



ISLSCP Initiative II global data sets: Surface boundary conditions and atmospheric forcings for land-atmosphere studies

Forrest G. Hall,¹ Eric Brown de Colstoun,² George J. Collatz,³ David Landis,² Paul Dirmeyer,⁴ Alan Betts,⁵ George J. Huffman,² Lahouari Bounoua,³ and Blanche Meeson⁶

Received 3 April 2006; revised 25 September 2006; accepted 23 October 2006; published 30 November 2006.

[1] We report herein the publication and evaluation of the International Satellite Land Surface Climatology Project (ISLSCP) Initiative II global interdisciplinary data record. The record consists of 52 data sets, with a common series in the 10-year period 1985 to 1996. Selected data series extend well beyond this period. All series are coregistered to a common grid and gap-filled for continuity using uniform procedures. We describe briefly the individual data sets within the collection; provide user guidance; and contrast, compare and evaluate those data sets containing similar parameters (land cover, NDVI, albedo, precipitation and near-surface meteorology). We also describe the process used to develop the Initiative II collection which involved a broad international scientific community focused on addressing a well-defined set of carbon, water and energy cycle questions within the context of a specific set of analysis tools. The communities that drove the definition of the Initiative II collection were investigators within the international scientific communities of the Global Energy and Water cycle Experiment, GEWEX, program (<http://www.gewex.org/>); the International Geosphere/Biosphere Program IGBP (<http://www.igbp.kva.se/>); and the U.S. Global Change Research Program, USGCRP (<http://www.usgcrp.gov/>). Finally, we report usage statistics based on access and download of files from the ISLSCP Initiative II collection available at <http://www.daac.ornl.gov>.

Citation: Hall, F. G., E. Brown de Colstoun, G. J. Collatz, D. Landis, P. Dirmeyer, A. Betts, G. J. Huffman, L. Bounoua, and B. Meeson (2006), ISLSCP Initiative II global data sets: Surface boundary conditions and atmospheric forcings for land-atmosphere studies, *J. Geophys. Res.*, *111*, D22S01, doi:10.1029/2006JD007366.

1. Background

[2] The International Satellite Land Surface Climatology Project (ISLSCP) Initiative I data collection, a pilot project, produced the first interdisciplinary Earth Science collection of global data to support land-atmosphere exchange studies [Sellers *et al.*, 1996a]. Initiative I produced a 2-year data set spanning 1987–1988 and containing global, monthly, 1° spatial resolution fields of vegetation attributes, near-surface meteorology, atmospheric radiation and clouds, precipitation, river routing, runoff, soils, and snow/ice data. Each data series in the collection was peer reviewed, registered to a common grid, reprocessed to a common format, and carefully documented. The collection was published in a 5 CD

set and distributed by the Goddard Space Flight Center (GSFC) Data Analysis and Archive Center (DAAC). Over 13,000 sets have been ordered from the DAAC and over 267,000 files have been downloaded. There are over 500 citations in the scientific literature supporting a wide variety of uses. Given the success and unique contributions of the ISLSCP Initiative I collection, it was recognized that such collections should be continued and expanded to at least 10 years to enable studies of interannual variability and to include newer state-of-the-art data sets needed to more fully address specific Earth science issues. Accordingly a follow-on effort led by National Aeronautic and Space Administration (NASA), involving a host of national and international partners, was initiated to produce the ISLSCP Initiative II collection, and fulfill the scientific community requirements.

1.1. ISLSCP Data Initiative Process

[3] It takes a community to build a data collection. A community focused on addressing a well-defined set of science questions using a well-defined set of models, analysis tools and data. The communities that drove the definition of the Initiative II collection were investigators within the international Global Energy and Water cycle Experiment, GEWEX, program (<http://www.gewex.org/>),

¹University of Maryland Baltimore County, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.

²Science Systems and Applications Inc., NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.

³NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.

⁴Center for Ocean-Land-Atmosphere Studies, Calverton, Maryland, USA.

⁵Atmospheric Research, Pittsford, Vermont, USA.

⁶Ocean US, Arlington, Virginia, USA.

