

A comparison of the IGBP DISCover and University of Maryland 1 km global land cover products

M. C. HANSEN†

Department of Geography, University of Maryland, College Park, MD 20742, USA

and B. REED

Raytheon ITSS, EROS Data Center, Sioux Falls, SD 57105, USA

Abstract. Two global 1 km land cover data sets derived from 1992–1993 Advanced Very High Resolution Radiometer (AVHRR) data are currently available, the International Geosphere–Biosphere Programme Data and Information System (IGBP-DIS) DISCover and the University of Maryland (UMd) 1 km land cover maps. This paper makes a preliminary comparison of the methodologies and results of the two products. The DISCover methodology employed an unsupervised clustering classification scheme on a per-continent basis using 12 monthly maximum NDVI composites as inputs. The UMd approach employed a supervised classification tree method in which temporal metrics derived from all AVHRR bands and the NDVI were used to predict class membership across the entire globe. The DISCover map uses the IGBP classification scheme, while the UMd map employs a modified IGBP scheme minus the classes of permanent wetlands, cropland/natural vegetation mosaic and ice and snow. Global area totals of aggregated vegetation types are very similar and have a per-pixel agreement of 74%. For tall versus short/no vegetation, the per-pixel agreement is 84%. For broad vegetation types, core areas map similarly, while transition zones around core areas differ significantly. This results in high regional variability between the maps. Individual class agreement between the two 1 km maps is 49%. Comparison of the maps at a nominal 0.5° resolution with two global ground-based maps shows an improvement of thematic concurrency of 46% when viewing average class agreement. The absence of the cropland mosaic class creates a difficulty in comparing the maps, due to its significant extent in the DISCover map. The DISCover map, in general, has more forest, while the UMd map has considerably more area in the intermediate tree cover classes of woody savanna/woodland and savanna/wooded grassland.

1. Introduction

There are currently two global 1 km resolution land cover products available, both derived from data from the Advanced Very High Resolution Radiometer (AVHRR). The first was produced by the U.S. Geological Survey for the International Geosphere–Biosphere Programme (IGBP) and the second by the University of Maryland (UMd). The purpose of this paper is to compare the characteristics of these maps in order to clarify the similarities and differences in the development of each product. A preliminary numerical comparison is also included to illuminate areas of agreement and disagreement.

The two global land cover data sets were created for the same fundamental

purpose of providing improved global land cover information for environmental modellers. The IGBP product, DISCover, was designed to meet the various global land cover needs of IGBP core science projects (IGBP 1992, Rasool 1992). The IGBP-DIS Land Cover Working Group (LCWG) developed a programme to create a global land cover product based on 1 km AVHRR data which culminated in the DISCover land cover product (Loveland *et al.* 2000).

At the University of Maryland, global land cover maps have been produced for the modelling community and, as finer resolution global data sets have become available, for researchers working on a variety of applications requiring land cover information. Recent research has included the generation of one degree (DeFries and Townshend 1994), 8 km (DeFries *et al.* 1998) and 1 kilometer (Hansen *et al.* 2000) global land cover maps.

As explained by Merchant *et al.* (1993), the evaluation of large-area land cover products is very difficult, primarily because of a lack of corroborating evidence and the relatively high cost of conducting a statistically meaningful validation. Existing regional land cover data that may be available for comparison are often of undocumented accuracy and have been developed with non-standardized classification legends and unknown methodologies. Using such information often only compounds the problems in an evaluation of global products. *In lieu* of a rigorous statistical validation, a comparative overview of the methodologies along with areal and per-pixel comparisons is offered here to help users understand the differences between the products and allow them to make better informed decisions on how to use these data sets.

2. Methodological similarities and differences

Table 1 summarizes the key similarities and differences between the IGBP DISCover and UMD products. Both use data from the National Oceanic and Atmospheric Administration (NOAA) AVHRR satellite sensor. The data derived from the AVHRR that were used in the two classification sequences were collected based on monthly maximum Normalized Difference Vegetation Index (NDVI) composites collected from April 1992 to March 1993 inclusive. The DISCover project used the 12 monthly maximum NDVI data while the UMD system used all five AVHRR channels as well as the NDVI in deriving 41 multi-temporal metrics from the 12 monthly composites.

Table 1. Similarities and differences of the IGBP-DISCover and University of Maryland global land cover products.

