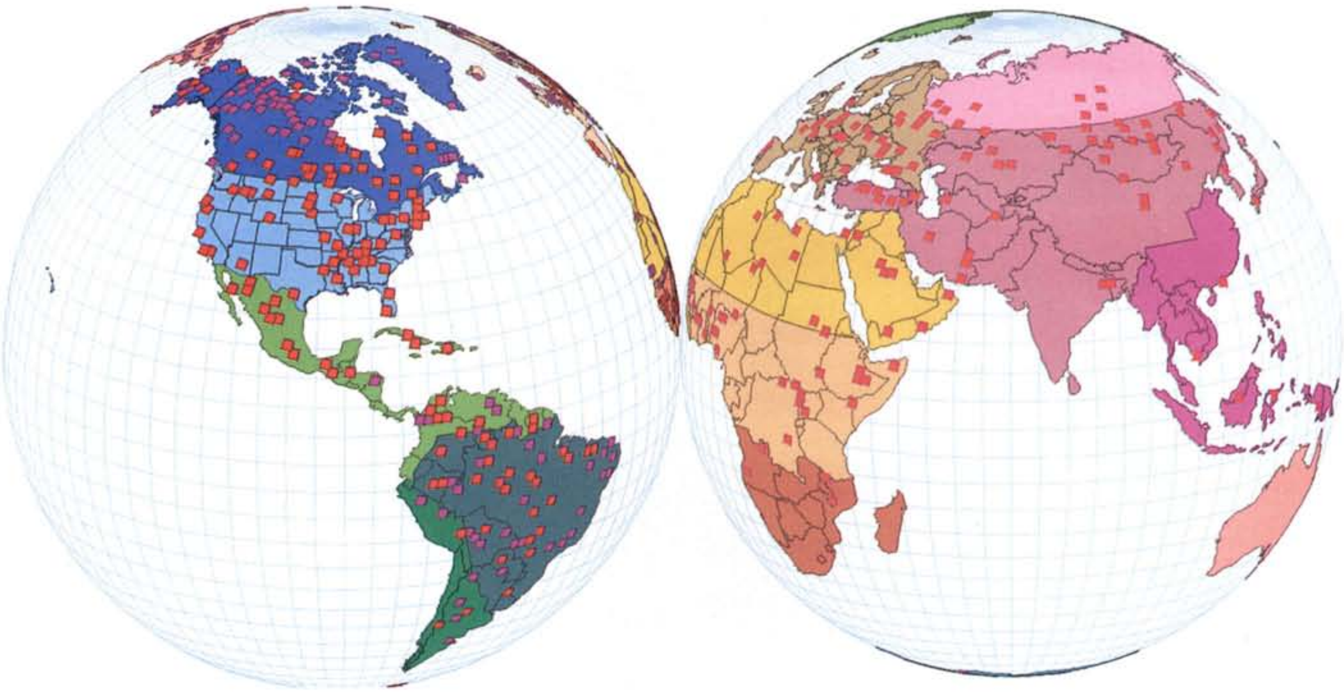

Plate 2 IGBP sample distribution by interpretation region.

A series of image interpretation keys were developed to provide EIIs with representative examples of land-cover classes in a variety of spatial and temporal regimes. Researchers at the Remote Sensing Research Unit (RSRU), University of California, Santa Barbara (UCSB) developed keys in an effort to establish a common reference framework for the core sample validation process (Kelly *et al.*, 1999, in this issue). The development of these keys was based upon the following assumptions:

- A correlation exists between image properties and land-cover classes;
- AVHRR and validation imagery can be consistently identified, registered, and cataloged, and
- Current state-of-the-practice interpretation techniques are adequate to statistically validate land-cover classes and the methodology.

Owing to logistics and financial considerations, interpretation keys were not completed for all classes in all regions for all major class phenologic states. Keys that were developed were provided to the EII during the GVW in the form of hardcopy graphics organized by IGBP interpretation region. EIIs employed these image interpretation keys as a part of the suite of ancillary data that were available to the VWG (Kelly *et al.*, 1999, in this issue).