Table 1. Science Foci of GEWEX and USGCRP/IGBP

GEWEX FOCI	USGCRP, IGBP FOCI
How are global precipitation, evaporation and the cycling of water changing?	What are the magnitudes and distributions of carbon sources and sinks on seasonal to centennial timescales, and what are the processes controlling their dynamics?
What are the effects of clouds and surface hydrologic processes on Earth's climate?	What are the magnitudes and distributions of ocean carbon sources and sinks on seasonal to centennial timescales, and what are the processes controlling their dynamics?
How are variations in local weather, precipitation and water resources related to global climate variation?	What are the effects on carbon sources and sinks of past, present, and future land use change and resource management practices at local, regional, and global scales?
What are the consequences of land cover and land use change for human societies and the sustainability of ecosystems?	How do global terrestrial, oceanic, and atmospheric carbon sources and sinks change on seasonal to centennial timescales, and how can this knowledge be integrated to quantify and explain annual global carbon budgets?
What are the consequences of climate change and increased human activities for coastal regions? How can weather forecast duration and reliability be improved?	What will be the future atmospheric concentrations of carbon dioxide, methane, and other carbon-containing greenhouse gases, and how will terrestrial and marine carbon sources and sinks change in the future?
How can predictions of climate variability and change be improved?	How will the Earth system, and its different components, respond to various options for managing carbon in the environment, and what scientific information is needed for evaluating these options?
How will water cycle dynamics change in the future?	

the International Geosphere/Biosphere Program IGBP (<http://www.igbp.kva.se>), and the US Global Change Research Program, USGCRP (<http://www.usgcrp.gov/>). The scientific foci of these organizations are defined in terms of specific sets of science questions (Table 1). To address these questions quantitatively, the community has developed both an analysis framework and data requirements to feed and validate its elements (Figure 1). The Initiative II data

collection was defined and developed by this community, meeting in regular twice-yearly workshops (Figure 2). The process was coordinated by the ISLSCP Initiative II staff located at the Goddard Space Flight Center and guided by a science working group (Figure 2) through monthly teleconferences. The GSFC staff coordinated the project but also developed the FASIR-NDVI data set with associated biophysical parameters for the period 1982–1998. The staff

