# 1. TITLE

### 1.1 Data Set Identification

ISLSCP II IGBP DISCover and SiB Land Cover, 1992-1993

### 1.2 File Name(s)

The data sets are provided as a contribution to the International Satellite Land Surface Climatology (ISLSCP) Initiative II data collection at three spatial resolutions of 0.25, 0.5 and 1degrees lat /long and for two different classification schemes. For each spatial resolution there is adominant land cover type classification layer, ranging from 0 to 15 for the SiB (Simple Biosphere) classification scheme, and from 0 to 17 for the IGBP classification scheme (see Section 8.2 for a list of cover types). For each classification scheme used, there are layers that provide the fraction, from 0 to 100, of each land cover type per cell. The dominant land cover type files are named:

edc_lcover_*scheme*_XX_domclass.asc
where *scheme* the land cover classification scheme (IGBP or SiB)
XX qd, hd, or 1d, denoting a spatial resolution of 1/4, 1/2 or 1 degrees, respectively.

The fractional files are called:
edc_lcover_*scheme*_XX_cZZ.asc
where *scheme* the land cover classification scheme (IGBP or SiB)
XX qd, hd, or 1d, denoting a spatial resolution of 1/4, 1/2 or 1 degrees, respectively.

ZZ a number from 00 to 15 (or 17 for IGBP) which represents the land cover type code as described in Section 8.2.

As an example, the file named edc_lcover_igbp_qd_c02.asc is the fraction of Evergreen Broadleaf Forest at a quarter degree spatial resolution. This file is associated with the edc_lcover_igbp_qd_domclass.asc dominant land cover type file. At this revision the BATS (Biosphere-Atmosphere Transfer Scheme) classification products are not available.

**1.3 Revision Date of this Document**

October 30, 2009

## 2. INVESTIGATOR(S)

**2.1 Investigator(s) Name and Title**

Tom Loveland, Jess Brown, Don Ohlen, Brad Reed, and Zhiliang Zhu (U.S. GeologicalSurvey/EROS Data Center), Limin Yang (University of Nebraska-Lincoln)

**2.2 Title of Investigation**

Global Land Cover Characterization Project.

**2.3 Contacts (For Data Production Information)**

| | Contact 1 | Contact 2 |
|---|---|---|
| 2.3.1 Name | Stephen M. Howard | Tom Loveland |
| 2.3.2 Address | EROS Data Center<br>47914 252nd St | EROS Data Center<br>47914 252nd St |
| City/St. | Sioux Falls, SD | Sioux Falls, SD |
| Zip Code | 57198 | 57198 |
| Country | USA | USA |
| 2.3.3 Tel. No. | 605 594 6114 | 605 594 6066 |
| Fax No. | 605 594 6529 | 605 594 6529 |
| 2.3.4 E-mail | lcac@usgs.gov | Loveland@usgs.gov |

| | Contact 3 |
|---|---|
| 2.3.1 Name | Eric Brown de Colstoun |
| 2.3.2 Address | NASA/GSFC<br>Code 614.4 |
| City/St. | Greenbelt, MD |
| Zip Code | 20771 |
| Country | USA |
| 2.3.3 Tel. No. | (301) 614-6597 |
| Fax No. | (301) 614-6695 |
| 2.3.4 E-mail | ericbdc@ltpmail.gsfc.nasa.gov |

**2.4 Data Set Citation**

Loveland, T., J. Brown, D. Ohlen, B. Reed, Z. Zhu, L. Yang, and S. Howard . 2009. ISLSCP II IGBP DISCover and SiB Land Cover, 1992-1993. In Hall, Forrest G., G. Collatz, B. Meeson, S. Los, E. Brown de Colstoun, and D. Landis (eds.). ISLSCP Initiative II Collection.Data set. Available on-line [http://daac.ornl.gov/] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A. doi:10.3334/ORNLDAAC/930

