1. TITLE

1.1 Data Set Identification

ISLSCP II Continuous Fields of Vegetation Cover, 1992-1993.

1.2 File Name(s)

The data sets in this directory are provided at three spatial resolutions of 0.25, 0.5 and 1 degrees lat/long. For each spatial resolution there are eight files describing the percentage, from 0 to 100, of the following global continuous fields:

1) Bare cover
2) Herbaceous cover
3) Tree cover
4) Water
5) Deciduous tree cover
6) Evergreen tree cover
7) Broadleaf tree cover
8) Needleleaf tree cover

The files for 1) are called **bare_percent_xx.asc**, where xx is qd, hd, or 1d, denoting a spatial resolution of 1/4, 1/2 or 1degree, respectively. The files for 2) are called **herb_percent_xx.asc**, with xx as above, and so on for the different continuous fields.

1.3 Revision Date of this Document
October 26, 2009

2. INVESTIGATOR(S)

2.1 Investigator(s) Name and Title

Dr. Ruth DeFries
Department of Ecology, Evolution and Environmental Biology (E3B)
Columbia University, New York

Dr. Matthew C. Hansen
Geographic Information Science Center of Excellence
South Dakota State University, Brookings

2.2 Title of Investigation
Characterizing heterogeneity of global vegetative cover.

2.3 Contacts (For Data production Information)

	Contact 1	Contact2
2.3.1 Name	Dr. Ruth DeFries	Dr. Matthew C. Hansen
2.3.2 Address	Dept. of Ecology, Evolution and Environmental Biology Columbia University 10th Floor Schermerhorn Ext.	Geographic Information Science Center of Excellence South Dakota State University 1021 Medary Ave., Wecota Hall Box 506B
City/St.	New York, NY	Brookings, SD
Zip Code	10027	57007
Country	USA	USA
2.3.3 Tel. No.	301 405-4884	605 688-6848
2.3.4 E-mail	rdefries@geog.umd.edu	Matthew.Hansen@sdstate.edu

	Contact 3
2.3.1 Name	Dr. Eric Brown de Colstoun
2.3.2 Address	NASA/GSFC Code 614.4
City/St.	Greenbelt, MD
Zip Code	20771
Country	USA
2.3.3 Tel. No.	(301) 614-6597
2.3.4 E-mail	eric.c.browndecolsto@nasa.gov

2.4 Data Set Citation:
DeFries, R., M.C. Hansen. 2009. ISLSCP II Continuous Fields of Vegetation 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/931

2.5 ISLSCP Initiative II Collection References
Users of the International Satellite Land Surface Climatology (ISLSCP) Initiative II data collection are requested to reference the following publications when these data are used:

Hall, F.G., E. Brown de Colstoun, G. J. Collatz, D. Landis, P. Dirmeyer, A. Betts, G. Huffman, L. Bounoua, and B. Meeson, The ISLSCP Initiative II Global Datasets: Surface Boundary Conditions and Atmospheric Forcings for Land-Atmosphere Studies, *J. Geophys. Res.*, 111, doi:10.1029/2006JD007366, 2006.

DeFries, R. S., Townshend, J. R. G., and Hansen, M. C., 1999, Continuous fields of vegetation characteristics at the global scale at 1km resolution, *Journal of Geophysical Research*, 104, 16 911-16 925.

DeFries, R. S., Hansen, M. C., Townshend, J. R. G., Janetos, A. C., and Loveland, T. R., 2000, A new global 1-km dataset of percentage tree cover derived from remote sensing. *Global Change Biology*, 6, 247-254.

3. INTRODUCTION

3.1 Objective/Purpose

The objective of this study was to derive continuous fields of vegetation cover from multi-temporal Advanced Very High Resolution Radiometer (AVHRR) data using all available bands and derived Normalized Difference Vegetation Index (NDVI). The continuous fields describe sub-pixel proportions of cover for tree, herbaceous, bare ground and water cover types. For tree cover, additional fields describing leaf longevity (evergreen and deciduous) and leaf morphology (broadleaf and needleleaf) were also generated. The modeling of carbon dynamics and climate require knowing tree characteristics such as these. These products were resampled and aggregated to 0.25, 0.5 and 1.0 degree grids for the International Satellite Land Surface Climatology Project (ISLSCP) data initiative II.

3.2 Summary of Parameters

The data set describes the geographic distributions of three fundamental vegetation characteristics: tree, herbaceous and bare ground cover, plus a water layer. For tree cover, leaf longevity and morphology layers were produced. The data sets are provided at three spatial resolutions of 0.25, 0.5 and 1degrees (lat/long).