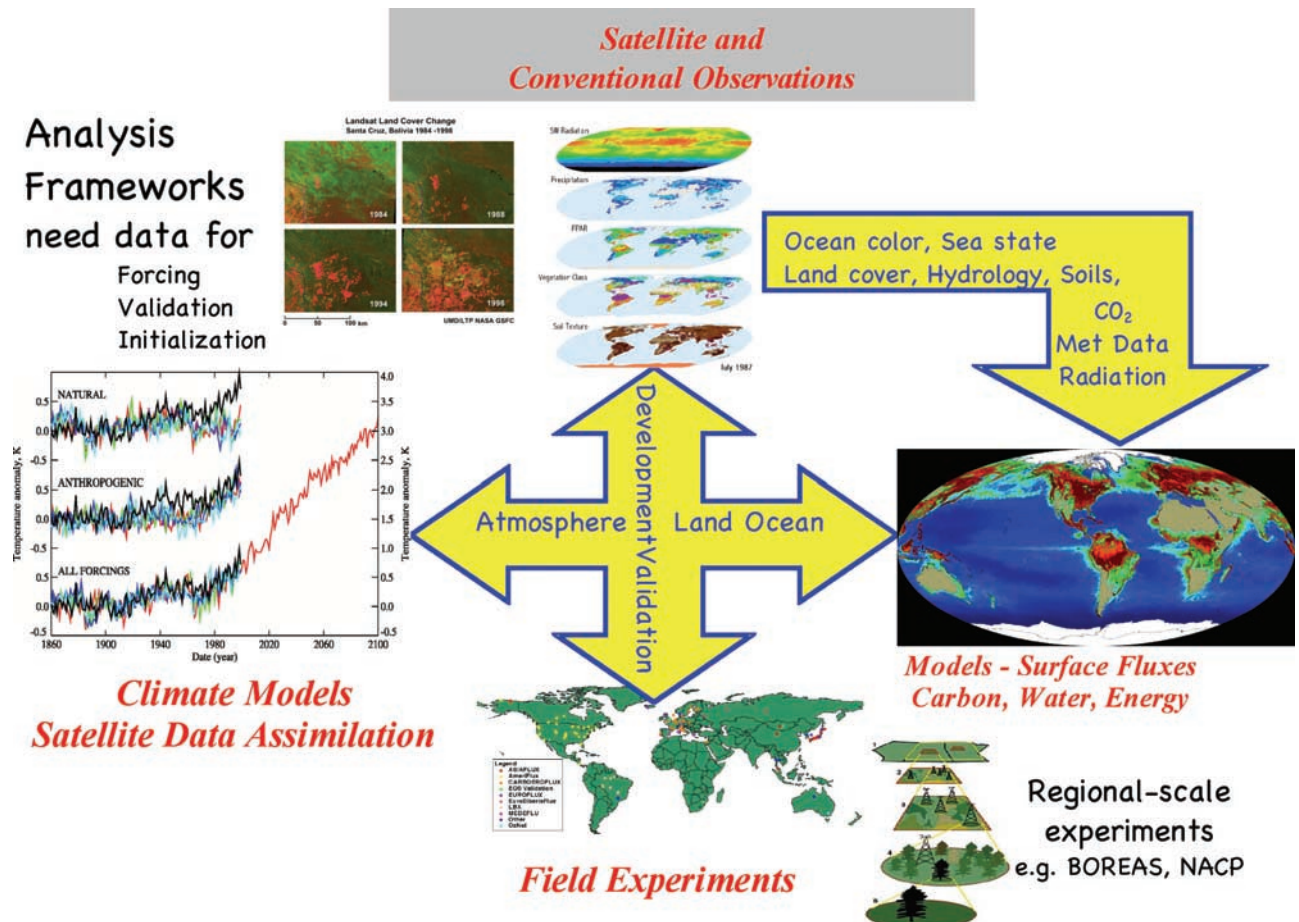


Figure 1. Analysis framework for GEWEX and IGBP/BAHC.