Methodology Test

The validation methods and specifications originally developed by the VWG were evaluated and refined during an early Validation Test Workshop. This meeting was held on 01-03 February 1998 in Santa Barbara, California, and led to the final definition of the validation tools and protocol. IGBP Validation Regions 1, 2, and 3 were used for this test (Canada, the U.S., Mexico, Central America, and a portion of northern South America). The Validation Test Workshop included participants from the UCSB; Joint Research Center Isrpa, Italy; Department of Geography, Boston University; Desert Research Institute, University of Nevada; USGS/EROS Data Center, Sioux Falls, South Dakota; MeteoFrance, Toulouse, France; and the University of Maryland, College Park, Maryland.

To test the core sampling procedure, 37 DISCover Data Set samples were validated by interpretation of Landsat TM imagery. In accord with the IGBP validation procedure, each sample was interpreted by three separate interpreters, and statistics were summed for each point.

To evaluate confidence site mapping procedures, 15 TM subscenes, each approximately 40 by 40 kilometers in size and centered on a DISCover sample, were manually mapped and classified using the DISCover legend. Mapped polygons were also coded with several specific vegetation cover parameters of interest to confidence site investigators. Mapping times and comments were recorded during completion of each subscene.

Following these activities, discussions were held among participants to address the procedures and technical issues relevant to the validation. Four topics were of principal interest:

- Global Validation Workshop schedule (including data acquisiton),
- DISCover/high-resolution (TM-SPOT) registration issues,
- Validation interpretation issues, and
- Issues related to Confidence Site mapping.

A schedule was also developed for the DISCover Global Validation Workshop (GVW). A listing of potential GVW invitees was compiled. Following the validation test, a procedure to identify any systematic bias in the co-registration of the Landsat TMverification imagery and the DISCover data set was discussed. A suite of suggested ancillary data to be provided for the GVW was specified, including Digital Elevation Models (DEM), the Digital Chart of the World (DCW), and a variety of vegetation and other thematic maps covering each validation region.

Global Validation Workshop

The IGBP Global Validation Workshop (GVW) was held during 07–18 September 1998 at the USGS EROS Data Center in Sioux Falls, South Dakota. GVW Session 1 took place during 07–11 September and included Validation Regions 3 (Central America), 4 (Northern South America), 7 (North Africa), 8 (Central Africa), 9 (Southern Africa), 11 (Central Asia/Japan), and 12 (Southeast Asia/China). Group 2 was composed of Validation Regions 1 (North America/Canada), 2 (North America/US), 5 (Southern South America), 6 (Europe), 10 (Russia), and 13 (Australia/New Zealand). This group met during 14–18 September.

The core sampling procedure did not include the Water and Snow and Ice classes due to data constraints and DISCover data set characteristics. The required minimum 25 samples were validated for 13 of the 15 classes. The Deciduous Needleleaf Forests and Permanent Wetlands classes (both relatively small in spatial extent within DISCover 1.0) included fewer than 25 samples. The principal reason for this lack of data is the location of these classes relative to image availability. Confidence Site mapping was accomplished for approximately 73 percent (302 of 415) of the confidence sites.

High-resolution image interpretation was accomplished using digital subscenes presented to the interpreter on video display. Each TM subscene was provided to the EIIs as seven-band data sets.

All verification activities were performed using commercially available hardware and image processing and geographic information system software packages. The USGS/EDC facilities include IBM compatible personal computers supporting the Microsoft Windows-NT operating system. Raster/vector image processing activities for the GVW were performed using ERDAS/Imagine (Version 8.3.1) image processing software.

Within each region, all PSP/SSP sets were displayed and verified by each EII. A standardized hardcopy form was used by EIIs to record each verified sample. Information recorded for each sample included EII name, date, interpreted class, and interpretation confidence for each sample. The original VWG protocol specified that at least two of the three EIIs must agree on the land-cover type before a given AVHRR sample was accepted as correct. If the three EIIs all disagreed, and all identified three different land-cover types for a sample, the DISCover classification was considered to be in error.

The results of the GWV core sample analysis are presented as a series of tables, contingency matrices, and summary discussions. Individual class user accuracies (including confidence intervals) are summarized along with the overall DISCover 1.0 data set accuracy. Contingency matrices are also provided in order to identify confusion classes. These tables have been developed in two ways: by a summary which treats each sample interpretation independently and a summary which uses a majority rule for samples determined to be in error as well as correctly classed samples.

