# The IGBP DISCover Confidence Sites and the System for Terrestrial Ecosystem Parameterization: Tools for Validating Global Land-Cover Data

**Douglas Muchoney, Alan Strahler, John Hodges, and Janet LoCastro**

## Abstract

*The IGBP Validation Confidence Site database provides a set of 379 land-cover maps, each containing an IGBP Core Validation sample. Each map is 448 km² in area and is delineated and labeled by photointerpretation of Landsat or SPOT satellite imagery at a scale of 1:125,000. Within each map, land-cover types and polygons are assigned descriptive labels and parameter codes for vegetation attributes, including life form, cover, height, and phenology for canopy and ground layers. These attributes are a subset of parameters defined by the System for Terrestrial Ecosystem Parameterization (STEP), a site model and database that characterizes land surface and vegetation for use in global algorithm training, testing, and validation of land-cover data. Because the maps are linked to the core samples, they provide a large, consistent dataset that is stratified to represent equally all of the world's major vegetation form classes and land-cover types. The confidence site database has three primary applications: (1) as a set of validation benchmarks for alternate regional or global land-cover classifications emphasizing vegetation attributes, (2) as a secondary information source for studying core sample accuracy issues, and (3) as a source of training and test sites for regional and global supervised classification of coarse-resolution satellite imagery.*

## Introduction

The preparation of the IGBP DISCover global land-cover map and database product (Loveland *et al.*, 1999a, in this issue) has provided the opportunity to develop a sample design for validation of the accuracy of this product and research into related issues concerning the quality of the information it contains (Scepan, 1999, in this issue). The primary strategy for this validation was the acquisition of a global random stratified sample of one-square-kilometer pixels for which the land-cover class label was verified by photographic interpretation of high-resolution Landsat or SPOT digital imagery. Approximately 400 such samples were selected, necessitating the acquisition of nearly as many Thematic Mapper and SPOT images.

In addition to this core sample, the IGBP validation strategy proposed the assessment of the quality of the database at a series of confidence sites, that is, a set of locations at which fine-resolution land-cover information was available (Belward, 1997). In this report, confidence sites were envisioned not as random samples, but rather as sites of opportunity where fine-resolution data were acquired to provide a testing ground for technique development and allow further studies of accuracy and intrinsic land-cover properties.

To help meet this need for confidence sites, the acquisition of core validation data was expanded to include photointerpretation and labeling of a 448-square-kilometer site containing each core validation pixel. These land-cover patches thus provide a dataset of site-specific land-cover information that is inherently stratified to sample the full range of global land covers and environments.

A key feature of the classification of land-cover types within each individual confidence site is that photointerpreters were charged with delineating land-cover types that were natural and well-suited to the individual site. In addition to a descriptive type label devised by the photointerpreter, each type, and in many cases each polygon, was further coded with a set of attributes describing vegetation form, coverage, and phenology by strata. These attributes, or site and vegetation parameters, were based on a subset of parameters used by the System for Terrestrial Ecosystem Parameterization (STEP). STEP is a model and database that has been developed by Boston University to ascribe parameters to global sites based on field and high-resolution remote sensing data for use in training and testing land-cover classification algorithms, and for validating the EOS-MODIS global land-cover map products (Strahler *et al.*, 1999).

The STEP attributes, along with the type description, allow the relabeling of the site within any system of broad land-cover classes, for example, those used by the Simple Biosphere Model SiB, SiB2 (Sellers *et al.*, 1996); the Biosphere-Atmosphere Transfer Scheme, BATS (Dickenson *et al.*, 1986; Dickenson *et al.*, 1993); or the IGBP scheme itself (Belward and Loveland, 1995; Loveland *et al.*, 1999b). By using a subset of the STEP parameters, the confidence site database is actually free of any specific land-cover classification system and may be used for validation of any broad global land-cover classification system.

Confidence site activities also included the validation of a pair of secondary core sample pixels at each confidence site. These secondary pixels were located at opposite and adjacent corners of a 20- by 20-km rectangle, including the core site at one corner. These pixels thus represent a sample of individual pixels centered 19 and 26.9 kilometers away from the initial

Photogrammetric Engineering & Remote Sensing
Vol. 65, No. 9, September 1999, pp. 1061–1067.

0099-1112/99/6509–1061$3.00/0


Department of Geography and Center for Remote Sensing, Boston University, 675 Commonwealth Avenue, Boston, MA 02215-1401 (muchoney@bu.edu).

core sample. When corrected for spatial auto-correlation and conditional probability of selection given the core sample type, these samples can serve to reduce the confidence interval on individual entries in the core site confusion matrix.

## Description of Confidence Site Database

Plate 1 shows the layout of a typical confidence site as it is superimposed on a Landsat Thematic Mapper image for a sample point in Mato Grosso, Brazil. The yellow line bounds the confidence site area. The core sample is located in the upper left corner and is shown by a black square outlining the 1-square-km sample pixel. At the upper right, centered 19 kilometers immediately east of the primary core sample pixel, is the first secondary core sample pixel. The second secondary core sample pixel is located in the lower right hand corner. At each core sample, the confidence area is expanded so that the core sample is centered in a 5- by 5-km region. This extension ensures that each core sample will be placed in its geographical context of land covers.