Product characteristics	IGBP DISCover	University of Maryland
Sensor	AVHRR	AVHRR
Time of data collection	April 1992–March 1993	April 1992–March 1993
Input data	12 monthly NDVI composites	41 metrics derived from NDVI and bands 1–5
Classification technique	Unsupervised clustering	Supervised classification tree
Processing sequence	Continent-by-continent	Global
Classification scheme	IGBP (17 classes)	Simplified IGBP (14 classes)
Refinement/update schedule	Annual	Currently being updated
Validation	September 1998	Evaluated using other digital data sets

The IGBP-DISCover product was created using the 12 monthly maximum NDVI values, representing the annual phenology of vegetation, as inputs into an unsupervised clustering program. The clusters resulting from the algorithm were then labelled and refined at the continental scale according to available ancillary digital and map-based information. Over 250 maps and atlases of ecoregions, soils, vegetation, land use and land cover were used in the interpretation phase of the study and served as reference data to guide class labelling. This approach provided a highly flexible methodology in creating a product which best reproduced the thematic information in the ancillary maps while adding the spatial detail inherent in the remotely sensed data. For more information on the cluster refinement and labelling techniques see the work of Loveland *et al.* (2000).

The DISCover classes are of the IGBP classification scheme. A tabular comparison of the DISCover and UMD class lists can be found in Hansen *et al.* (2000). The processing sequence for the IGBP-DISCover product was continent-by-continent as the data became available through the IGBP-DIS global 1 km project (Eidenshink and Faudeen 1994). The DISCover product was released via the World Wide Web (<http://edcwww.cr.usgs.gov/landdaac/glcc/glcc-na.html>) free of charge to all potential users in July 1997 and is scheduled to be improved annually based on peer-review, user feedback and results from the validation study.

The UMD product used a supervised classification tree based on 41 temporal metrics calculated to represent the phenology of global vegetation. Training sites were derived from over 150 interpreted Landsat scenes distributed throughout the world. A complete list of the Landsat training scenes, along with ancillary data used to interpret them, can be found at the UMD website (www.geog.umd.edu/landcover/global-cover.html).

The UMD classes largely conform to the IGBP scheme. However, permanent wetlands, cropland/natural vegetation mosaic and ice and snow IGBP classes are not included in the UMD product. Signatures derived by the classification tree algorithm were extrapolated worldwide. Most signatures act globally, while others represent unique regional characterizations of a single class or subclass. The UMD product is available at the previously mentioned website, and an improved future version will be generated based on evaluations using ancillary data and user feedback.

The IGBP LCWG established a validation team who developed a strategy and methodology for validating the DISCover land cover product (Belward 1996). The strategy was based on a stratified random sample of the DISCover land cover classes. The random samples were taken from higher resolution satellite imagery (mainly Landsat and SPOT (System Probatoire pour l'Observation de la Terre)). There is no formal validation programme of comparable statistical rigour established for the UMD product due to the considerable cost involved. As an alternative, comparisons with other regional digital data sets are being used to evaluate the map. The DISCover validation sites will also be used in this fashion. While these sites will not provide accuracy statements with known confidence levels, (due to the IGBP DISCover based sampling strategy) they will certainly provide valuable information.

3. Areal and per-pixel comparisons

Figure 1 shows the areal totals for classes as aggregated into physiognomically similar groupings. The totals are quite similar, the only exception being the apparent disagreement in grass/shrub cover totals. This difference can be explained by the absence of the agriculture mosaic class in the UMD classification. However, when

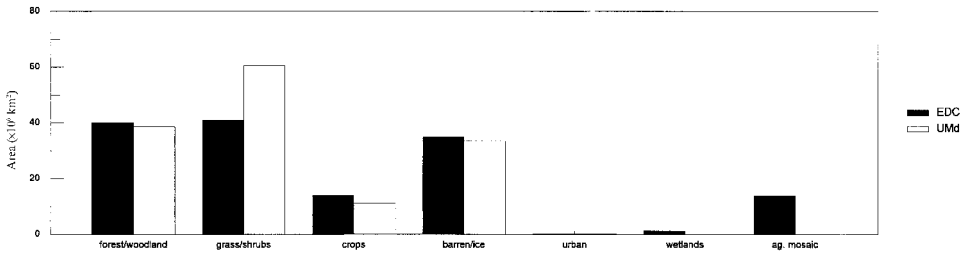


Figure 1. Global area totals for aggregated classes of the EDC DISCover and UMD 1 km maps. Forest/woodland represents all forest classes plus the woody savanna class for the DISCover map and all forest classes plus the woodland class for the UMD map. Grass and shrubs represents both shrub classes, the grassland class and the savanna class for the DISCover map and both shrub classes, the grassland class and the wooded grassland class for the UMD map. Barren/ice is the combined barren class and permanent snow or ice class for the DISCover map and the bare ground class for the UMD map. Crops, urban and wetlands represent individual classes from the respective maps.