Results

Table 2 shows that the highest individual class accuracies were established in Class 2 (Evergreen Broadleaf Forests; 0.840), Class 7 (Open Shrublands; 0.778), and Class 16 (Barren; 1.00). Classes 2 and 16 meet the accuracy goal established by IGBP for DISCover 1.0 of 0.85 accuracy (at 95 percent confidence).

The accuracies for DISCover Class 4 (Deciduous Broadleaf Forests) and Class 9 (Savannas) are the lowest of the 15 classes verified. Class 3 (Deciduous Needleleaf Forests) and Class 11 (Permanent Wetlands) also have low accuracy, but the number of samples validated for these classes was well below the minimum 25 samples specified in the validation protocol.

Measuring overall data set accuracy based upon the spatial extent of each class is the appropriate estimate to use for stratified sampling. Such an area-weighted accuracy measurement combines results over strata to construct an overall accuracy estimate (Cochran, 1977). Using this technique, the overall accuracy of DISCover based on the original VWG protocol is 0.669 (Table 2). A bivariate analysis shows a positive correlation between class accuracy and class coverage fraction ($r = 0.733$), indicating that those DISCover classes that cover larger porportions of area are classified with higher accuracy (Figure 1).

Thematic map accuracy is also commonly summarized through a contingency table or "confusion matrix" approach (Story and Congalton, 1986) where individual samples are plotted as they were mapped on one of the *x-y* axes and as they were verified on the other. Producing a confusion matrix for DISCover 1.0 using samples based on the original protocol was complicated by the use of a majority decision rule for verifying core samples as correct. The validation of each core sample was comprised of three individual interpretations, and DISCover classification errors were sometimes not interpreted by each interpreter as the same class. Thus, for core samples verified as incorrect, there is not always a single appropriate entry that may be made in an error matrix.

Additional contingency tables were compiled using two alternate methods. The first counted each EII interpretation

TABLE 2. IGBP DISCOVER 1.0 OVERALL ACCURACY

DISCover Class	Verified Samples	Samples Verified Correct	User's Accuracy	Confidence Interval (.95)	Percent Cover	Overall DISCover Accuracy
1-Evergreen Needleleaf Forests	26	15	.58	0.38–0.77	.0482	.669
2-Evergreen Broadleaf Forests	25	21	.84	0.69–0.99	.0916	
3-Deciduous Needleleaf Forests	11	5	.46	0.15–0.76	.0123	
4-Deciduous Broadleaf Forests	25	10	.40	0.20–0.60	.0284	
5-Mixed Forests	27	15	.56	0.36–0.75	.0471	
6-Closed Shrublands	27	15	.56	0.36–0.75	.0198	
7-Open Shrublands	27	21	.78	0.62–0.94	.1489	
8-Woody Savannas	31	18	.58	0.40–0.76	.0750	
9-Savannas	26	11	.42	0.23–0.62	.0755	
10-Grasslands	26	15	.58	0.39–0.77	.0830	
11-Permanent Wetlands	17	5	.29	0.07–0.52	.0075	
12-Cropland	28	18	.64	0.46–0.82	.1028	
13-Urban and Built-up	30	16	.53	0.35–0.72	.0032	
14-Cropland/Natural Vegetation Mosaics	26	13	.50	0.30–0.70	.1114	
16-Barren	27	27	1.00	0.87–1.00	.1452	

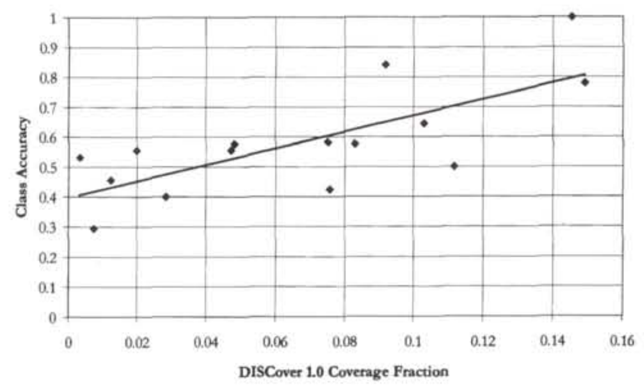

Figure 1. Bivariate plot of DISCover class accuracy and cover fraction.