ISLSCP Data Collection

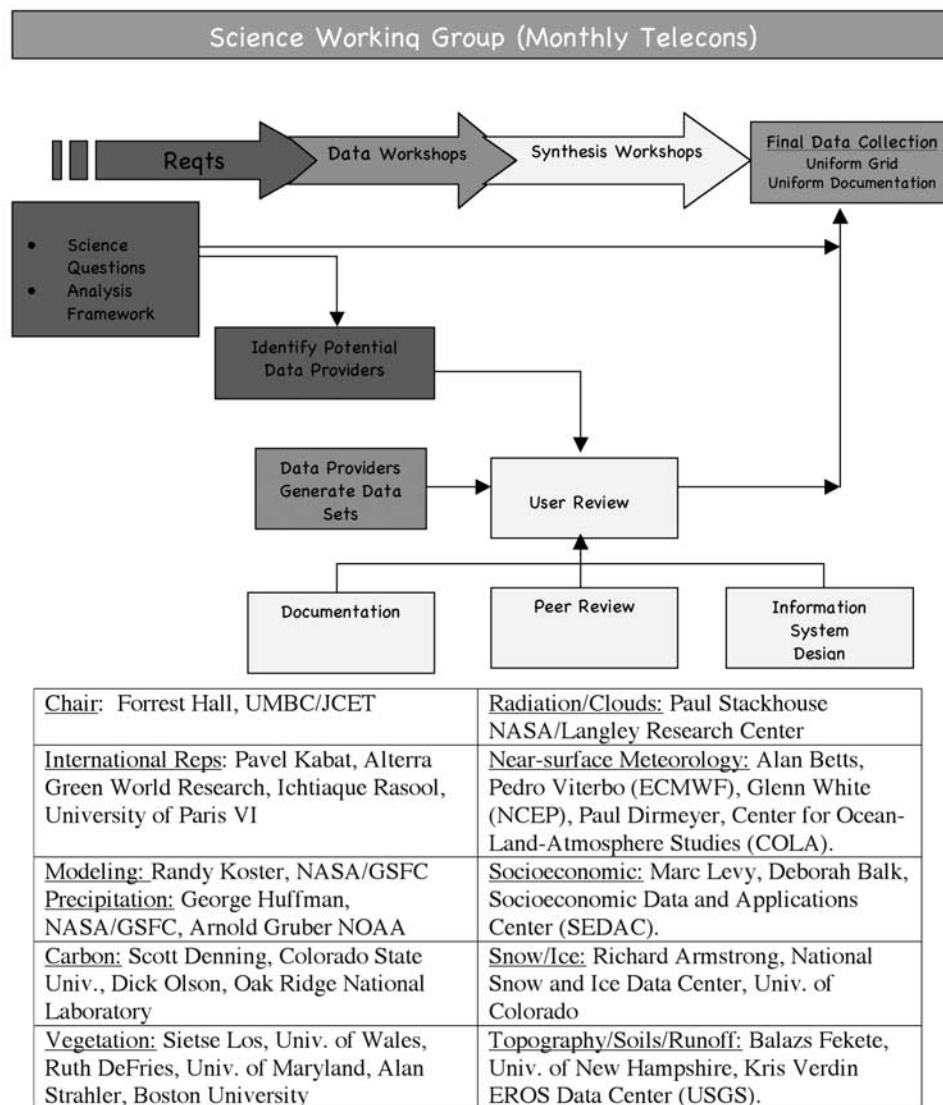


Figure 2. ISLSCP Initiative II process. Workshops were held twice yearly and involved users and providers. Monthly teleconferences with the ISLSCP Science Working Group were critical as data requirements and data production evolved.

organized and coordinated the first Initiative II workshop in October of 1999 and each six months thereafter, culminating in an Initiative II data evaluation workshop in May of 2005.

[4] The workshops were attended on average by about 50 scientists, consisting of the future users of the data including hydrologists, meteorologists, ecologists and the data providers covering expertise from remote sensing to meteorology and Earth and soil sciences. Without these meetings, some of the Initiative II data sets would not have been generated, and those that were, likely would not have corresponded to user’s precise needs. Certainly they would not have been fully documented, gridded, and been in compatible units or in compatible formats.

1.2. Framework Motivating the Initiative II Collection

[5] The 52 Initiative II data sets and their properties (Appendix A) were selected to develop, provide inputs to,

or validate the results of the elements of a well-defined analysis framework including models and climate observations (Figure 1). Developed over the past few years by the science community, this analysis framework addresses a specific set of science questions (Table 1) focused on quantifying how the Earth is changing and the consequences of these changes for life on Earth. The science questions and analysis framework that informed Initiative I, focused to a large extent on water and energy cycling, have evolved in the intervening years. Initiative II changed accordingly to include the carbon cycle and its interannual variability, and in a limited but important first attempt, human dimensions.

[6] An important development over the past few years within the science community that motivated the content and structure of the Initiative II collection was the development of an analysis framework to address surface-atmosphere exchanges and transport of carbon, in addition to the

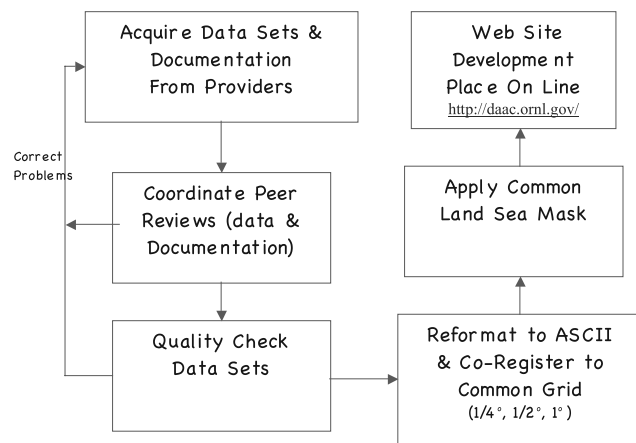


Figure 3. ISLSCP Initiative II data processing flow. Data were placed on line as beta test data as soon as data were received from provider.

water and energy cycle. This new framework described in Figure 4 in section 2.1 integrates (1) “top-down” (atmospheric) approaches that use atmospheric carbon dioxide concentration measurements and transport models to quantify surface sources and sinks of carbon and (2) “bottom-up” (process) approaches that use biogeochemical models and field studies to elucidate the ecological, biological and physical processes involved in the surface carbon sources and sinks. Each of these elements can independently produce estimates of land-atmosphere fluxes, but comparisons between their outputs are essential to quantify, understand the underlying causes of the sources and sinks and to validate estimates of their strength. Described in the various sections to follow, are how the Initiative II data series support this new analysis framework.