## 3. INTRODUCTION

### 3.1 Objective/Purpose

The DISCover data set was developed under the auspices of the International Geosphere-Biosphere Programme (IGBP). DISCover development was co-ordinated by the IGBP-Data and Information System's (IGBP-DIS) Land Cover Working Group (LCWG). A number of IGBPinitiatives required global land cover data that improved on the existing databases. The IGBP's Data and Information System began the 1 km DISCover land cover project to obtain data from the Advanced Very High Resolution Radiometer (AVHRR) for all terrestrial surfaces and then derive land cover data sets from this archive for the IGBP's core science elements. The development of this dataset was endorsed by the Committee on Earth Observations Satellites,and implemented by the U.S. Geological Survey (USGS), NASA, National Oceanic and Atmospheric Administration, U.S. Environmental Protection Agency, United Nations Environment Programme, and the European Commission's Joint Research Centre. Over 4.4 Terabytes of 1 km resolution AVHRR data from 23 receiving stations were collected, assembled and processed to create the 1 km resolution DISCover global land cover product. DISCover, based on IGBP requirements and the known limitations of the AVHRR data concerning separation of cover types, was created from these data. The specifications for DISCover included a consistent and validated global characterization of global land cover using the 17 category DISCover land cover classification legend. This legend was designed to represent the general structural elements of land cover and vegetation, be consistent with previous global data sets, provide categories suited for use in a wide range of models, and represent broad categories ofland use. The DISCover legend was intended for use in global-scale modelling of climate,biogeochemistry and other Earth system processes. The original 1km resolution data set was aggregated to spatial resolutions of 1/4, 1/2 and 1 degree as a contribution to the ISLSCP Initiative II data collection.

### 3.2 Summary of Parameters

The data set describes the geographic distributions of 17 classes of land cover based on the International Geosphere-Biosphere DISCover land cover legend (Loveland and Belward 1997) and the 15 classes of the SiB model. Specifically, the resampled DISCover datasets were derived from the 1km DISCover dataset compiled by the USGS. The 1km datasets for each classification scheme were aggregated to 1, 0.5 and 0.25 degree spatial resolutions for this ISLSCP II data collection. Each layer of the aggregated products corresponds to a single DISCover land cover category and the values represent the percentage of the coarse resolution cell (1 degree, etc…) occupied by that land cover category. The dominant class data show the land cover category that occupies the majority of the cell and is derived from the percentage files for each cover type.

### 3.3 Discussion

The original data set is derived from 1-km Advanced Very High Resolution Radiometer (AVHRR) data spanning a 12-month period (April 1992-March 1993) and is based on a flexible data base structure and seasonal land cover regions concepts. Seasonal land cover regions provide a framework for presenting the temporal and spatial patterns of vegetation in the data base. The regions are composed of relatively homogeneous land cover associations (for example, similar floristic and physiognomic characteristics), which exhibit distinctive phenology (that is, onset, peak, and seasonal duration of greenness), and have common levels of primary production. Rather than being based on precisely defined mapping units in a predefined land cover classification scheme, the seasonal land cover regions serve as summary units for both descriptive and quantitative attributes. The seasonal land cover regions were then evaluated and assigned to a land cover category of the IGBP land cover mapping scheme (i.e. DISCover). The accuracy of the global land cover data set was determined through an independent accuracy assessment sponsored by the IGBP, NASA, and the USGS. The results of the validation can be found in Scepan (1999) with further elaboration by DeFries and Los(1999), and Loveland et al. (1999). Brown et al. (1999) provides details on applications of the DISCover and other global land cover data sets.

In addition, the original USGS global land cover database includes 1km resolution datasets based on the Simple Biosphere Program, Biosphere-Atmosphere Transfer Scheme, Olsen Global Land Cover, and U.S. Geological Survey Anderson System land cover classification schemes. For a complete discussion of the background of the 1 km dataset and access to the original 1km data sets, please see http://edc2.usgs.gov/1KM/ and Loveland et al. (1999, 2001).

Because the land/water boundaries of these resampled maps did not always agree with those of the ISLSCP II land/water masks, both the thematic land cover type files and the land cover fraction files have been modified to agree with the land and water fractions of the ISLSCP II land/water mask. This has not yet been done for the BATS scheme because this scheme contains an inland water category that does not agree with the ISLSCP II inland water mask for this task.

## 4. THEORY OF ALGORITHM/MEASUREMENTS

The overall classification strategy is a multi-temporal unsupervised classification with postclassification refinement using multi-source digital ancillary spatial data. The masked monthly NDVI composites representing a 12-month period were classified using an unsupervised technique to define preliminary greenness classes. The translation of the preliminary greenness classes into seasonal land cover regions required the addition of digital ancillary data, such as elevation and ecoregions data, as well as a collection of other land cover/vegetation reference data. The interpretation was based on extensive use of computer-assisted image processing tools; however, the classification process was far from automated and more closely followed a traditional manual image interpretation philosophy. The image classification process relied on the skills of the human interpreters to make the final decisions regarding the relationship between spectral classes defined using unsupervised methods and landscape characteristics that lead to specific land cover definitions. The computer-assisted land cover analysis was a complex task requiring sophisticated tools to explore, visualize, integrate, and extract information from a wide variety of spatial and non-spatial data. To develop the global land cover database, AVHRR data were integrated with other sources of data in order to create regions of land cover, for labeling and interpreting the regions, and to resolve occurrences where more than one type of land cover was represented in the same spectrally-defined cluster. Ancillary data were essential as aids to labeling, interpretation, and post-classification refinement of a classified satellite data set. However, because these ancillary data existed in a variety of forms including text, tables, spreadsheets, images, and other types of graphics, their use in the interpretation process was logistically challenging.