comparing the per-pixel accuracies for these groupings, it is clear that the internal arrangement of the classes as represented on the globe is quite different between the maps. The per cent per-pixel agreement for these groupings, excluding the agricultural mosaic and wetlands classes which do not nest into the UMD classes, is 74%.

Per-pixel agreement and disagreement can be seen in figure 2, which displays the agreement between the two maps for tall (forest and woody savanna/woodlands) and short/no vegetation (all other classes). As shown in figure 1, the overall areas of these classes differ very little. The global totals for tall vegetation differ by less than 1.6 million km², or less than 4% of the total tall woody land cover as expressed in each map. However, the per-pixel agreement for tall versus short vegetation is 84%. Figure 2 shows how most core forested areas are mapped similarly, while disagreements occur mostly along the edges of these areas and constitute wide regional variability. For example, west Africa in the DISCover map is woodier than in the

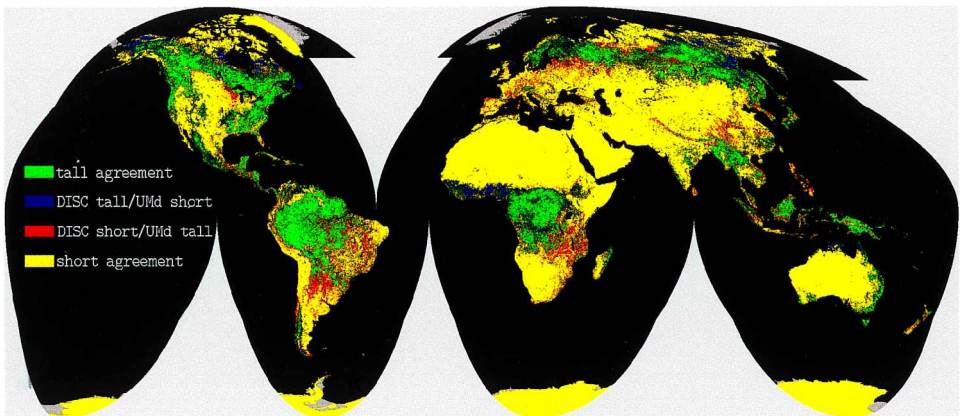


Figure 2. Global agreement of tall versus short vegetation for the EDC DISCover and UMD 1 km maps. Tall vegetation represents all forest classes plus the woody savanna class for the DISCover map and all forest classes plus the woodland class for the UMD map. Short vegetation is all other classes combined.

UMd map, while the opposite is true in southern Africa. Users whose models are sensitive to regional variability should be aware of these disagreements between the two products.

Figure 3 shows the areal totals for all classes and reveals two significant differences between the maps. First, while the aggregate forest/woody savanna and woodlands totals may be similar, the DISCover map has more forest, of all types, than the UMd map. The woody savanna/woodland class, conversely, has greater presence in the UMd map. Also, the wooded grassland class for the UMd map is more than twice the size of its savanna counterpart in the DISCover map. Much of this disagreement is related to the mosaic class used in the DISCover product. The overall result from these two differences is the increased presence of intermediate woody classes such as woody savannas/woodlands and savannas/wooded grasslands in the UMd map than in the DISCover map. Excluding the three classes not present in the UMd map, the per-pixel agreement for the remaining classes equals 48%. The fact that the agreements diminish greatly when viewing all of the classes versus aggregates is not surprising, but users who need this level of detail should examine the data themselves in order to judge which map is most useful for their purposes.