The core and confidence sites are overlain on a Landsat Thematic Mapper image that has been geolocated using ephemeris information (Husak *et al.*, 1999, in this issue). Where the primary core sample fell close to the margin of a Landsat scene, the L-shaped placement of secondary core samples was rotated to keep the confidence site area within the high-resolution image.

For the delineation and labeling of land-cover classes within the confidence site, each interpreter was presented with a 27- by 29-cm color composite image covering an area of about 35 by 40 km (Plate 1). This image included a digital overlay of the confidence site and core sample locations. Placing the image in a clear plastic sleeve, the interpreters then manually delineated land-cover regions within the confidence site. The photointerpreters were encouraged to extend the delineation boundaries beyond the edges of the confidence seam where it was easy and natural to do so. Each resulting polygon within the confidence site area was then numbered, and its attributes were recorded.

## Confidence Site Attributes

The confidence site attributes are a subset of those of the STEP, which was devised for parameterizing training sites for global land-cover mapping using the MODIS instrument (Strahler *et al.*, 1999). STEP provides for explicit description of the structural, functional, and compositional components of the vegetation and landscape tied to specific sites and plots. Its primary purpose is to provide a comprehensive model of the land-surface cover that can be used to train and test algorithms and to validate land-surface products. STEP has been applied to vegetation classification in Central America (Muchoney *et al.*, 1999), where it is now being used for regional vegetation characterization and monitoring. It has recently been used for supervised

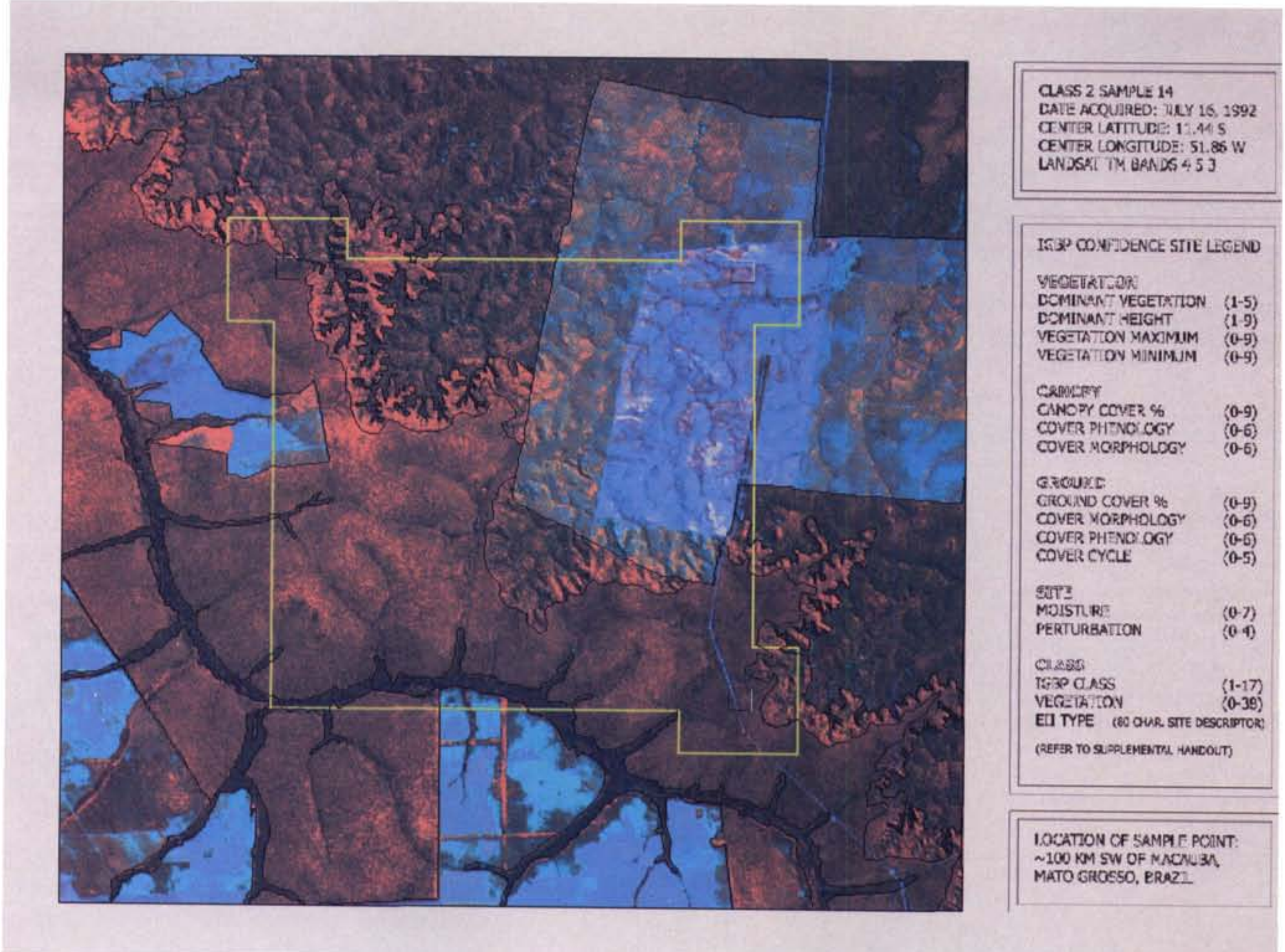


Plate 1. Mato Grosso, Brazil core and confidence sites.