## 5. EQUIPMENT

### 5.1 Instrument Description

The global land cover data set was based on AVHRR maximum monthly composites for 1992-93 bands 1-5 and derived NDVI at approximately 1 km resolution (see Eidenshink and

Faundeen 1994).

**5.1.1 Platform (Satellite, Aircraft, Ground, Person)**

The AVHRR instrument is flown on the National Oceanic and Atmospheric Administration (NOAA) series of satellite platforms.

**5.1.2 Mission Objectives**

AVHRR was designed for the instantaneous observation of clouds, ocean, land, ice and snow cover for weather analysis purposes. The multi-spectral measurements have been proven to be suitable for the quantitative measurement of a number of parameters that AVHRR was originally not designed for. The long data record also allows the use of AVHRR data for climate analysis purposes.

**5.1.3 Key Variables**

All 5 spectral bands of the AVHRR were used as inputs: channel 1 (visible red reflectance, 0.58-0.68 microns), channel 2 (near infrared reflectance, 0.725-1.1 microns), channel 3 (thermal infrared, 3.55-3.93 microns), channel 4 (thermal, 10.3-11.3 microns), channel 5 (thermal, 11.5-12.5 microns) and the NDVI (channel 2- channel 1)/(channel 2 + channel 1).

**5.1.4 Principles of Operation**

AVHRR, a scanning radiometer, is operated and maintained by the National Environmental Satellite Data and Information Service (NOAA/NESDIS).

**5.1.5 Instrument Measurement Geometry**

AVHRR operates with a cross-track scanning system with a maximum of 55.4 degrees scan angle from the nadir. The nominal resolution of the sub-satellite point is 1.1 km for Local Area Coverage (LAC) and 4 km for Global Area Coverage (GAC) data. The spatial resolution decreases substantially towards the edges of the orbital swath.

**5.1.6 Manufacturer of Instrument**

ITT, Fort Wayne, IN.

**5.2 Calibration**

5.2.1 Specifications

**5.2.1.1 Tolerance**

See Eidenshink and Faundeen (1994) for more details on the production of the global 1km AVHRR data set.

**5.2.2 Frequency of Calibration**

See Eidenshink and Faundeen (1994) for more details.

**5.2.3 Other Calibration Information**

None.

## 6. PROCEDURE

**6.1 Data Acquisition Methods**

One-kilometer AVHRR NDVI composites spanning April 1992 through March 1993, are the core data set used in land cover characterization. In addition, other key ancillary geographic data include digital elevation data, ecoregions interpretations, and country or regional-level vegetation and land cover maps were used to help characterize the global land cover. The global land cover characteristics database was developed on a continent-by-continent basis at 1 km resolution. The geographic projection was obtained by resampling the 1 km data to 30 arc seconds. The resampled global datasets share the same map projection (Geographic). For a detailed discussion on the creation of the 1km land cover data sets, see: http://edc2.usgs.gov/1KM/.

**6.2 Spatial Characteristics**

**6.2.1 Spatial Coverage**

The coverage is global. Data in files are ordered from North to South and from West to East beginning at 180 degrees West and 90 degrees North.

**6.2.2 Spatial Resolution**

The data are given in an equal-angle lat/long Earth grid that has three spatial resolutions of 0.25 x 0.25, 0.5 x 0.5 and 1.0 x 1.0 degree lat/long.

**6.3 Temporal Characteristics**

**6.3.1 Temporal Coverage**

The data set is derived from data collected from April 1992 to March 1993.

**6.3.2 Temporal Resolution**

This data set represents the land cover types present during the period from April 1992 to March. The temporal resolution is thus one year.

## 7. OBSERVATIONS

**7.1 Field Notes**

Not applicable to this data set.

## 8. DATA DESCRIPTION

**8.1 Table Definition with Comments**

Not applicable to this data set.