3.3 Discussion

The geographic distribution of vegetation over the land surface is traditionally described using classification schemes with discrete numbers of vegetation types. When such land cover data sets are used as boundary conditions in earth system models, an overly homogeneous land cover depiction with overly abrupt boundaries is introduced into parameter estimates. This study involves an alternative approach to describe global land cover with continuous fields of vegetation characteristics. A linear mixture model is applied to 1 km AVHRR data to estimate proportional cover for three important vegetation characteristics: life form (percent woody vegetation, percent herbaceous vegetation and bare ground), leaf type (percent needleleaf and percent broadleaf), and leaf duration (percent evergreen and percent deciduous). Linear discriminants for input into the mixture model are derived from 30 metrics representing the annual phenological cycle, such as maximum annual NDVI and minimum annual red reflectance. Through comparison with training data derived from a global network of Landsat Multi-Spectral Scanner (MSS) scenes, we conclude that the linear assumption implicit in the linear mixture model is not severely violated. The linear relationships between percent cover as determined from the training data and the linear discriminants are used to estimate endmember values and the mixture model is applied to derive the seven layers of global continuous fields (DeFries et al. 1999 and 2000). The original 1km continuous fields data produced by DeFries et al. (2000) were resampled and aggregated to coarser resolutions for ISLSCP II. Because the land/water boundaries of these aggregated continuous fields did not always agree with those of the ISLSCP II land/water masks, all data layers have been modified to agree with the land and water fractions of the ISLSCP II land/water mask.

4. THEORY OF ALGORITHM/MEASUREMENTS

Mixture models are used to partition a data set into percent contributions of ‘pure’ land cover types. These land cover types are delineated in multi-spectral space and are referred to as spectral endmembers. Every combination in multi-spectral space can then be modeled linearly as a percent contribution of each endmember. In this manner, the output map retains much of the variability seen in the original imagery. Only that variability relating directly to the mixtures being depicted is retained. The result is a map product with much greater detail than that found in a discrete classification.

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° scan angle from the nadir. The nominal resolution of the sub-satellite point is 1.1 km for LAC and 4 km for GAC. The spatial resolution decreases substantially towards the edges of the orbital swath.

5.1.6 Manufacturer of Instrument

ITT.

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

The AVHRR 1km data set was processed at the EROS Data Center under the guidance of the IGBP (International Geosphere Biosphere Programme) (Eidenshink and Faundeen 1994) .

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 grid that has spatial resolutions of 0.25 x 0.25, 0.5 x 0.5 and 1.0 x 1.0 degrees lat/long.

6.3 Temporal Characteristics

6.3.1 Temporal Coverage

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

6.3.2 Temporal Resolution

This data set represents vegetation conditions present during the period from April 1992 to March. The temporal resolution is yearly.

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 Units of Measurement	8.2.5 Data Source
Tree Cover	Percentage of each cell covered with trees	0-95 No Data= -999	Percent DeFries	et al. (1999, 2000)
Herbaceous Cover	Percentage of each cell covered with grasses/herbs	0-100 No Data= -999		
Bare Ground Cover	Percentage of each cell covered with bare ground	0-100 No Data= -999		
Deciduous Cover	Percentage of each cell covered with deciduous trees	0-85 No Data= -999		
Evergreen Cover	Percentage of each cell covered with evergreen trees	0-92 No Data= -999		
Needleleaf Cover	Percentage of each cell covered with needleleaved trees	0-95 No Data= -999		
Broadleaf Cover	Percentage of each cell covered with broadleaved trees	0-85 No Data= -999		
Water	Percentage of each cell covered with water	0-100		

NOTE***: The sum of the percent deciduous and evergreen layers equals the percent tree cover layer as does the sum of the needleleaf and broadleaf layers. Needleleaf deciduous forest is allowed to occur only east of 60 degrees East and north of 40 degrees North (largely Siberia). See section 9.3.1.

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 folloed by numerical fields of varying length, which are delimited by a single space and arranged in columns and rows. All values in these files are written as real numbers. A value of –999 is used to indicate cells with no data.

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 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.

8.5 Related Data Sets

For other global land cover data sets, see (DeFries and Townshend, 1994, DeFries et al. 1998, Loveland et al. 2000, Hansen et al. 2000). Other land cover data sets in the ISLSCP II collection include the EROS Data Center (EDC) Classifications, the MODIS classification, and the C_4 vegetation fraction data set.