[7] As can be seen in Appendix A, there are multiple data series available for some variables, for example, there are five different products for albedo (33 through 38), eight products for land cover (39,40, 45, 46, 48, 49, 50, and 51) three products for vegetation biophysical parameters and NDVI (42,43 and 44), two products for near-surface meteorology fields (22 and 23), and four products for precipitation fields (22 through 25). One of the principal reasons for including multiple data series for single parameters was that no single available data set met all of the requirements of ISLSCP Initiative II in terms of spatial (at least 1°) and temporal (1986–1995 coverage at monthly time step) resolutions and quality. This paper will briefly describe the individual data sets, their generation and contrast and compare the multiple data series and provide user guidance in their selection.

1.3. Processing and Data Preparation

[8] Figure 3 diagrams the data processing and preparation approach employed by a small GSFC staff (about 1.5 full-time equivalents) to coordinate the production, acquisition, peer review, evaluation, rework, gap-filling, and gridding of the data. The staff also produced the FASIR biophysical parameter data set. Each Initiative II data set was placed online in beta test mode as soon as it was provided. The project consisted of two phases; the assembly of a beta

version of the collection lasting 3 years, and the evaluation of the collection lasting 1 1/2 years.

[9] Data and documentation underwent two separate peer reviews: a scientific review by independent producers of similar types of data and a usability review by a scientist or group of scientists who would be using the data in their research. This was an essential step and uncovered a number of problems with both data and documentation that were then fed back to the providers for rework and completion. Staff members also provided data quality checks to ensure that the data were properly formatted and spatially and temporally complete. A common land sea mask was defined and applied to the applicable data sets. A gap-filling procedure was developed to deal with discrepancies between the provider’s land-water mask and the Initiative II common land sea mask. All cells that have been modified from the original data are made available to the users in separate files. Each data set was then gridded to a common grid varying in resolution from $1/4^\circ$ to 2.5° latitude and longitude. For compatibility, data sets with native resolutions greater than 1° were regridded to 1° . Except for data sets based on point measurements, all parameters in the entire data collection are available in common 1° versions.

2. ISLSCP Initiative II: Data

[10] In this section we describe the individual elements of the data collection and the rationale for their inclusion. In the Comparison and Evaluation section to follow, we contrast, compare and evaluate those data sets that contain common parameter sets (land cover, NDVI, albedo, precipitation and near-surface meteorology) and provide guidance for their use.

2.1. Carbon and Socioeconomic Data

[11] The carbon analysis framework described in the section above, Framework Motivating the Initiative II Collection and in Figure 4, motivated the data selections for this Initiative II category.

[12] To scale our understanding of the physical and biological mechanisms underlying the surface-atmosphere exchange of carbon water and energy across the entire continent the modern analysis framework utilizes bottom-up biogeochemical process models that utilize leaf-level photosynthesis relationships to couple the leaf-level and landscape level carbon, water and energy cycles. These models permit the direct computation of surface-atmosphere carbon dioxide exchange as a function of remote sensing and climate inputs such as those contained in the vegetation and near-surface meteorology data (12 through 51). Furthermore, the models rely on relationships deduced from plot-level ecology studies that can be developed and validated using the Initiative II flux tower data (5) measured from towers. Unfortunately, flux tower data were available to Initiative II for only selected U.S. flux towers. For additional validation of the model outputs, gridded Net Primary Productivity (NPP) estimates from 12 biogeochemical models were included in the Initiative II collection (8 and 11). A complete terrestrial carbon budget requires data on emissions from fossil fuel combustion (1, 3 and 4),