**8.2 Type of Data**

| 8.2.1 Parameter/ Variable Name | 8.2.2 Parameter/ Variable Description | 8.2.3 Data Range | 8.2.4 Unit of Measurement | 8.2.5 Data Source |
|---|---|---|---|---|
| | | | | |
| edc_lcover_igbp | Dominant IGBP land cover class within each grid cell. See class definitions below this table. | 0-17 | See 8.2.2 | IGBP DISCOver |
| edc_lcover_igbp | Fraction of each IGBP land | 0-100 | Unitless | IGBP |

| | | | | |
|---|---|---|---|---|
| _cXX | cover class contained within each grid cell. There is one layer each for each IGBP land cover land cover class listed below this table. | | | DISCOver |
| edc_lcover_sib | Dominant SiB land cover class within each grid cell. See class definitions below this table. | 0-15 | See 8.2.2 | IGBP DISCOver |
| | | | | |
| edc_lcover_sib_cXX | Fraction of each SiB land cover class contained within each grid cell. There is one layer each for each SiB land cover class listed above. | 0-100 | Unitless | IGBP DISCOver |

IGBP Land Cover Type Codes and Definitions

| IGBP Land Cover Code | Definition | Explanation |
|---|---|---|
| 1 | Water Bodies | Oceans, seas, lakes, reservoirs, and rivers. Can be either fresh or salt water bodies. This has been adjusted to match the ISLSCP II land/water masks. |
| 2 | 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. |
| 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 with 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 | Lands with woody vegetation less than 2 meters tall and |

| | | |
|---|---|---|
| | Shrublands | 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 forest canopy cover of 30-60%. The forest cover height exceeds 2 meters. |
| 9 | Savannas | Lands with herbaceous and other understory systems and forest canopy cover of 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 surface water and herbaceous or woody vegetation. The vegetation can be present in either salt, brackish, or fresh water. |
| 12 | Croplands | 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 primarily covered by buildings and other man-made structures. Note that this class has not been mapped directly from AVHRR data. It is overlaid from the populated places layer from the Digital Chart of the World. |
| 14 | Cropland/ Natural Vegetation Mosaic | Lands with a mosaic of croplands, forests, shrublands, and grasslands in which no one component comprises more than 60% of the landscape. |
| 15 | Permanent snow and Ice | Lands under snow and/or ice cover throughout the year. |
| 16 | Barren or Sparsely Vegetated | Lands with exposed soil, sand, rocks, or snow that never has more than 10% vegetated cover during any time of the year. |
| 17 | unclassified | Found in some coastal zones and small islands. |

**SiB Land Cover Type Codes and Definitions**

| **SiB Land Cover** | **Definition** |
|---|---|

| Code | |
|---|---|
| 0 | Water Bodies |
| 1 | Evergreen Broadleaf trees |
| 2 | Broadleaf Deciduous Trees |
| 3 | Deciduous and Evergreen Trees |
| 4 | Evergreen Needleleaf Trees |
| 5 | Deciduous Needleleaf Trees |
| 6 | Ground Cover with Trees and Shrubs |
| 7 | Ground Cover with Trees and Shrubs |
| 8 | Broadleaf Shrubs with Perennial Ground Cover |
| 9 | Broadleaf Shrubs with Bare Soil |
| 10 | Groundcover with Dwarf Trees and Shrubs |
| 11 | Bare Soil |
| 12 | Agriculture or C3 Grassland |
| 13 | Persistent Wetland |
| 14 | Ice Cap and Glacier |
| 15 | Missing Data |

**8.3 Sample Data Record**

Not applicable to this data set.

**8.4 Data Format**

All of the files in the ISLSCP Initiative II data collection are in the standard ARC GIS ASCII Grid format. The file format consists of six lines of header information followed by numerical fields of varying length, which are delimited by a single space and arranged in columns and rows. The files at different spatial resolutions each contain the following numbers of column and rows:

One degree: 360 columns by 180 rows
1/2 degree: 720 columns by 360 rows
1/4 degree: 1440 columns by 720 rows
All values are written as floating point values. Missing values are assigned the value of –99 on all data layers. Missing values over land are assigned the value of -88.

All files are gridded to a common equal-angle lat/long grid, where the coordinates of the upper left corner of the files are located at 180 degrees W, 90 degrees N and the lower right corner coordinates are located at 180 degrees E, 90 degrees S. Data in the files are ordered from North to South and from West to East beginning at 180 degrees West and 90 degrees North.

The ASCII map files (with the extension of ".asc") have all had the ISLSCP II land/water mask applied to them. All points removed from the original files are stored in "differences" files (with the extension ".dif"). These ASCII files contain the Latitude and Longitude location of the cell-center of each removed point, and the data value at that point. There is one ".dif" file for each ASCII map file.

**8.5 Related Data Sets**

Other land cover data sets in the ISLSCP II collection include the University of Maryland (UMD) land cover and vegetation continuous fields data sets, the MODIS land cover classification for 2001, and the C4 vegetation fraction data set. Historical land cover and cropland cover data sets are also available in this collection.