separately; there are typically three interpretations for each core sample, although, due to various factors, some core samples were not verified by all three EIIs. Individual verifications were then charted and the matrix was produced (Table 3). An examination of Table 3 shows that there are few DISCover 1.0 classes which were systematically confused with other classes. In DISCover 1.0, Class 4 (Deciduous Broadleaf Forests) is most often confused with Classes 12 (Cropland) and 14 (Cropland/Natural Vegetation Mosaics). Class 8 (Woody Savannas) is most often confused with Class 9 (Savannas). Class 11 (Permanent Wetlands) is confused with Class 7 (Open Shrublands). Errors in classification within other DISCover classes are clearly non-systematic. The greatest disparity is found in EII interpretations for Class 1 (Evergreen Needleleaf Forests), with interpretations shown in 13 other classes.

In order to remove an element of interpretation inconsistency in the accuracy analysis, a second method of producing contingency tables was also employed. This method applied the majority rule to the core samples that were verified in error as well as those verified as correct. If a DISCover sample was to be determined in error, at least two of the three EIIs had to verify the sample as the same class. If two of the three EIIs could not agree on the sample class (i.e., all three interpreters identified a different land-cover type), the sample was not included in the accuracy assessment. This method has the advantage of summarizing verified core samples (rather than EII interpretations) in a confusion matrix and applies the majority rule to samples (those incorrectly mapped) which do not lie on the confusion matrix diagonal as well as samples that do (correctly mapped samples). While this approach reduces the number of samples that are included in the matrix, it increases the consistency of the sample set by directly coupling each sample to a land-cover class assigned by a majority of interpreters. If no majority land cover emerged for a sample during the verification, the sample was not included in this analysis. The results of this analysis are shown in Table 4. Calculated using this method, the overall accuracy of DISCover 1.0 is 73.5 percent. Individual class accuracies range between 38.5 percent and 100 percent.

Finally, it is useful to assess the thematic accuracy of DISCover 1.0 within the context of the intended use of the data set. DeFries and Los (1999, in this issue) examined the accuracy of DISCover 1.0 within the specific context of global climate modeling. In this study, the authors examined the land-cover parameters required in the Simple Biosphere Model (SiB2) land-atmosphere interactions model (Sellers *et al.*, 1996) and related these to the characteristics of the thematic classes that make up the DISCcover data set. Errors of commission were then examined for each DISCover class to determine whether these errors impact the specific parameters required for SiB2. The results of this study show that the accuracy of the DISCover data set may be as high as 90 percent for SiB2 model parameters.

In addition to the analysis of the accuracy of DISCover 1.0, the interpretability of the high-resolution image data used to validate the core samples was evaluated from the information gathered during the conduct of this exercise, as well as the reactions and comments of the EIIs themselves.

As a part of the GVW interpretation process, each EII placed a confidence value on the interpretation of each core sample: 1 (Low Confidence), 2 (Medium Confidence), and 3 (High Confidence). These confidence metrics were then compiled by class in order to identify the IGBP classes which the EIIs believed were relatively more or relatively less interpretable. The overall interpretation confidence level for the DISCover data set is medium to high. Only Class 10 (Grasslands) and Class 11 (Per

TABLE 3 DISCOVER 1.0 VALIDATION ERROR MATRIX (SUMMED FOR ALL EII INTERPRETATIONS BY PROPORTIONS)