Three snapshots of local areas at full resolution have been included to reflect the level of concurrency between the two maps (the DISCover product is on the left-hand side and the UMd product is on the right). Figure 4(a) shows an area near Perth, Australia where both maps exhibit general thematic agreement with the consistent delineation of forest/woodlands and crops. Figure 4(b) shows an area along the United States/Canada border in the Pacific northwest where there is general agreement for the class aggregates of figure 1, but confusion within the aggregates themselves. There is disagreement within the forested area between evergreen needleleaf and mixed forests and between the open shrublands and grasslands of the Columbia River basin. The increased presence of forest in the DISCover map and of intermediate tree cover classes within the UMd map can also be seen along the forest/non-forest boundaries. Figure 4(c) shows an example in France of the identification of the same geographic entity, the Massif-Central, but of two different thematic depictions due to the use of different classification schemes. The absence of the cropland mosaic class in the UMd map (right-hand side) creates a significant divergence in the portrayal of this area for both maps. The DISCover map characterizes the area as an agricultural mosaic, while the UMd map has forests, woodlands, wooded grasslands and croplands present. At coarse resolutions, mixed pixels dominate in many areas such as Europe, and a consistent characterization between maps

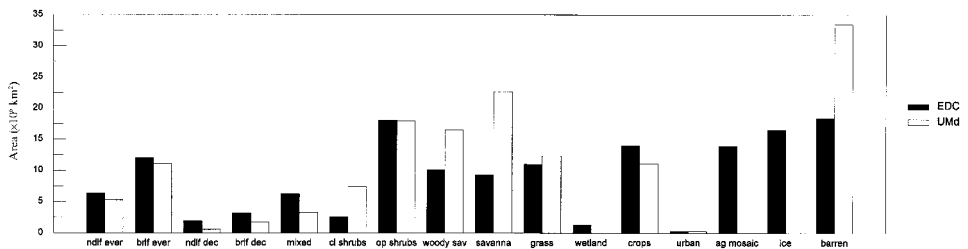


Figure 3. Global area totals for individual classes of the EDC DISCover and UMd 1 km maps, with the IGBP name listed. The UMd woodlands class is plotted as woody savanna, the wooded grassland class is plotted as savanna and the bare ground class is plotted as barren.

can be difficult to realize. Not having a common classification scheme assures dissimilar depictions and creates problems for evaluators and users.

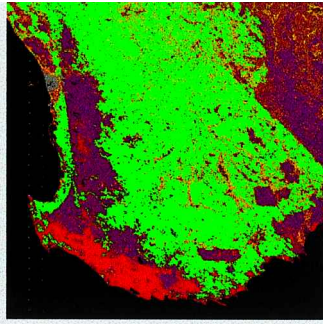
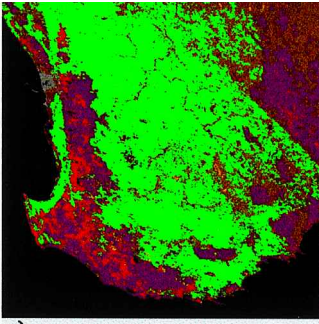
4. Comparison to ground-based maps

One of the primary reasons for developing land cover maps derived from remotely sensed data is the ability to improve upon traditional ground-based mapping methods. DeFries and Townshend (1993) revealed the level of disagreement present among traditional ground-based land cover maps and advocated the use of remote sensing to map global land cover as a way to create more consistent map products. A comparison of the agreement between the DISCover and UMd maps and two land cover maps derived from ground-based data revealed a gain in thematic agreement. Two maps, a one degree global vegetation map compiled by Matthews (1983) and 'Carbon in Live Vegetation of Major World Ecosystems' by Olson *et al.* (1983) were examined to make this comparison. The nesting of the data sets into individual classes allowed for a number of possible permutations, so each ground-based map was aggregated into the classes present in figure 1. Of these classes, the forest/woodland, grass/shrubs, crops, and barren classes are common to all four maps. The Olson and Matthews maps were reprojected into the Interrupted Goode Homolosine

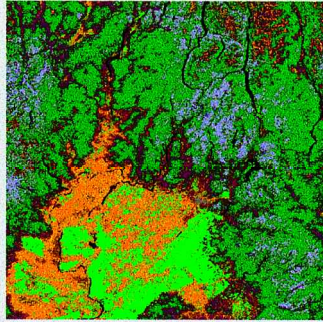
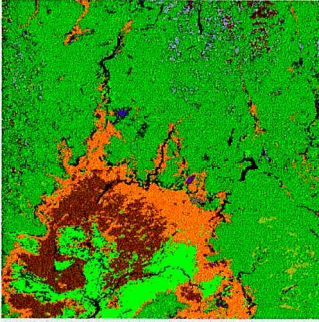
Table 2. Comparison of thematic agreement (in %) at nominal 0.5° resolution grid of remotely sensed derived DISCover and UMd maps and ground-based 'Global Distribution of Vegetation at 1° × 1°' compiled by Matthews (1983) and 'Carbon in Live Vegetation of Major World Ecosystems' by Olson *et al.* (1983).