## 9. DATA MANIPULATIONS

### 9.1 Formulas

#### 9.1.1 Derivation Techniques/Algorithms

The global land cover characteristics database was developed on a continent by continent basis through the unsupervised classification of AVHRR 1km NDVI monthly composites, followed by extensive post-classification refinement using ancillary data. This approach was considered to be appropriate because: (1) a relatively detailed land cover characterization was required; (2) the global landscape is vast and heterogeneous; (3) the AVHRR composites are of variable quality; and (4) there is no systematic source of consistent and detailed reference data from which to base training decisions.

Data Processing Sequence

For a full description of the processing of the AVHRR data, see http://edc2.usgs.gov/1KM/.

#### 9.2.1 Processing Steps and Data Sets

The classification methodology involves a sequence of six analysis steps:

*AVHRR Recompositing and Quality Assessment* The 10-day composites were consolidated into monthly composites using maximum value compositing and evaluated to document image characteristics and problems. Previous experience has shown that through the recompositing process, the overall data volume can be reduced by two-thirds and composite quality can be improved, while still providing phenological information relevant to the land cover classification process. Data volume reduced from 23.84 gibabytes for the 12-month global 10-day composites to 7.95 gigabytes for monthly composites while still preserving the annual sequence of phenological development. Improvements in data quality can be expected by using monthly composite because a longer time span increases both the likelihood of cloud-free coverage and overall data completeness. This is particularly important in areas affected by ground station acquisition problems or policies and persistent clouds. For some 10-day periods, there have been gaps in the composites where no data were available. Expanding the temporal window reduced these problems. The AVHRR quality assessment process provided a means to identify correctable deficiencies in the AVHRR composites that would adversely affect the land cover classification. Examples of image characteristics that were identified include: (1) gross image misregistration where landscape features (i.e., stream channels, coastlines, lakes) were clearly offset; (2) gross radiometric anomalies caused by processing blunders (i.e., application of incorrect radiometric calibration coefficients; absence of atmospheric corrections); (3) gaps resulting from missing data; (4) presence of mosaic or composite lines; (5) problems associated with the inclusion of images outside the composite period; and (6) excessive cloud contamination.

*Mask Preparation* Masks of barren or sparsely vegetated areas (water bodies, barren, and snow and ice) were generated. While NDVI is appropriate for the identification of vegetated land cover patterns and characteristics, it is not suited to the discrimination of cover patterns within non-vegetated landscapes. The NDVI of non-vegetated lands has a high variance, and when statistical clustering techniques are applied, the non-vegetation pixels are segmented into an excessive number of irrelevant classes that detract from the segmentation of vegetated landscapes. Consequently, the NDVI data for non-vegetated areas were masked prior to the classification of the 12-month set of monthly composites. or DCW. In some cases, the DCW hydrography is quite dated and exaggerates the actual spatial extent of some water bodies (e.g., Aral Sea of Kazakstan and Uzbekistan, Lake Chad of Chad, Nigeria, and Niger, and Lake Turkana of Kenya). Barren, sparsely vegetated, and snow and ice cover areas were identified through the thresholding of a maximum greenness composite representing the April 1992-March 1993 period. When maximum greenness NDVI values are less than 0.04-0.10, it can be assumed that vegetation densities are below 5-10% cover. The interpretation of non-vegetated surfaces used a conservative strategy that initially led to underestimation of barren areas. However, subsequent classification and segmentation of the vegetated landscape provided an additional opportunity to identify barren and sparsely vegetated lands. Therefore, differences caused by variable soil color, illumination levels, or other environmental factors could still be defined and interpreted.

*Unsupervised Classification/Preliminary Greenness Classes* The set of monthly NDVI composites were clustered using an unsupervised classification strategy, cluster attributes were generated, and preliminary greenness class interpretations were developed. The interpretation was based on the use of computer-assisted image processing tools; however, the classification process was far from automated and more closely followed a traditional manual image interpretation approach. Unsupervised techniques have been applied with success in other classification problems in which AVHRR data were used. In this study, unsupervised classification provided meaningful mapping units that complemented the database concept used in this project. It can be argued that the relationship between spectral data and specific land cover classes is, at best, ambiguous. As a result, it can also be argued that the process of assigning land cover type labels to spectral classes, whether done through an unsupervised or supervised process, is a significant potential source of classification error. Therefore, spectral regions defined using an unsupervised strategy can be treated as classic geographic regions (i.e., having patterns more similar within than outside their boundaries), and can serve as models of local landscape diversity. Considering that digital image classification is far from automated (only the image segmentation is automated), classifications, whether supervised or unsupervised, are based on a series of local decisions that aggregate to a global land cover classification. Those local interpretations provide a means to document the environmental diversity represented within individual spectral regions. Thus, the unsupervised strategy used in this research was intended as the starting point for segmenting the global AVHRR data into local regions that can be interpreted using a combination of traditional and automated image analysis techniques and tools. Clustering of the masked continental AVHRR monthly composite sets was done using the “Los Alamos” clustering algorithm developed by Kelly and White (1993). This algorithm is optimized for efficient use with large data sets. It uses a Monte Carlo random sampling approach in which a new sample is selected for each clustering iteration. In addition, the sample used for each iteration is quite small, typically one-percent of the input data set, which increases clustering performance. The clustering is based on the KMeans technique to develop cluster centers and vectors, and the minimum distance to the mean classifier is used to assign pixels to corresponding classes. The number of clusters created for each continent was based the collective judgement of the project team, and considered continental data set size, data quality, and environmental variability.