			Verified Class (Proportion)																	Total Samples In Class
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
Mapped Class	Evergreen Needleleaf Forests	1	**.557**	.013		.025	.089	.013	.013	.038	.025	.038	.025	.038	.013	.076		.038		79
	Evergreen Broadleaf Forests	2		**.775**						.013		.038	.038			.088		.025	.013	80
	Deciduous Needleleaf Forests	3			**.485**		.061						.152	.212		.030		.030	.030	33
	Deciduous Broadleaf Forests	4	.013	.025		**.443**	.089	.013	.025	.025	.063	.013	.013	.127		.114		.013	.025	79
	Mixed Forests	5	.013			.050	**.538**	.038	.088	.038	.013	.013	.013	.038		.050		.025		80
	Closed Shrublands	6	.013				.013	**.488**	.113	.025	.063	.088	.038	.013		.038		.050	.062	80
	Open Shrublands	7						.024	**.735**	.036	.024	.048	.012	.012		.012		.096		83
	Woody Savannas	8		.043			.022	.043	.054	**.505**	.183	.043		.086		.022				93
	Savannas	9		.012		.012	.012	.037	.062	.160	**.444**	.037		.049	.012	.111		.037	.012	81
	Grasslands	10				.018	.038	.018	.088		.075	**.513**	.025	.113		.025		.088	.018	80
	Permanent Wetlands	11	.053			.018	.018	.105	.298			.105	**.316**					.070	.018	57
	Cropland	12		.010	.012	.012	.012	.012		.023	.058	.081	.023	**.605**	.035	.116				86
	Urban and Built-up	13	.013			.038	.064					.038		.090	**.584**	.115		.026	.038	78
	Cropland/Natural Vegetation Mosaics	14	.077	.128			.038	.026	.013	.026	.038	.038		.103	.051	**.449**		.013		78
	Snow and Ice	15																		
	Barren	16								.012		.012					.012	.952	.012	84
	Water Bodies	17																		

TABLE 4 DISCOVER 1.0 VALIDATION ERROR MATRIX (MAJORITY RULE ERROR ANALYSIS BY PERCENT)

			Varified Class (Proportion)																	Total Samples In Class with Majority Agreement
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
	Evergreen Needleleaf Forests	1	**.750**				.050						.050	.050		.050		.050		20
	Evergreen Broadleaf Forests	2		**.875**									.042			.042		.042		24
	Deciduous Needleleaf Forests	3			**.556**								.222	.222						9
	Deciduous Broadleaf Forests	4				**.625**	.063							.188		.063			.063	16
	Mixed Forests	5	.045			.045	**.681**		.136					.045		.045				22
	Closed Shrublands	6						**.750**				.050	.050					.050	.100	20
	Open Shrublands	7							**.875**			.042						.083		24
Mapped	Woody Savannas	8								**.615**	.154	.038		.115						26
Class	Savannas	9								.176	**.647**			.059		.118				17
	Grasslands	10							.050			**.750**		.100		.050		.050		20
	Permanent Wetlands	11	.077						.385			.077	**.385**					.077		13
	Cropland	12										.048		**.857**	.048	.048				21
	Urban and Built-up	13					.083							.083	**.667**	.083		.042	.042	24
	Cropland/Natural Vegetation Mosaics	14	.087	.130			.043							.130	.043	**.565**				23
	Snow and Ice	15																		
	Barren	16																**1.00**		27
	Water Bodies	17																		

Total Number of Samples = 379
Number of Samples with Interpreter Majority Agreement = 306
Number of Samples Correct by EII Majority Agreement = 225
Total Number of Samples Incorrect = 154
Number of Samples Incorrect with EII Majority Class Agreement = 81
Percent of Samples Incorrect Samples with EII Majority Agreement = .526 (81/154)
Percent of Majority Agreement Samples Correct = .735 (225/306)

manent Wetlands) have average confidence values indicating relatively low EII confidence.

To identify any regional differences in EII interpretation confidence, ratings were also compiled by IGBP Validation Region. Average EII confidence values are below 2.0 (medium) confidence for only Regions 1 (North America/Canada) and 11 (Central Asia/Japan).

Following the analysis of EII interpretation confidence by region and DISCover class, the data were correlated in order to determine whether a relationship existed between the verified individual DISCover class accuracies and EII confidence. The results of this analysis show that DISCover classes which were verified to have the highest accuracies also tended to be interpreted with higher confidence by EIIs (Scepan *et al.*, 1999, in this issue).

This exercise demonstrates that Landsat TM and SPOT imagery can be efficiently used to validate global land-cover products such as DISCover. The utility of, and confidence that may be placed in, this technique depends principally upon the land-cover classification scheme in use or a subset of categories. The 15 validated IGBPland-cover classes were not equally interpretable on the TM and SPOT imagery. Interpreter confidence was highest for Evergreen Broadleaf Forests and Urban and Built-up DISCover classes while Grasslands and Permanent Wetlands were interpreted with relatively less confidence. Analysis of image interpretation in each of the 13 validation regions indicates that confidence in interpretations for North America and Canada (Region 1) and Central Asia/Japan (Region 11) are lower than average. Confidence in interpretations is significantly higher than average for North America and the United States (Region 2), Northern and Southern South America (Regions 4 and 5), and Southeast Asia and China (Region 12). A correlation of 0.733 is observed between the accuracy of individual DISCover classes and the confidence of the Expert Image Interpreter in class interpretation. Variations in interpretation confidence are also noted between regions or based upon the geographic location of samples.