classification algorithm training, testing, and validation of North American land cover (Friedl *et al.*, 1999). The current STEP manual and database template can be accessed at **http://crs-www.bu.edu/Faculty/Muchoney/step.html**.

STEP is based on vegetation physiognomy, morphology, phenology, periodicity, and pattern derived from remote sensing, ancillary, and field data. STEP accommodates five strata for field data collection but collapses to two strata for the global site database. Parameters are intended to be pure for each form and level of description. For example, vegetation cover fraction is explicit and not inferred from a generalized physiognomic class or other thematic classes.

From a classification perspective, important parameters can be merged to yield classification units that take advantage of the specific biogeophysical gradients used to define a class. For example, physiognomic and hydrologic attributes can be combined to discriminate herbaceous and forested wetlands. Mapping and description are based on homogeneous patches and stands at the global 1-km level. Horizontal and vertical structure, leaf morphology, and physiognomy need to be characterized in relation to the phenology of each stratum. If these dimensions are not modeled in tandem, it is difficult to benefit from the temporal resolution of sensors such as AVHRR (Advanced Very High Resolution Radiometer) and MODIS (Moderate Resolution Imaging Spectroradiometer) to derive parameters such as leaf area index (LAI) and net primary productivity (NPP).

Because of the need for classification systems and models that allow for better inference and estimation of a number of classification systems and parameters, STEP was developed as an approach to map multiple classification systems directly and to develop global surface parameters directly from site parameters. Rather than only create another classification system, we have restructured the classification and parameter elements into a lowest common denominator set that is classification-free yet allows for approximation and cross-walking of multiple classifications at their fundamental levels.

Appendix A describes the land-cover attributes and codes recorded for each polygon within a confidence site. We present a description of the attributes that were used to characterize the confidence sites, keeping in mind that these are a subset of the full range of STEP parameters. Table 1 provides the site parameters for the Mato Grosso example.

#### Vegetation Parameters

Vegetation structure and geometry parameters include horizontal and vertical structure. In the case of vegetation index (VI), there is a need to describe vegetation cover fraction to understand the contribution of the bare ground fraction, which is disproportionately strong due to the nonlinearity of the VI/LAI relationship, with as little as 25 percent bare ground cover fraction significantly effecting response (Sellers *et al.*, 1986). Because vegetation type influences surface fluxes as a function of leaf area, leaf size, and canopy height and spatial arrangement, the best strategy for specifying land surface parameters is by modeling the vegetation (Sellers *et al.*, 1986; Bonan, 1996).

For field plot data, STEP accommodates up to five vegetative strata. For global remote sensing-derived plots, we have generalized vertical structure into ground (<2m) and aboveground (>2m) components because scrub/shrub canopies can have similar canopy structure, leaf physiology, and morphology. The canopy stratum is assumed to be perennial and woody. The ground stratum may include woody, non-woody, herbaceous, graminoid and non-vascular elements. It includes dwarf shrub communities and krummholz or creeping phanerophytic vegetation. This two-strata generalization of canopy vertical structure is consistent with models such as SiB (Sellers *et al.*, 1986). When vertical structure can be directly observed or inferred, this information is tracked by physiognomic class. The STEP ground layer includes bare ground elements, including rock, soil, water, vegetation, wood, and other organic matter. Canopy height can be inferred from vegetation physiognomy but is also included as an additional parameter. Canopy depth is important to describing canopy attenuation of PAR (Dickinson, 1995) and requires estimation of canopy top height and bottom height.

Horizontal structure parameters include estimated cover fraction at each vertical stratum for each vegetative and non-vegetative element, distribution, and a general description of physiognomy. When coupled with vertical structure, physiognomy, and phenology, STEP can be used to model both leaf-on and leaf-off or dry/wet season reflectance and provide a model of proportional mixing. The horizontal structure is defined as the proportion of surface area of each phenological, morphological, and physiognomic class of each stratum.

Leaf phenology is the normal status of whether a class of plants (vegetation type) drops all or a portion of its leaves periodically and therefore must regrow all or a portion of its photosynthetic mechanism from year to year. Leaf phenology is important from a model perspective for characterizing productivity, LAI, and canopy radiative properties. From a remote sensing perspective, it is important for characterizing the temporal dynamics of the land surface and vegetation reflectance (Loveland *et al.*, 1995; Reed *et al.*, 1994). By formalizing the phenological relationship, we can determine the nature of the scene for both leaf-on and leaf-off conditions, wet season/dry season, or winter snow.