For example, Eurasia was the largest and most diverse continent, so 150 clusters were defined. However, only 80 clusters were defined for South America, the smallest continent. The remaining three continents had 100 clusters created. Originally, a small number of clusters (50-70) were planned for Australia-Pacific. However, because of poor data quality, the islands of Borneo, Sumatra, and Papua-New Guinea were clustered separately using a smaller number of clear composites. This increased the total number of Australia-Pacific clusters to 100.

The clustered data represented preliminary greenness classes that corresponded to homogeneous patterns of seasonality and were related to relative patterns of productivity. However, in many cases the preliminary greenness classes represented multiple disparate land cover classes. Preliminary labels were developed that provided a general understanding of the characteristics of each cluster or preliminary greenness class, and classes with two or more disparate land cover classes were identified. The image interpretation process followed the traditional "convergence of evidence" strategy that is commonly used in air photo interpretation. Image features, including the spatial pattern of classes and their location and association with other classes, the annual NDVI sequence, site characteristics, and relationships between patterns and reference materials were all used in the labeling of the preliminary greenness classes. The draft descriptions were based on a wide range of references, including digital and hardcopy land cover maps, atlases, and Landsat imagery. Because individual interpreters may have a biased perspective of particular areas based on their discipline background, interpretation experience, and familiarity with the study area, three or more interpreters independently labeled each class. Where differences existed, the interpreters compared decisions and consulted reference materials in order to arrive at a consensus.

*Post-Classification Refinement/Seasonal Land Cover Regions* Several methods were used to stratify and transform preliminary greenness classes representing multiple disparate land cover characteristics into seasonal land cover regions. Seasonal land cover regions, by definition, have similar mosaics of land cover types and common seasonal properties. The seasonal land cover regions development process involved splitting the heterogeneous preliminary greenness classes into relatively homogenous land cover using a post-classification refinement process, and then creating land cover descriptions and attributes for each region. Much of the land cover confusion was the result of spectral similarities between natural and agricultural land cover. Developing criteria based on the relationship between the confused seasonal greenness classes and selected ancillary data sets solved these problems. The classes with multiple cover types were split into smaller, more homogenous regions using a variety of methods, including: (1) Ancillary data splits – Selected ancillary data sets (i.e., elevation, ecoregions) were used to subdivide the heterogeneous classes into "pure" seasonal land cover regions. (2) User-defined polygons – Also referred to as "on-screen digitizing", this approach was used when ancillary data did not provide the appropriate spatial context for dividing the class into homogenous parts. (3) Multi-source combinations – In some cases, ancillary data were augmented with user-defined polygons to develop the necessary mask for post-classification refinement. This typically involved developing new regions based on a combination of both elevation and ecoregions. (4) Spectral reclustering – Reclustering, using either different clustering parameters or a smaller set of NDVI composites were used to break single classes into a number of smaller classes. This stratification approach was used the least. After the preliminary greenness classes were stratified into seasonal land cover regions, final land cover attributes were formulated. As before, each seasonal land cover region was materials was generally unknown, agreement between multiple references and consensus among the interpreter teams was required before class descriptions were finalized. A total of 961 seasonal land cover regions were produced for the five vegetation-dominated continents. Eurasia, the largest and most diverse landmass, had 255 seasonal land cover regions, while Australia-Pacific, the smallest, had 137 regions. For North America, 205 regions were defined, 197 were developed for Africa, and 167 seasonal land cover regions were recognized for South America. The total number of Eurasia seasonal land cover regions was constrained by the need to keep the number of classes in an 8-bit range so that the data set could be used in most image processing and geographic information systems. Once the final seasonal land cover regions were defined, attributes were developed for each individual seasonal land cover region. This included land cover assignments for Olson Global Ecosystems, IGBP DISCover, SiB, SiB2, BATS, Running, and USGS Anderson classification systems. It should be noted that the seasonal land cover regions are continent-specific and definitions are not standardized between continents. These seasonal land cover regions are the fundamental spatial unit of the database and the basis for developing general land cover categorizations. They provide a detailed representation of the global patterns of land cover and environmental diversity, and offer a unique way of visualizing the interplay of land cover, seasonality, and productivity. However, they are descriptive and must be used with considerable caution because they lack the standardization necessary for some scientific applications. While they indicate important landscape properties, they require validation to verify their contents.