Summary and Conclusions

There is still a great deal not known about validating global-scale thematic geospatial data sets. The methodology and the procedures that were employed in this effort were not ideal in approach or implementation, but a great deal was learned in the course of this research. Many significant issues were addressed in the conduct of this work; many of these issues remain open for further study. What is clear, however, is that, to the extent practical, the goals of this study were accomplished. A statistical validation was performed of the IGBP DISCover (Version 1.0) 1-kilometer global land-cover data set. While all DISCover classes were not adequately validated to the standards that were set, the classes that were validated cover 86.4 percent of the global land surface. The validation did not include snow and ice (11.4 percent of land surface), and there were inadequate samples to validate permanent wetlands (0.9 percent of the Earth's land surface) and deciduous needleleaf forests (1.3 percent). Water was the fourth class not validated.

The required minimum 25 samples were validated for 13 of the 15 DISCover classes. Deciduous Needleleaf Forests and Permanent Wetlands classes (both relatively small in spatial extent within DISCover 1.0) included fewer than 25 samples, principally due lack of high-resolution verification data. If we are to continue to use stratified random sampling procedures for the validation of global-scale land-cover products, methods must be found to improve access to imagery of sample locations across the globe.

For the 15 DISCover classes validated, the average class accuracy was 59.4 percent, with accuracies for the 15 verified DISCover classes ranging between 40.0 percent and 100 percent. The overall area-weighted accuracy of the data set was determined to be 66.9 percent. When only samples which had a majority interpretation for errors as well as for correct samples are considered, the average class accuracy of the data set is 73.5 percent.

The highest individual class accuracies were established in Class 16 (Barren; 1.00), Class 2 (Evergreen Broadleaf Forests; 0.840), and Class 7 (Open Shrublands; 0.778). Classes 2 and 16 meet the accuracy goal established by IGBP for DISCover 1.0 of 0.85 accuracy (at 95 percent confidence). The accuracies for DISCover Class 4 (Deciduous Broadleaf Forests) and Class 9 (Savannas) are the lowest of the 15 classes verified. Class 3 (Deciduous Needleleaf Forests) and Class 11 (Permanent Wetlands) also have low accuracy, but the number of samples validated for this class was well below the minimum 25 samples specified in the validation protocol.

A bivariate analysis of class accuracy and class coverage fraction shows a positive correlation between these variables ($r = 0.733$); larger and less fragmented classes in DISCover 1.0 have higher thematic accuracy. In addition, work by DeFries and Los (1999, in this issue) indicates that, for certain climate model parameters, the user accuracy of the DISCover data set may be as high as 90 percent.

Based on this study, it can no longer be assumed that all DISCover classes are equally interpretable on satellite spectral imagery. This reinforces the desirability of having both discipline scientists and remote sensing specialists involved in the development of classification schemes for global geospatial products. The results also clearly demonstrate the difficulty associated with the interpretation of many of these classes from remotely sensed data sets. In these instances, the question remains regarding the strength of the spectral relationship between the AVHRR-based DISCover classes and a number of the classes as seen on the TM and SPOT imagery. Impacts of registration, interpreter consistency, and the availability and quality of ancillary data may also be questioned based upon the results described here. This study demonstrates that there is still much to learn.

The IGBP-DIS Global Validation program represents an important first step towards the development of procedures to operationally validate global-scale thematic land-cover products at regular intervals. This research has demonstrated that such validation is possible, but depends upon the good will, support, cooperation, and collaboration of interested organizations and institutions. This effort can serve as a foundation for future systematic global land-cover validation efforts. Currently, there are no plans in place to extend this validation effort. Altough efforts are underway to garner support for future validation activities, it remains easier to support production of these data sets than to support their validation.