DISCover/UMd	Forest/woodland	Grass/shrubs	Crops	Bare ground
Forest/woodland	88.6%	9.0	2.4	0.0
Grass/shrubs	15.8	69.0	10.0	5.2
Crops	8.6	12.2	79.2	0.0
Bare ground	0.0	9.3	0.0	90.7
Average class agreement = 81.87%				
Overall agreement = 80.32%				
Olson/Matthews	Forest/woodland	Grass/shrubs	Crops	Bare ground
Forest/woodland	70.2%	22.0	7.1	0.7
Grass/shrubs	14.9	60.0	6.7	18.5
Crops	16.1	29.5	51.1	3.2
Bare ground	0.5	14.6	1.0	83.9
Average class agreement = 66.30%				
Overall agreement = 68.35%				

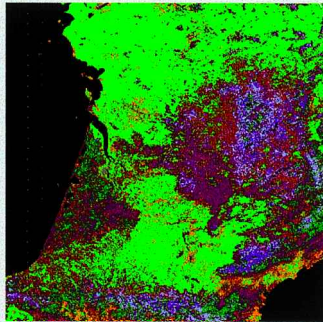
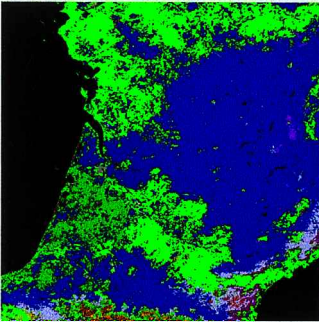
Figure 4. Full-resolution snapshots of areas depicting the general differences between the DISCover and UMd map products. The permanent wetlands and snow and ice classes are absent from these windows, leaving only the cropland/natural vegetation mosaic class as the only category not in common between the data sets. The DISCover product is on the left, and the UMd product is on the right with both representing 500 km by 500 km squares. (a) An area in south-west Australia centred at 117°29' E, 33°04' S. (b) An area along the Canada/United States border in the Pacific northwest centred at 118°10' W, 48°16' N. (c) An area in the south of France centred at 0°33' E, 44°57' N.











a)



b)



c)

- | | |
|-----------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------|
|  water |  closed shrubland |
|  needleleaf evergreen forest |  open shrubland |
|  broadleaf evergreen forest |  grassland |
|  needleleaf deciduous forest |  cropland |
|  broadleaf deciduous forest |  barren/bare ground |
|  mixed forest |  urban and built-up |
|  woody savanna/woodland |  cropland/natural vegetation mosaic |
|  savanna/wooded grassland | |

projection and the 1 km data sets were resampled to this coarser grid. The results comparing the Matthews/Olson and DISCover/UMd agreement are shown in table 2.

The ground-based maps have an average class disagreement of 33.7%, while that of the remotely sensed derived maps is 18.1%, indicating a reduction in the areas of disagreement by 46% ($(35.7 - 18.1)/33.7$). The overall disagreement is reduced by 38%. However, these are not measures of accuracy. In fact, two maps could have 100% agreement and still be entirely wrong. It is posited here that the synoptic view provided by remote sensing allows for a more consistent depiction of the Earth surface than do traditional approaches, even when given the greatly disparate classification approaches used in the making of the DISCover and UMd maps.

5. Discussion and conclusions

The approaches taken by the IGBP and the UMd to complete the task of characterizing the globe into a set of similar classes are very different. From the algorithms to the input variables, there is little in common between the two methods. However, while many differences do exist, there is also an amount of thematic agreement, especially at the class aggregate level. Aside from the input variables, algorithms and classification schemes, a number of external factors create variability which make it difficult to clearly compare the methodologies.