9. DATA MANIPULATIONS

9.1 Formulas

9.1.1 Derivation Techniques/Algorithms

The principal steps used to produce the data sets are: 1) to use the 30 AVHRR metrics and the training data to calculate linear discriminants for use as variables in the linear mixture model, 2) determine endmember values of the linear discriminants based on training data, and 3) apply the linear discriminants and endmembers to a linear mixture model to derive the continuous fields.

Training data are required to derive the linear discriminants and to test the linear assumptions of the mixture model. Ideally, training data would be available for all ranges of percentage woody vegetation, herbaceous vegetation, and bare ground and percentages of each leaf type and leaf duration. Such information is not available, particularly with global coverage. Consequently, we used training data developed in previous research for the purpose of land cover classification derived from the AVHRR 8 km Pathfinder data set (DeFries et al. 1998). The linear mixture model works on the following principle. With three components and two linear discriminants (the case of woody, herbaceous and bare ground cover), the mixture model becomes:

$$R_1 = r_{1w}x_w + r_{1h}x_b + r_{1b}x_b \quad (1)$$

$$R_2 = r_{2w}x_w + r_{2h}x_b + r_{2b}x_b \quad (2)$$

$$x_w + x_h + x_b = 1 \quad (3)$$

where R_1 and R_2 are the first and second linear discriminants; r_{iw}, r_{ih} and r_{ib} are endmember values for woody vegetation, herbaceous vegetation and bare ground respectively; and x_w, x_h and x_b are fractional cover for woody vegetation, herbaceous vegetation and bare ground respectively. These simultaneous equations are then solved to determine the proportions of woody vegetation, herbaceous vegetation and bare ground for each pixel. For a detailed description of the methodology, see DeFries et al. (1999).

Preliminary mixture model results were observed to be overestimating tree coverage in locations known to have intermediate values characteristic of savannas and wooded grasslands. Consequently, we adjusted the mixture model results based on two land cover classifications previously derived from the same AVHRR data set, one of which (Hansen, et al. 2000) employed a supervised decision tree algorithm using all available spectral bands according to the methodology described in DeFries et al. (1998) and the other using an unsupervised classification of monthly values of the Normalized Difference Vegetation Index in conjunction with ancillary data (Loveland et al. 2000). Estimates of percent tree cover were scaled so that the lower and higher bounds were within the range of tree coverage defined for the cover type. Pixels classified as agriculture in the tropics were adjusted to range from 10 to 25 percent tree cover. Islands not present in the 1km data set were taken from the AVHRR 8km Pathfinder data (James and Kalluri 1994).

9.2 Data Processing Sequence

9.2.1 Processing Steps and Data Sets

Core training areas of 100 percent cover for each respective layer were delineated within the global training data set. These were used to calibrate the mixture model using the 30 AVHRR multi-temporal metrics. Inputs to the linear mixture model were linear discriminants derived from these metrics. The metrics were derived to be insensitive to time of year, so as to be globally applicable.

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 original continuous fields products. To address these issues, the original products were made to match with the water fractions of the ISLSCP II land/water mask. Two general cases were addressed: 1) The ISLSCP II mask is water and the continuous fields map is land, 2) The ISLSCP II mask is land and the continuous fields product is water. For 1), the original fractions for each continuous field (e.g. bare, tree, herb) were adjusted using the land fractions of the ISLSCP II mask. For cases in 2), if the original water fraction was less than 100%, the existing continuous fields values were adjusted as in 1). In cases were the original water fraction was 100% in 2), the cell(s) in all continuous fields files was(were) 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 given a value of –999 (No Data).

9.3 Calculations

9.3.1 Special Corrections/Adjustments

See section 9.1. The sum of the percent deciduous and evergreen layers equals the percent tree cover layer as does the sum of the needleleaf and broadleaf layers. For example, if a pixel has 50 percent needleleaf and 50 percent broadleaf trees with an overall tree cover of 40 percent, both the needleleaf and broadleaf layers will equal 20 percent. Also, note that needleleaf deciduous tree cover does not exist in the same pixels as broadleaf evergreen tree cover. To implement this rule, needleleaf deciduous forest is allowed to occur only east of 60 degrees East and north of 40 degrees North (largely Siberia). Outside of this area, all deciduous forest is broadleaf. This rule allows for successfully allocating the tree cover characteristics.

9.4 Graphs and Plots

See DeFries et al. (1999 and 2000).

10. ERRORS

10.1 Sources of Error

The definition of tree cover is any woody vegetation equal to or greater than 5 meters in height. As such, shrub cover and short trees are included in the herbaceous layer. Most errors are due to training data and algorithm limitations. The training data were derived from a diverse set of imagery using ancillary map sources for labeling. These map sources had widely varying land cover definition legends and levels of accuracy. Also, the linear model oversimplifies the complex relationships between the cover types. However, this was largely overcome by adjusting the result with the classification overlays. The data filling procedures described in Section 9.2.3 also introduce some errors.