Acknowledgments

The effort to validate the DISCover 1.0 land-cover data set was international in scope. It would not have been possible without the dedicated participation and support of a large number of individuals, groups, and organizations. Principal support for the study was provided by two research grants from the National Aeronautics and Space Administration Earth Science Enterprise: Land-Cover and Land-Use Change Program Research Grant NAG5-6259 and Research Grant NAG5-3620. Dr. Thomas Loveland and a number of others made it possible for the U.S. Geological Survey, EROS Data Center, Sioux Falls, South Dakota to play a critical role in this research by generously providing data, technical expertise, and logistic support. Dr. Gerard Szejwach and Ms. Martin Michou coordinated important support for this work from the IGBP Data and Information System Office in Tolouse, France. Drs. Alan Belward and J.P. Malingreau led the significant contributions to this work made by the European Commission Joint Research Center, Ispra, Italy.

Validating the DISCover Data Set would have not been possible without the energy and leadership provided to the effort by Dr. John Estes of the University of California Santa Barbara. A number of other university researchers and faculty also provided direction and technical expertise to the project. Particular appreciation is expressed to Dr. Alan Strahler and Douglas Muchone at Boston University; Dr. Kenneth McGwire at the Desert Research Institute, University of Nevada, Reno; Drs. John Townshend and Ruth Defries at the University of Maryland; and Dr. Chris Justice of the Department of Environmental Sciences, University of Virginia.

The contribution made by the Earth Resouces Display and Analysis System, Inc., of Atlanta, Georgia, in allowing the use of their IMAGINE image processing software during the global validation workshop, is gratefully acknowledged.

Deep appreciation is expressed to the individuals who served as Expert Image Interpreters, as well as to the group of graduate student researchers and other staff of the Remote Sensing Research Unit, Department of Geography, University of California Santa Barbara. Without their expertise and enthusiastic participation, the Global Validation Workshop would have been impossible.

References

Belward, A., (editor), 1996. *The IGBP-DIS Global 1Km Land Cover Data Set "DISCover": Proposal and Implementation Plans*, IGBP-DIS Working Paper 13, International Geosphere Biosphere Programme, European Commission Joint Research Center, Ispra, Varese, Italy, 66 p.

Borella, H.M., J.E. Estes, C.E. Ezra, J. Scepan, and L. Tinney, 1982. *Image Analysis for Facility Siting: A Comparison of Low- and High-Altitude Image Interpretability for Landuse/Landcover Mapping*, United States Nuclear Regulatory Commission Publication, NUREG/CR-2861, S-744-R, 61 p.

Cochran, W.G., 1977. *Sampling Techniques*, Wiley Series in Probability and Mathematical Statistics, New York, N.Y., 84 p.

Congalton, R.G., and Kass Green, 1999. *Assessing the Accuracy of Remotely Sensed Data: Principals and Practices*, Lewis Publishers, New York, N.Y., 137 p.

DeFries, R., and J.R.G. Townshend, 1994. NDVI-derived land cover classifications at a global scale, *International Journal of Remote Sensing*, 15(7):3443–3462.

DeFries, R.S., and S.O. Los, 1999. Implications of land cover misclassification parameter estimates in global land surface models: An example from the Simple Biosphere Model (SiB2), *Photogrammetric Engineering & Remote Sensing*, 65(9):1083–1088.

Eidenshink, J.C., and J.L. Faundeen, 1994. The 1km AVHRR Global Land Cover Data Set: First Stages in Implementation, *International Journal of Remote Sensing*, 15(7):3567–3586.

Estes, J.E., J. Scepan, and L. Ritter, 1987. *Evaluation of Low Altitude Remote Sensing Techniques for Obtaining Environmental Information*, U.S. Nuclear Regulatory Commission Publication NUREG/CR-3583, S-762-R, Washington, D.C., 67 p.

Estes, J.E., and W. Mooneyhan, 1994. Of Maps and Myths, *Photogrammetric Engineering & Remote Sensing*, 60(5):517–524.

Fitzpatrick-Lins, Katherine, 1980. *The Accuracy of Selected Land Use and Land Cover Maps at Scales of 1:250,000 and 1:100,000*, Geological Survey Circular 829, U.S. Geological Survey, Washington, D.C., 24 p.