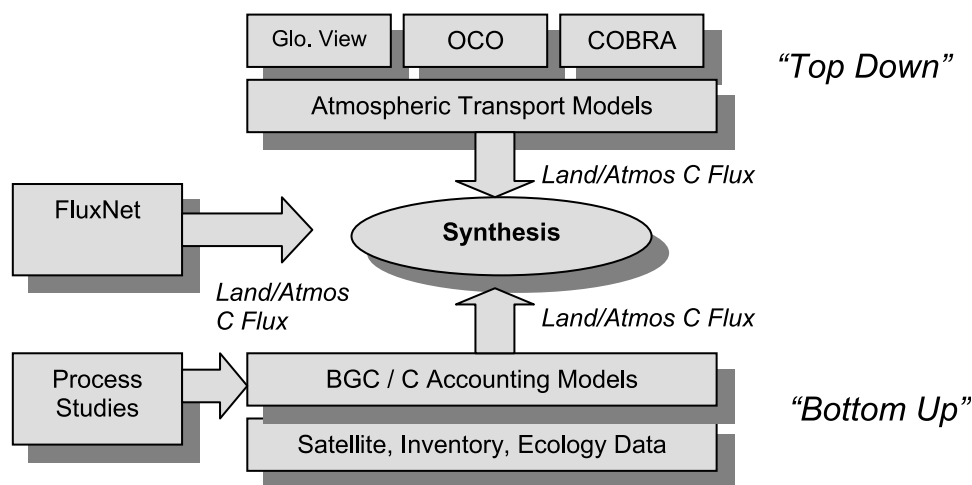


Figure 4. Conceptual experimental framework carbon analyses based on process models (“bottom up”) and atmospheric measurements (“top down”), both capable of estimating continental-scale carbon fluxes, once parameterized via process studies and FluxNet data.

continental erosion (2) and carbon loss from the land by riverine transport (6).

[13] Continental-scale predictions from the process models must then be reconciled against “top-down” analyses based on atmospheric measurements made at scales of 100–1000+ km. Top-down techniques utilize the temporal and spatial variations in atmospheric methane and CO₂ concentration data (9) and reanalysis winds provided by the Initiative II near-surface data (22 and 23) to track these variations back to their surface origin. This analysis component of the framework is essential to quantify the location and timing of surface-atmosphere exchanges of carbon, water and energy.

[14] To assess the contribution of the human dimension to carbon fluxes the Initiative II collection includes gridded population and Gross Domestic Product (GDP) data (sets 31 and 32) from the Center for International Earth Science Information Network (CIESIN) at Columbia University. These are global, gridded data at three spatial resolutions of 1/4°, 1/2° and 1° and for the reference years of 1990 and 1995. The Initiative II GDP data are distributed regionally to facilitate the integration of GDP with other data at a subnational level and to promote interdisciplinary studies that include socioeconomic aspects. This data set estimates GDP density on a grid at three resolutions: 1/4°, 1/2° and 1°.

2.2. Vegetation

[15] Vegetation is an important pathway through which soil water is transferred to the atmosphere during the process of photosynthesis. Carbon uptake is coupled to water diffusion from plants stomata and affects surface climate. Vegetation’s control on surface climate is exerted through at least three distinct mechanisms. (1) The structure of the vegetation affects the aerodynamic exchanges through roughness elements and alters the partitioning of incoming energy. (2) The optical properties of leaves determine the amount of energy absorbed by plants hence albedo, and (3) through its photosynthetic function, and the amount of leaves on the canopy, vegetation alters the partitioning of surface water and energy fluxes. Vegetation type, morphological, optical and physiological properties

are therefore crucial parameters used by Surface Vegetation transfer Schemes (SVAT) to estimate surface fluxes of carbon, water and energy. Other difficult to measure parameters such as land cover history and ecosystem rooting depth are also important for landscape dynamics, primarily in determining the soil carbon stores, nutrient levels and hydrological characteristics.

[16] To accommodate these needs, the Initiative II collection contains 8 different state-of-the-art data sets dealing with various aspects of land cover and land use. The principal land cover data sets for Initiative I are the University of Maryland (UMD) Land Cover data set (50) of Hansen *et al.* [2000], the IGBP-DIScover vegetation classification (551), and the UMD continuous fields of vegetation cover (40). All were derived from global AVHRR data at a 1 km spatial resolution [Eidenshink and Faundeen, 1994]. The 1-year data set spanned 1992–1993 and was produced under the auspices of the Data and Information System of the International Geosphere Biosphere Programme (IGBP-DIS). In the process of aggregating land cover type to coarser Initiative II grid cells, the percentages of each land cover type within the cell are retained and used to calculate the dominant type of the cell; the dominant type and the percentage of each land cover type in each cell are provided as multiple separate layers for the users.

[17] To track monthly, annual and interannual variations in land vegetation, two monthly NDVI time series over land, FASIR (42 and 43) covering the years 1981–1998 and (44) GIMMS (1991–2002) are also included in the Initiative II collection. As discussed below in the section entitled Multiple Data Series the GIMMS and FASIR data series are processed using different approaches and have been evaluated and compared with each other by Hall *et al.* [2006] where these approaches and their evaluation are reported in detail.