TABLE 1. CONFIDENCE SITE PARAMETERS, MATO GROSSO SITE

| | IGBP Global Validation Workshop Confidence Site Mapping Worksheet | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Vegetation | | | | Canopy | | | Ground | | | | Site | | Class | |
| Polygon # | Dominant | Dominant Height | Maximum % Cover | Minimum % Cover | % Cover | Cover Phenology | Cover Morphology | Vegetation Cover | Cover Morphology | Cover Phenology | Cover Cycle | Moisture | Perturbation | IGBP LC Class | Vegetation |
| 1 | 1 | 7 | 9 | 7 | 9 | 1 | 2 | — | — | — | — | 4 | 1 | 2 | 3 |
| | EII Type: Primary Forest | | | | | | | | | | | | | | |
| 2 | 1 | 6 | 8 | 6 | 8 | 2 | 2 | — | 4 | 2 | 4 | 4 | 1 | 2 | 3 |
| | EII Type: Escarpment Forest Cerrado | | | | | | | | | | | | | | |
| 3 | 2 | 3 | 7 | 6 | 3 | 2 | 2 | 6 | 4 | 2 | 4 | 4 | 1 | 7 | 20 |
| | EII Type: Cerrado | | | | | | | | | | | | | | |
| 4 | 4 | 2 | 8 | 6 | 0 | — | — | 8 | 4 | 2 | 4 | 4 | 3 | 10 | 27 |
| | EII Type: Pastoral Cerrado-recent | | | | | | | | | | | | | | |
| 5 | 4 | 2 | 8 | 6 | 0 | — | — | 8 | 4 | 2 | 4 | 4 | 3 | 10 | 27 |
| | EII Type: Primary Forest Pastoral | | | | | | | | | | | | | | |
| 6 | 1 | 7 | 9 | 8 | 8 | 1 | 2 | — | — | — | — | 3 | 1 | 2 | 3 |
| | EII Type: Gallery Forest | | | | | | | | | | | | | | |

Phenology applies to the overall dominants of each of the ground and canopy strata. In the case of a mixed deciduous broadleaf canopy (phenology equals deciduous) and evergreen needleleaf forest canopy (phenology equals evergreen), the canopy phenological type would be mixed. Drought deciduous species are problematic as they are subject to leaf fall due to water balance rather than temperature. The mixed class is distinct from semi-evergreen and semi-deciduous in that it is intended to represent co-occurrence of evergreen, semi-evergreen, semi-deciduous, and deciduous vegetation. Canopy phenology and ground-cover phenology are tracked independently, with phenological categories provided in Appendix A.

Vegetation periodicity or life-cycle relates to energy expenditure by plants. While annuals must produce root and above ground vegetation annually, and storage is in seed, perennials can store energy above and below ground. We use life cycle rather than periodicity to avoid confusion with phenological phenomena such as flowering, and leaf generation and senescence. Life cycle is tracked for the ground stratum while vegetation in the canopy stratum implies that it is perennial. Life cycle can be used in tandem with phenology, for example, to characterize perennial grasslands that die back seasonally.

The STEP vegetation structure, morphology, and phenology parameters relate to the dominant lifeform, canopy, and ground strata.

*Dominant Vegetation*

The confidence site parameters for dominant vegetation are dominant type and height. The dominant life or growth form or life form is the form of plant that dominates the vegetation or ecosystem type. The dominant life-form classes are defined by their stature and whether the are ligneous (woody) vascular, nonligneous (nonwoody) vascular, or nonvascular plants (Appendix A). The percent vegetation cover of dominant life form is the estimated surface-cover fraction achieved by viewing a horizontal cross-section of the dominant stratum. While STEP describes dominant life-form taxonomy, phenology, and morphology, these attributes were not defined for the confidence sites.

The height of the dominant life or growth form is important for describing structure and physiognomy which are related to biomass and productivity and influence parameters such as surface roughness. The height categories are average values of the height of the dominant life form (Appendix A).

*Canopy*

The confidence site canopy parameters are canopy cover fraction, canopy morphology, and canopy phenology. Percent vegetation cover is the estimated surface cover fraction at each stratum obtained by viewing a horizontal cross-section of the stratum. The cover fraction categories are ordinal, ranging from 0 to 9 and representing 10 percent cover classes of 0 to 100 percent. These values can be generalized to five classes: (1) 0 to 10, (2) 10 to 40, (3) 40 to 60, (4) 60 to 90, and (5) 90 to 100 percent to coincide with other classification systems. In this physiognomic classification system, cover class 5 denotes a closed canopy. In the case of tree-cover fraction, class 5 represents closed forest, class 4 is forest, class 3 is woodland, class 2 is woodland savanna, and class 1 is treeless.

*Ground Cover*

STEP ground-cover attributes comprises an estimate of the ground-cover fractions of vegetation, soil, wood, rock, nonwoody organic matter, water, and snow/ice. Vegetation ground-cover fraction, morphology, and cycle (periodicity) were estimated for the confidence sites. Further estimates of maximum and minimum vegetation cover, including both canopy and ground strata, were also made.