Husak, G.J., B.C. Hadley, and K.C. McGwire, 1999. Registration Accuracy of High-Resolution Satellite Data Used in the IGBP Validation, *Photogrammetric Engineering & Remote Sensing*, 65(9):1033–1039.

Kelly, M., J.E. Estes, and K.A. Knight, 1999. Image Interpretation Keys for Validation of Global Land-Cover Data Sets, *Photogrammetric Engineering & Remote Sensing*, 65(9):1041–1049.

Loveland, T.R., and A.S. Belward, 1997. The IGBP-DIS Global 1 km Land Cover Data Set, DISCover First Results, *International Journal of Remote Sensing*, 18(15):3289–3295.

Loveland, T.R., Z. Zhu, D.O. Ohlen, J.F. Brown, B.C. Reed, and L. Yang, 1999. An Analysis of the IGBP Global Land-Cover Characterization Process, *Photogrammetric Engineering & Remote Sensing*, 65(9):1021–1032.

Lunetta, Ross S., Russell G. Congalton, Lynn K. Fenstermaker, John R. Jensen, Kenneth C. McGwire, and Larry R. Tinney, 1991. Remote Sensing and Geographic Information System Data Integration: Error Sources and Research Issues, *Photogrammetric Engineering & Remote Sensing*, 47(6):677–687.

Muchoney, D., A. Strahler, J. Hodges, and J. Locastro, 1999. The IGBP DISCover Confidence Sites and the System for Terrestrial Ecosystem Parameterization: Tools for Describing and Validating Global Land-Cover Data, *Photogrammetric Engineering & Remote Sensing*, 65(9):1061–1067.

Pearson, E.S., and H.O. Hartley, (editors), 1966. *Biometrika Tables for Statisticians, Volume 1, Third Edition*, Cambridge University Press, Cambridge, England.

Rosenfield, George H., K. Fitzpatrick-Lins, and H.S. Ling, 1981. Sampling for Thematic Map Accuracy Testing, *Photogrammetric Engineering & Remote Sensing*, 48(1):131–137.

Scepan, J., G. Menz, and M.C. Hansen, 1999. The DISCover 1.0 Image Interpretation Process, *Photogrammetric Engineering & Remote Sensing*, 65(9):1075–1081.

Sellers, P.J., S.O. Los, C.J. Tucker, C.O. Justice, D.A. Dazlich, G.J. Collatz, and D.A. Randall, 1996. A Revised Land Surface Parameterization (SiB2) for Atmospheric GCMs-Part II: The Generation of Global Fields of Terrestrial Biophysical Parameters from Satellite Data, *Journal of Climate*, 9:706–737.

Story, M., and R.G. Congalton, 1986. Accuracy Assessment: A User's Perspective, *Photogrammetric Engineering & Remote Sensing*, 53(3):397–399.

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*, 35:243–256.

Townshend, J.R.G., C.O. Justice, D. Skole, J.-P. Malingreau, J. Cihlar, P. Teillet, F. Sandowski, and S. Ruttenberg, 1994. The 1Km Resolution Global Data Set: Needs of the International Geosphere Biosphere Programme, *International Journal of Remote Sensing*, 15(17):3417–3442.

Tucker, C.J., J.R.G. Townshend, and T.E. Goff, 1985. African Land Cover Classification Using Satellite Data, *Science*, 227:369–375.

benefits of membership?

asprs
THE IMAGING & GEOSPATIAL INFORMATION SOCIETY

The benefits of membership in ASPRS: The Imaging and Geospatial Information Society far exceed the initial investment.

Member benefits and services include:

- Monthly subscription to *Photogrammetric Engineering & Remote Sensing (PE&RS)*
- Discounts of 25-40% on all ASPRS publications
- JOB FAIR access
- Discounts on registration fees for ASPRS Annual Meetings and Specialty Conferences
- Discounts on ASPRS Workshops
- Receipt of your Region's Newsletters
- Region specialty conferences, workshops, technical tours and social events
- Local, regional, national, and international networking opportunities
- Eligibility for over $18,000 in National and Regional awards, scholarships and fellowships
- Opportunity to obtain professional certification
- And many more

Plus, ASPRS offers two levels of membership. See page 1105 or call 301-493-0290x109 for more information and an application.

www.asprs.org