One important variable is the reliance on a wide and varying set of ancillary data sources within both techniques (DeFries *et al.* 1998). For the DISCover product, ancillary sources were used to label clusters resulting from the unsupervised classification algorithm. For the UMd map, ancillary sources were used to aid in interpretation of the original high-resolution data sets which, in turn, were used to create the 1 km training data set. Regional variability in the quality and reliability of these data sources is very high and introduces variability in the output map products. The IGBP LCWG validation workshop is a first step in attempting to generate a standardized global validation data set for use with coarse-resolution map products. The production of such data sets allows for the possibility of testing the differences between the methodologies by providing a global reference standard.

Another variable is the 1 km data set, which is also a first generation product. Artifacts exist within these data due to a variety of factors which, in turn, differentially affect map outputs. The presence of clouds, data gaps, misregistrations and other anomalies increase the probability of errors being portrayed in the final land cover products. The extent to which noisy data are manifested or ignored within the two mapping approaches and, thus, in the final maps is currently unknown. A more rigorous comparison of the two maps would include a discussion of the degree to which the two approaches handle noisy data. Future production of global data sets will employ new techniques for generating global satellite composites which will reduce many of the undesired effects associated with past approaches (El Saleous *et al.* 2000), and allow for comparisons relating classification methodologies directly to multi-spectral information.

The DISCover and UMd 1 km land cover products represent the first ventures into mapping global land cover at a moderate spatial resolution. Questions regarding appropriate methodologies, data sources and evaluation techniques are still under investigation. The future task is to discern areas of weakness within the present set of products and identify ways to produce improved iterations of these maps. This first review shows general agreement for broad vegetation categories, with low per-

pixel agreement for individual classes and significant regional variability. The determination of why the disagreements exist and how to improve correlation between maps are important research topics that might include the identification of core areas for major land covers, areas of mixed pixels, and the identification of multi-spectral information which best discriminates global land covers.

References

- BELWARD, A. S., (editor) 1996, The IGBP-DIS Global 1 km Land Cover Data Set 'DISCOVER': proposal and implementation plans. Report WP No. 13, IGBP-DIS, Stockholm, Sweden.
- DEFRIES, R. S., and TOWNSHEND, J. R. G., 1994, NDVI-derived land cover classifications at a global scale. *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 8 km spatial resolution: the use of training data derived from Landsat imagery in decision tree classifiers. *International Journal of Remote Sensing*, **19**, 3141–3168.
- EIDENSHINK, J. C., and FAUDEEN, J. L., 1994, The 1 km AVHRR global land data set: first stages in implementation. *International Journal of Remote Sensing*, **15**, 3443–3462.
- EL SALEOUS N. Z., VERMOTE, E. F., JUSTICE, C. O., TOWNSHEND, J. R. G., TUCKER, C. J., and GOWARD, S. N., 2000, Improvements in the global biospheric record from the Advanced Very High Resolution Radiometer (AVHRR). *International Journal of Remote Sensing*, **21**, 1251–1277.
- HANSEN, M. C., DEFRIES, R. S., TOWNSHEND, J. R. G., and SOHLBERG, R., 2000, Global land cover classification at 1 km spatial resolution using a classification tree approach. *International Journal of Remote Sensing*, **21**, 1331–1364.
- IGBP, 1992, Improved Global Data for Land Applications. IGBP Global Change Report No. 20, edited by J. R. G. Townshend, International Geosphere Biosphere Programme, Stockholm, Sweden.
- 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 1 km AVHRR data. *International Journal of Remote Sensing*, **21**, 1303–1330.
- MATTHEWS, E., 1983, Global vegetation and land use: new high resolution data bases for climate studies. *Journal of Climate and Applied Meteorology*, **22**, 474–487.
- MERCHANT, J. W., YANG, L., and YANG, W., 1993, Validation of continental-scale land cover data bases developed from AVHRR data. *Proceedings of the Pecora 12 Symposium, Land Information from Space-Based Systems*, Sioux Falls, South Dakota, 24–26 August 1993 (Bethesda, MD: ASPRS), pp. 63–72.
- OLSON, J. S., WATTS, J., and ALLISON, L., 1983, Carbon in live vegetation of major world ecosystems. Report W-7405-ENG-26, U.S. Department of Energy, Oak Ridge National Laboratory, Oak Ridge, TN, USA.
- RASOOL, S. I., 1992, Requirements for terrestrial biospheric data for IGBP Core Projects. IGBP-DIS Working Paper #2, International Geosphere–Biosphere Programme—Data and Information Systems, Paris, France.