### Site Parameters

Physical site and landscape parameters are disturbance and moisture regime. The level and nature of vegetation disturbance greatly influences our understanding of land-cover and land-use processes, and provides a further indication of floristics, physiology, function, and structure. Data on the nature of human and natural disturbance regimes and their effects on ecosystems, though largely undeveloped, are important to ecosystem models (Schimel *et al.*, 1997). While successional status is difficult to ascertain, the level of modification can be inferred and provides important indications of biological integrity and functioning. Disturbance has significant impacts on biodiversity as well as on the radiative properties of vegetation. For example, selective forest cutting can open up a canopy, which increases light trapping and decreases albedo (Dickinson, 1983). Alternatively, clearcutting can increase albedo, indicating the need to characterize the type of disturbance.

Because the effects of human disturbance are so important to biological conservation and understanding land-use and land-cover change processes, it is important to estimate the state of disturbance for a site. In many cases, the magnitude of human intervention is evident, while in others it is more inconspicuous. Natural systems are defined as those with minimal human impact to vegetation structure and floristics. Modified types are semi-natural lands where the natural vegetation structure and species are largely intact though the influence of human management, including forest high-grading and grazing by domestic livestock, does occur. We do not explicitly include a pastoral class which can be inferred from the perturbation categories. Agricultural systems include continuous, systematic conversion or use of the site for row and cereal crop, permanent intensive pasture, and industrial forest manage-ment.

Moisture regime is intended to be used to describe the prevailing and seasonal water balance of sites, especially to describe wetlands. The nature of hydrologic processes is critical to defining vegetation type and condition. Moisture regime is used to differentiate wetland vegetation and to characterize seasonal vegetation classes such as drought-deciduous vegetation. The seasonal condition of drought-deciduous species, especially in areas of the tropics that are influenced by seasonally variable phenomena such as El Niño, are critical indicators of impacts of global change. The site moisture regime categories are defined in Appendix A.

### *A Priori* Classifications

Physiognomic classifications (Kuchler, 1949) include strict physiognomic criteria (appearance) as well as modified systems based on physiognomic-floristic and physiognomic-structural criteria (Schimper, 1898; Schimper and von Faber, 1935; UNESCO, 1973 Beard, 1978;). Physiognomy is generally described for dominant life or growth forms (Penfound, 1969). Physiognomic systems have the advantage of being relatively intuitive, simple, and globally applicable. Physiognomic classifications based on life or growth form can be considered as being species independent but indicative of environmental effects which may influence the functional attributes of vegetation. In this case, the classification is more correctly physiognomic-ecological or physiognomic-climatic in nature (Schultz, 1995). Physiognomic classes or formations describe the dominant life forms and relate to standing above-ground woody and vegetative biomass.

Structural systems are based on the spatial arrangement of vegetation components (Fosberg, 1967). Structural systems convey stand-level data which are independent of community taxonomy and which can include life form, size defined as vertical extension, function, leaf morphology and texture, and horizontal distribution (Dansereau, 1951). Although the STEP attributes allow the development of physiognomic and structural criteria independent of a generalization of these criteria, we have defined a set of 14 physiognomic-structural formations based on dominant life form, and vertical and horizontal structure and distribution.

While STEP is classification-free in that it can be used to approximate many types of classification systems without requiring that they be specified in the database, it includes a descriptive field and five other classifications for which the confidence polygons were labeled. For the confidence sites, the IGBP class, a vegetation class, and a user-defind class or description were used.

### *IGBP*

The IGBP system label is included because of its wide use, and the IGBP label is assigned to each site based on its attributes instead of using the global IGBP classification of Belward and Loveland (1995).

### *Vegetation Type*

In addition to the IGBP type, we further see the need for a more detailed vegetation classification system, as well as an ecosystem classification, i.e., the Boston University STEP-Vegetation system. This system is in the class of physiognomical/structural classification systems but is more detailed than other classifications of this type.

### *Common Name/Description*

The common name (EIITYPE) is used to describe the specific site ecosystem or land-cover type without regard to any classificatory structure. It may be a local or common name (e.g., Cockpit Country Jamaica Mesic Limestone Forest), or just an *ad hoc* description (e.g., secondary limestone forest, small farmsteads). This variable is included as an aid to developing site parameters and for understanding relationships between modeled and site parameters.

### Quality Assurance Data

The photointerpreters were also requested to provide a general level of confidence in their assignment of parameters to each polygon. Assignment of land-cover type labels was facilitated by the use of ancillary information, including maps, atlases, and vegetation descriptions from the literature. These materials were supplied both by the photointerpreters themselves and by the library at the EROS data center (Kelley *et al.*, 1999, in this issue; Scepan, 1999, in this issue). Secondary core samples were validated by the photointerpreters against the IGBP land-cover type descriptions in a procedure that was identical to that used for the core samples themselves.