The derived land cover data sets were developed using a methodology that first related the individual seasonal land cover regions to Olson's Global Ecosystems (Olson, 1994), an update of the 49 class system developed by Olson and Watts (1982), which were then cross-walked into the other land cover legends. The Olson legend was used as the bridging system because it: (1) has sufficient thematic detail (94 potential classes) and was developed for global applications; (2) has been used for large area modeling and has links to landscape productivity, particularly carbon stocks; (3) recognizes anthropogenic elements of the landscape; (4) recognizes landscape mosaics that occur at coarse resolutions; and (5) includes attributes on climate and physiognomy, and implies floristic elements. A look-up table was developed that provided the relationship of each Olson class to the corresponding classes in the other general land cover legends. Once a seasonal land cover region was assigned to an Olson category, it was automatically related to any one of the six other land cover classification systems. The advantage of this approach was that it increased the efficiency and improved the consistency for assembling the global derived land cover layers. In addition, it facilitates the development of new data sets when additional user requirements are identified. However, while this strategy was generally effective, it was still necessary to review individual translations to verify class assignments. A review of the results of the Olson to other land cover legend translation process revealed that less than 5% of all original assignments required modification.

Urban land cover could not be consistently classified using multi-temporal NDVI data. The heterogeneous nature of urban land cover, resulting from the complex patterns of land use, as well as the coarse resolution of 1-km AVHRR data, make it practically impossible to map urban land cover using computer-assisted image classification methods. Thus, the required urban land cover data came from the populated places data layer in the Defense Mapping gency's Digital Chart of the World, or

DCW. This data layer was derived from 1:1,000,000-scale maps and is therefore at a compatible resolution. A significant limitation of the DCW urban data is that they are drawn from maps of varying ages, ranging from the 1960's to 1980's. Typically the data from the rapidly urbanizing developing world are from the oldest sources.

#### 9.2.2 Processing Changes

None.

#### 9.2.3 Additional Processing by the ISLSCP II Staff

Some discrepancies were found between the ISLSCP II land/water mask and the water/land values in the EDC land cover products. To address these issues, the original EDC products were made to match with the water fractions of the ISLSCP II land/water mask and a new dominant land cover type map was derived. Two general cases were addressed: 1) The ISLSCP II mask is water and the EDC map is land, 2) The ISLSCP II mask is land and the EDC product is water. For 1), the original EDC fractions for each land cover category were adjusted using the land fractions of the ISLSCP II mask. For all cells in this category, the original EDC land cover type (SiB or IGBP) was replaced with a value of 0 (water). For cases in 2), if the EDC water fraction was less than 100%, the existing EDC fractions were adjusted as in 1). In cases were the EDC water fraction was 100% in 2), the cell in all land cover fraction files was filled from an average of all surrounding cells in a 3 by 3 window. In a few instances such as small islands, no land values were available in the 3 by 3 window and the cell was labeled as "Unclassified" (i.e. SiB=15 or IGBP=17). The dominant land cover types were then derived using the new fraction files. The class with the largest fraction was arbitrarily assigned as the dominant land cover type in the cell.

### 9.3 Calculations

#### 9.3.1 Special Corrections/Adjustments

None.

### 9.4 Graphs and Plots

See http://edc2.usgs.gov/1KM/

## 10. ERRORS

### 10.1 Sources of Error

Sources of error include the data inputs (misplaced swaths, noise, misregistration, cloud contamination, missing data, etc...) Classification error also resulted from interpreter mistakes, incomplete or erroneous reference data, or the ambiguous relationship between land cover and multitemporal satellite reflectance values.

### 10.2 Quality Assessment

#### 10.2.1 Data Validation by Source

This dataset was validated through peer review in which expert users provided evaluations and feedback on classification strengths and weaknesses. In addition, the data were validated using an independent statistical validation. Landsat Thematic Mapper and SPOT satellite image data were used to verify 379 primary core samples selected from DISCover 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. A complete summary of the statistical validation and validation results can be found in Scepan (1999).