10.2 Quality Assessment

10.2.1 Data Validation by Source

The data set has not been systematically validated. However, comparisons with higher resolution data, as well as ancillary land cover products are made in DeFries et al. (1999 and 2000). Overall, comparisons with other satellite-derived databases reveal an internally consistent depiction of global tree cover. The only noticeable bias is the loss of heterogeneity in spatially complex areas. In these areas, forest is overestimated due to the favoring of greener, more densely vegetated pixels in the compositing process.

10.2.2 Confidence Level/Accuracy Judgment

Cursory validation indicates that the user should be aware of the following issues:

1) The absence of short woody vegetation in the data set.
2) The high end of tree cover (>=80 percent) is not well delineated due to saturation of the satellite signal and cloud contamination in the tropics.
3) The low end of tree cover (0-20 percent) is inconsistently depicted as the tree canopy represents a small portion of the overall satellite signal in these areas.
4) Interannual variability of herbaceous cover is not captured in a single year depiction.

10.2.3 Measurement Error for Parameters and Variables

Not available at this revision.

10.2.4 Additional Quality Assessment Applied

None.

11. NOTES

11.1 Known Problems with the Data

See section 10.2. Some areas with small values of herbaceous cover are found in central Greenland. Some areas of missing data have been identified from mismatches with the ISLSCP II land/water mask.

11.2 Usage Guidance

See section 10.2. When aggregating the 1km maps to coarser scales, many errors are reduced, such as the limitation of depicting spatial heterogeneity. Users at one-quarter degree and coarse grids should not be concerned with many of the error sources, save for the high-end saturation of tree cover at 80 percent.

Users should be aware that because this is a fractional product, the data are consistent with the land and water fractions of the ISLSCP II land/water mask and NOT the binary land/water mask. All areas that have a non-zero percentage of land within the cell will have a value.

11.3 Other Relevant Information

None.

12. REFERENCES

12.1 Satellite/Instrument/Data Processing Documentation

None.

12.2 Journal Articles and Study Reports

DeFries, R. S. and J. R. G. Townshend, 1994, NDVI-derived land cover classification at global scales. *International Journal of Remote Sensing*, 15:3567-3586.

DeFries, R. S., Hansen, M. C., Townshend, J. R. G. and Sohlberg, R. S., 1998, Global land cover classifications at 8km spatial resolution: the use of training data derived from Landsat imagery in decision tree classifiers, *International Journal of Remote Sensing*, 19, 3141-3168.

DeFries, R. S., Townshend, J. R. G., and Hansen, M. C., 1999, Continuous fields of vegetation characteristics at the global scale at 1km resolution, *Journal of Geophysical Research*, 104, 16 911-16 925.

DeFries, R. S., Hansen, M. C., Townshend, J. R. G., Janetos, A. C., and Loveland, T. R., 2000, A new global 1-km dataset of percentage tree cover derived from remote sensing. *Global Change Biology*, 6, 247-254.

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

Hansen, M. C., DeFries, R.S., Townshend, J.R.G., and Sohlberg, R., 2000, Global land cover classification at 1km spatial resolution using a classification tree approach, *International Journal of Remote Sensing*, 21, 1331-1364.

James, M. E. and S. N. V. Kalluri, 1994. The Pathfinder AVHRR land data set: An improved coarse resolution data set for terrestrial monitoring. *International Journal of Remote Sensing*, 15(17):3347-3363.

Loveland, T. R., Reed, B.C., Brown, J.F., Ohlen, D.O., Zhu, Z., Yang, L., and Merchant, J.W., 2000, Development of a global land cover characteristics database and IGBP DISCover from 1km AVHRR data, *International Journal of Remote Sensing*, 21, 1301-1330.

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 ORNL DAAC. There are no plans to update these data.

14. GLOSSARY OF ACRONYMS

AVHRR	Advanced Very High Resolution Radiometer
CD-ROM	Compact Disk (optical), Read Only Memory
DAAC	Distributed Active Archive Center
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
ISLSCP	International Satellite Land Surface Climatology Project
LAC	Local Area Coverage
MODIS	Moderate Resolution Imaging Spectroradiometer
MSS	Multi Spectral Scanner
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
NESDIS	National Environmental Satellite Data and Information Service (NOAA)
NOAA	National Oceanic and Atmospheric Administration
ORNL	Oak Ridge National Laboratory
UMD	University of Maryland