In total, the confidence sites provide 15 dominant vegetation, canopy, ground-cover, site, and class parameters for site polygons, with confidence estimates for each parameter. The completion and general confidence estimate for the confidence sites completed at the September 1998 validation workshop are summarized in Table 2. Note that Regions 1 and 2 were largely incomplete due to problems getting the interpreters to the workshop due to an airline strike. These sites are being completed by Boston University.

TABLE 2. CONFIDENCE SITE STATUS

| | | | | general confidence estimate | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Region | Total Sites | Number Complete | Percent Complete | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 1 | 44 | 7 | 15.9 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 67 | 22 | 32.8 | 0 | 0 | 1 | 4 | 3 | 1 | 6 | 6 | 1 |
| 3 | 14 | 14 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 4 | 6 |
| 4 | 34 | 34 | 100 | 0 | 0 | 1 | 1 | 0 | 1 | 6 | 8 | 17 |
| 5 | 6 | 6 | 100 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 2 | 1 |
| 6 | 45 | 45 | 100 | 3 | 6 | 5 | 8 | 1 | 6 | 4 | 10 | 2 |
| 7 | 28 | 28 | 100 | 0 | 0 | 1 | 1 | 3 | 6 | 9 | 8 | 0 |
| 8 | 52 | 23 | 44.2 | 2 | 3 | 3 | 2 | 2 | 3 | 4 | 3 | 1 |
| 9 | 15 | 15 | 100 | 0 | 0 | 0 | 0 | 0 | 2 | 4 | 6 | 3 |
| 10 | 21 | 21 | 100 | 1 | 2 | 5 | 2 | 5 | 0 | 4 | 2 | 0 |
| 11 | 60 | 60 | 100 | 20 | 15 | 9 | 3 | 2 | 2 | 4 | 3 | 2 |
| 12 | 13 | 11 | 84.6 | 0 | 0 | 6 | 5 | 0 | 0 | 0 | 0 | 0 |
| 13 | 16 | 16 | 100 | 2 | 0 | 0 | 0 | 1 | 3 | 5 | 5 | 0 |
| Total | 415 | 302 | 72.8 | 35 | 26 | 31 | 26 | 18 | 25 | 51 | 57 | 33 |

## Potential Applications of Confidence Sites

The IGBP confidence sites provide a rich resource for global land-cover classification and validation activities using remotely sensed data. A primary advantage of this dataset is that it is consistent. Produced in a two-week period by an international group of photointerpretation experts, the dataset is the product of a consistent methodology applied uniformly to the suite of about 400 image segments. A second advantage is that the dataset is stratified by IGBP land-cover type and so provides a representative sampling of all the world's primary land-cover types.

We anticipate that the confidence site database will have three primary applications. First, it will be used in accuracy assessment of other global land-cover maps beyond that of the IGBP DISC over data product. Given a set of broad vegetation classes that are defined using parameters largely similar to those coded for confidence site polygons, it should be possible to assign a label to each confidence site polygon in such an alternative classification and thereby validate a classification label assignment of a sample, or of a wall-to-wall mapping, within each confidence site. In this way, the confidence site database will serve as a benchmark against which future global digital classifications may be measured.

The second important use of the confidence site database will be as a secondary information source for resolving core sample accuracy issues. For example, accuracy statistics for core samples include potential geolocation errors. Although such errors are no different in principle from thematic classification errors, the question of whether significant numbers of errors are due to geolocation accuracy rather than incorrect thematic classification is a research topic of considerable interest. Geolocation error may be quantified by overlaying AVHRR data in image format on the extended confidence site regions. If geolocation errors are present, their impact on thematic classification can be assessed. Note that geolocation error may be either in the AVHRR dataset or in the location of the core sample pixel and associated confidence site (Husak *et al.*, 1999, in this issue).

The third potential application for the confidence site database is as a source of training sites for global digital classifications of remotely sensed data. Each confidence site has the potential to contribute a number of training sites for use in a global or regional supervised classification. As researchers charged with producing a global land cover map using data from the MODIS instrument (Strahler *et al.*, 1999; Justice *et al.*, 1998), the authors expect to draw freely from this valuable resource.

## Dataset Preparation and Release

The authors are currently undertaking the task of preparing the confidence site database for wider use, which includes updating and editing confidence site overlays and parameter attributes prepared by the photointerpreters, as well as making the data available in a consistent format at a convenient and easy-to-use web site location **(http://geography.bu.edu/landcover/IGBP/index.htm)**. The database will be released in two versions. In version one, each confidence site will be presented as a scanned image of the hardcopy scene and an overlay provided by the photointerpreter. Parameter coding sheets will be transcribed and made available digitally. In version two, each confidence site region will be provided as a Landsat or SPOT digital image subscene with digital overlays locating the confidence site pixel, core site pixels, and delineated polygons in GIS format. In version two, geolocation will also be validated and image corner coordinates will be updated from control points where geolocation is problematic. Version one data have been completed and are available on the DISCover web site (Scepan, 1999, in this issue). The present timetable calls for version two data to be completed and released by December 1999.