#### 10.2.2 Confidence Level/Accuracy Judgment

Scepan (1999) reports that for the 15 DISCover classes (excluding water and urban), the average class accuracy was 59.4% with accuracies for the 15 verified DISCover classes ranging between 40.0% and 100%. The overall area weighted accuracy of the data set was determined to be 66.9%. 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%.

#### 10.2.3 Measurement Error for Parameters and Variables

None.

#### 10.2.4 Additional Quality Assessment Applied

None.

## 11. NOTES

### 11.1 Known Problems with the Data

See http://edc2.usgs.gov/1KM/. Also see Section 10.2 and consult Scepan (1999) for more information.

### 11.2 Usage Guidance

The DISCover data set was developed to be consistent at global to continental scales. As a result, the data should be used for continental to global applications.

When aggregating the 1km maps to coarser scales, many errors are reduced, such as the limitation in depicting spatial heterogeneity. Users should note that the dominant land cover type files are consistent with the ISLSCP II binary land/water mask while the land cover fraction files are consistent with the land and water fraction files of the ISLSCP II land/water mask. Users can utilize the land cover fraction files to generate different land cover products with their own rules.

## 12. REFERENCES

### 12.1 Satellite/Instrument/Data Processing Documentation

See http://edc2.usgs.gov/1KM/.

### 12.2 Journal Articles and Study Reports

Brown, J.F., Loveland, T.R., Ohlen, D.O., and Zhu, Z., 1999. The global land-cover characteristics database--the users' perspective: *Photogrammetric Engineering and Remote Sensing* 65(9): 1,069-1,074.

DeFries, R.S. and S.O. Los, 1999. Implications of land-cover misclassification for parameter estimates in Biosphere-Atmosphere models. *Photogrammetric Engineering & Remote Sensing*, 65(9): 1083-1088.

Eidenshink, J.C. and Faundeen, J.L., 1994. The 1km AVHRR global land data set: first stages in implementation. *International Journal of Remote Sensing* 15: 3443-3462.

Loveland, T.R., Zhu, Z., Ohlen, D.O., Brown, J.F., Reed, B.C., and Yang, L., 1999. An analysis of the IGBP global land-cover characterization process. *Photogrammetric Engineering and Remote Sensing* 65(9): 1,021-1,032.

Loveland, T.R., Reed, B.C., Brown, J.F., Ohlen, D.O., Zhu, J, Yang, L., and Merchant, J.W., 2001. Development of a Global Land Cover Characteristics Database and IGBP DISCover from 1-km AVHRR Data. *International Journal of Remote Sensing* 21(6/7): 1,303-1,330.

Scepan, J., 1999. Thematic Validation of High-Resolution Global Land-Cover Data Sets. *Photogrammetric Engineering and Remote Sensing* 65(9): 1,051-1,060.

## 13. DATA ACCESS

### 13.1 Data Access Information

The ISLSCP Initiative II data are archived and distributed through the Oak Ridge National Laboratory (ORNL) DAAC for Biogeochemical Dynamics at http://daac.ornl.gov.

### 13.2 Contacts for Archive:

E-mail: uso@daac.ornl.gov
Telephone: +1 (865) 241-3952

### 13.3 Archive/Status/Plans

The ISLSCP Initiative II data are archived at the ORNNL DAAC. There are no plans to update these data.

## 14. GLOSSARY OF ACRONYMS

| | |
|---|---|
| AVHRR | Advanced Very High Resolution Radiometer |
| BATS | Biosphere-Atmosphere Transfer Scheme |
| CD-ROM | Compact Disk (optical), Read Only Memory |
| DAAC | Distributed Active Archive Center |
| DCW | Digital Chart of the World |
| EOS | Earth Observing System |
| EDC | EROS Data Center |
| GAC | Global Area Coverage |
| GCM | General Circulation Model of the atmosphere |
| GSFC | Goddard Space Flight Center |
| IDS | Inter-disciplinary Science |
| IGBP | International Geosphere-Biosphere Programme |
| IGBP-DIS | IGBP-Data and Information Service |
| ISLSCP | International Satellite Land Surface Climatology Project |
| LAC | Local Area Coverage |

| | |
|---|---|
| LCWG | Land Cover Working Group (IGBP-DIS) |
| MODIS | Moderate Resolution Imaging Spectroradiometer |
| NASA | National Aeronautics and Space Administration |
| NDVI | Normalized Difference Vegetation Index |
| NESDIS | National Environmental Satellite Data and Information Service |
| NOAA | National Oceanic and Atmospheric Administration |
| ORNL | Oak Ridge National Laboratory |
| SiB | Simple Biosphere model |
| UMD | University of Maryland |
| USGS | Unites States Geological Survey |