## Conclusion

The IGBP confidence site database will provide a very useful and valuable resource for the global land-cover mapping community. Because it provides a large number of samples that are consistently derived and formatted, the database can serve to benchmark the accuracy of continental- to global-scale classifications of coarse-resolution remotely sensed data. In addition, it will provide a foundation for obtaining answers to a number of important research questions regarding the statistical accuracy and validity of the IGBP core validation sample. Moreover, it will also provide a source of global training site data that will be of great interest to the community of global land-cover mapping researchers.

## References


Beard, J.S., 1978. The physiognomic approach, *Classification of Plant Communities* (R.H. Whittaker, editor), Junk Publishers, The Hague, pp. 33–64.

Belward, Alan S., (editor), 1997. *The IGBP-DIS Global 1-Km Land Cover Data Set "DISCover" Proposal and Implementation Plans.*, Report of the Land Cover Working Group of IGBP-DIS, IGBP-DIS Working Paper No. 13, October 1996, reprinted October, 1997.

Belward, A., and T.R. Loveland, 1995. The IGBP 1-km land cover project, *Proc. of the 21st Annual Conf. of the Remote Sensing Society*, Southampton, UK, pp. 1099–1106.

Bonan, G.B., 1996. *A Land Surface Model (LSM Version 1.0) for Ecological, Hydrological, and Atmospheric Studies: Technical Description and User's Guide*, NCAR Technical Note 417+STR, National Center for Atmospheric Research, Boulder, Colorado, 150 pp.

Dansereau, P., 1951. Description and recording of vegetation, *Ecology*, 32(2):172–229.

Dickinson, R.E., 1983. Land surface processes and climate-surface albedos and energy balance, *Advances in Geophysics*, 25:305–353.

———, 1995. Land processes in climate models, *Remote Sensing of Environment*, 51:27–38.

Dickinson, R.E., A. Henderson-Sellers, P.J. Kennedy, and M.F. Wilson, 1986. *Biosphere Atmosphere Transfer Scheme (BATS) for the NCAR Community Climate Model*, NCAR Technical Note, NCAR, 69 pp.

Dickinson, R.E., A. Henderson-Sellers, and P.J. Kennedy, 1993. *Biosphere Atmosphere Transfer Scheme (BATS) Version le as Coupled to the NCAR Community Climate Model*, NCAR Technical Note, NCAR, Boulder Colorado 72 p.

Fosberg, F.R., 1967. *A Classification of Vegetation for General Purposes*, IBP Handbook No. 4, [place of publication], 120 pp.

Friedl, M., W. Woodcock, S. Gopal, D. Muchoney, A. Strahler, and C. Barker-Schaaf, 1999. A note on procedures for accuracy assessment in land cover maps derived from AVHRR data, *Remote Sensing Letters, International Journal of Remote Sensing*, (in press).

Husak, G.J., B.C. Hadley, and K.C. McGwire, 1999. Registration accuracy of high-resolution data used in the IGBP validation, *Photogrammetric Engineering & Remote Sensing*, 65(9):1033–1039.

Justice, C., E. Vermote, J. Townshend, R. Defries, D. Roy, D.K. Hall, V.V. Salomonson, J.L. Privette, G. Riggs, A. Strahler, W. Lucht, R. Myneni, Y. Knyazikhin, S.W. Running, R.R. Nemani, Z. Wan, A.R. Huete, W. van Leeuwen, R.E. Wolfe, L. Giglio, J.-P. Muller, P. Lewis, and M.J. Barnsley, 1998. The Moderate Resolution Imaging Spectroradiometer (MODIS): Land remote sensing for global change research, *IEEE Transactions on Geoscience and Remote Sensing*, 36:1228–1249.

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.

Kuchler, A.W., 1949. A physiognomic classification of vegetation, *Annals of the Assoc. of Am. Geographers*, 39:201–210.

Loveland, T.R., J.W. Merchant, J.F.Brown, D.O.Ohlen, B.C.Reed, P.Olson, and J. Hutchinson, 1995. Seasonal land-cover of the United States, *Annals Assoc. of Am. Geographers*, 85(2):339–355.

Loveland, T.R., Z. Zhu, D.O. Ohlen, J.F. Brown, B.C. Reed, and L. Yang, 1999a. An Analysis of the IGBP global land-cover characterization process, *Photogrammetric Engineering & Remote Sensing*, 65(9):1021–1032.

Loveland, T.R., B.C. Reed, J.F. Brown, D.O. Ohlen, J. Zhu, L. Yang, and J.W. Merchant, 1999b. Development of a Global Land Cover Characteristics Database and IGBP DISCover from 1-km AVHRR data, *International Journal of Remote Sensing*, (in press).

Muchoney, D.M., J. Borak, H. Chi, M. Friedl, J. Hodges, N. Morrow, and A. Strahler, 1999. Application of the MODIS global supervised classification model to vegetation and land cover mapping of Central America, *International Journal of Remote Sensing*, (in press).

Penfound, W.T., 1969. A physiognomic classification of vegetation in conterminous United States, *Bot. Rev.*, 33:289–326.

Reed, B.D., J.F. Brown, D. VanderZee, T.R. Loveland, J.W. Merchant, and D.O. Ohlen, 1994. Measuring phenological variability from satellite imagery, *J. Vegetation Science*, 5:703–714.

Schimel, D.S., VEMAP Participants, and B.H. Braswell, 1997. Continental scale variability in ecosystem processes: Models, data, and the role of disturbance, *Ecological Monographs*, 67:251–271.

Schimper, A.F.W., 1898. *Pflanzengeographie auf physiogischer*, Grundlage Gustav Fischer, Jena.

Schimper, A.F.W., and F.C. von Faber, 1935. *Pflanzengeographie auf physiogischer*, Grundlage Gustav Fischer, Jena.

Schultz, J., 1995. *The Ecozones of the World*, Springer-Verlag, Berlin, 449 p.

Sellers, P.J., Y. Mintz, Y.C. Sud, and A. Dalcher, 1986. A simple biosphere model (SiB) for use within general circulation models, *J. Atmos. Sci.*, 43:503–531.

Sellers, P.J., D.A. Randall, and G.J. Collatz, 1996. A revised land surface parameterization (SiB2) for atmospheric GCMs. Part 1: Model formulation, *Journal of Climate*, 9:676–705.

Strahler, A., D. Muchoney, J. Borak, G. Feng, M. Friedl, S. Gopal, J. Hodges, E. Lambin, D. McIver, A. Moody, C. Schaaf, and Curtis Woodcock, 1999. *MODIS Land Cover Product Algorithm Theoretical Basis Document (ATBD) Version 5.0*, Boston University, Boston, 89 p.

UNESCO, 1973. *International Classification and Mapping of Vegetation, Series 6*, Ecology and Conservation, UNESCO, Paris, 93 p.


## Appendix A

IGBP Confidence Site Mapping Parameters

**VEGETATION**

*DOMINANT VEGETATION:* Dominant life form or growth form
- 1 tree: woody with height >5 meters
- 2 dwarf tree and shrub: woody with height 2–5 meters
- 3 dwarf shrub: woody with height <2 meters
- 4 herbaceous: nonwoody, height <2 meters
- 5 nonvascular plants: nonwoody, height <0.1 meter

*DOMINANT HEIGHT:* Height of dominant life form
- 1 <0.1 meter
- 2 0.1–2 meters
- 3 2–5 meters
- 4 5–10 meters
- 5 10–20 meters
- 6 20–30 meters
- 7 30–40 meters
- 8 40–50 meters

9 >50 meters

*VEGETATION MAXIMUM:* Percent of total vegetation cover at normal maximum of annual development (PERCENT COVER CLASS 0–9).

0 0–10 percent cover
1 10–20 percent cover
2 20–30 percent cover
3 30–40 percent cover
4 40–50 percent cover
5 50–60 percent cover
6 60–70 percent cover
7 70–80 percent cover
8 80–90 percent cover
9 90–100 percent cover

*VEGETATION MINIMUM:* Percent of total vegetation cover when normal minimum (PERCENT COVER CLASS 0–9, as above)

**CANOPY**

*CANOPY COVER PERCENT:* Percent cover of canopy layer (0–9, as above)

*CANOPY PHENOLOGY:* Canopy stratum phenology

0 unknown phenology
1 evergreen: <20% deciduous
2 semievergreen: 20–40% evergreen
3 semideciduous: 20–40% deciduous
4 deciduous: <20% evergreen
5 mixed: 40–60% evergreen, 40–60% deciduous
6 no phenology (non-vegetative)

*CANOPY MORPHOLOGY:* Canopy stratum leaf morphology

0 unknown
1 none (not applicable, non-vegetated)
2 broadleaf
3 needleleaf
4 herbaceous
5 non-vascular
6 mixed

**GROUND**

*GROUND COVER VEGETATION:* Ground cover (PERCENT COVER CLASS 0–9, as above)

*GROUND COVER MORPHOLOGY:* Ground stratum leaf morphology

0 unknown
1 none (not applicable, non-vegetated)
2 broadleaf
3 needleleaf
4 graminoid
5 non-vascular
6 mixed morphology

*GROUND COVER PHENOLOGY:* Ground stratum leaf phenology, coded as per canopy layer above.

*GROUND COVER CYCLE:* Life cycle or vegetation periodicity

0 unknown
1 no periodicity
2 ephemeral (1–4 month life span)
3 annual
4 perennial
5 mixed periodicity

**SITE**

*MOISTURE:* Moisture Regime

0 unknown
1 permanently inundated
2 periodically inundated
3 hydric (wet)
4 mesic (moist)
5 xeric (dry)
6 hyper-xeric (desert)
7 variable
8 irrigated

*PERTURBATION:* Disturbance characteristics

0 unknown
1 natural
2 modified natural
3 agricultural systems
4 urban/industrial

**CLASS LABELS**

*IGBP:* 17 classes

*VEGETATION CLASS:* 38 classes

*EII TYPE:* 80-character site descriptor