[18] Identification of the dominant photosynthetic pathway (C3 or C4) in terrestrial vegetation communities is essential for accurate calculations of exchanges of carbon, water, and energy. C3 and C4 pathways respond differently to light, temperature, CO₂, and nitrogen; they also differ in

physiological functions such as stomatal conductance and isotope fractionation. Thus a fine-scale distribution of these plant types is essential for Earth science modeling. For the Initiative II C3 C4 data series, the fraction of C4 vegetation in a community (39) was derived from climate and land cover data included in the Initiative II collection [Collatz *et al.*, 1998; Still *et al.*, 2003].

[19] Ecosystem rooting depth (41) is a key variable in the surface energy and water budget. Vertical root distributions influence the fluxes of water, carbon, and soil nutrients and the distribution and activities of soil fauna. Roots transport nutrients and water upward, but they are also pathways for carbon and nutrient transport into deeper soil layers and for deep-water infiltration. Roots also affect the weathering rates of soil minerals. Global distribution of plant rooting depths is based on the global aboveground vegetation structure and climate. For calculating such processes on a global scale, data on vertical root distributions are needed. Initiative II project procured the resources necessary to generate such data. Vertical rooting depths were collected from the literature in order to construct maps of global ecosystem rooting depths [Schenk and Jackson, 2002]. The parameters included in these data sets are estimates for the soil depths containing 50% and 95% of all roots, termed 50% and 95% rooting depths. Together, these variables can be used to calculate estimates for vertical root distributions, using a logistic equation provided in the documentation. The data represent mean ecosystem rooting depths for 1° grid cells.

[20] To provide a historical context for the landscape, two historical land cover data sets are also included: the Historical Croplands Fractional Cover data set (45) of Ramankutty and Foley [1998], with data from 1700 to 1992 and a related Historical Land Cover and Land Use (1700–1992) data set (46) from the National Institute of Public Health and the Environment (RIVM) in Netherlands [Klein Goldewijk, 2001]. To construct the Historical Croplands Fractional Cover data, Ramankutty and Foley [1998] derived a spatially explicit data set of croplands for 1992 using the IGBP-DIScover remotely sensed land cover data set (51) of Loveland and Belward [1997] together with contemporary land inventory data, then extended the 1992 data set to 1700 using historical land inventory data. By extending their data set back in time, Ramankutty and Foley [1999] were also able to produce a land cover map of “potential” vegetation (49), or the natural vegetation before human alteration, which is included in this collection. Klein Goldewijk [2001] used historical statistical inventories of agricultural land (census data, tax records, land surveys, etc.) and various spatial analysis techniques to create a geographically explicit data set of land use change with a regular time interval. These two new global data sets of historical land cover change compare fairly well over most of the Earth even though the modeling approaches and input data used are quite different [Klein Goldewijk and Ramankutty, 2004].

2.3. Near-Surface Meteorology

[21] Initiative II near-surface meteorology data sets contain a monthly climate data series (24) and two 3-hourly data sets from reanalysis (22 and 23). The reanalysis data are global, while the climate observations cover the Earth’s terrestrial surface, excluding Antarctica. The monthly climate variables cover the period 1986 to 1995. It was created

by the Climatic Research Unit (CRU) at the University of East Anglia, in the United Kingdom and is a subset of Version 5 of their data set [New *et al.*, 1999, 2000]. The meteorological reanalysis data are 3-hourly, at $1 \times 1^\circ$ spatial resolution. The 3-hourly data are also averaged to provide monthly, monthly 3-hourly (i.e., monthly mean diurnal cycle) for the forecast fields and monthly 6-hourly for the analysis fields. The ERA40 data set is from the European Center for Medium Range Forecasting (ECMWF) reanalysis. The other near-surface data set was derived by the Center for Ocean-Land-Atmosphere (COLA) for Initiative-II from the National Centers for Environmental Predictions (NCEP)/Department of Energy (DOE) Atmospheric Model Intercomparison Project (AMIP)-II Reanalysis; hereafter named NCEP2 reanalysis.

[22] The CRU monthly climate mean climatology (25) is a temporal subsample of the longer 1901 to 1996 CRU Version 5 (CRU05) monthly time series data set. This climatology was constructed using station data climatological normals between 1961 and 1990. A total of 19,800 precipitation and 3615 wind speed station observations were included [New *et al.*, 1999, 2000]. The station data were interpolated as a function of latitude, longitude and elevation using thin-plate splines. The accuracy of the interpolations was assessed using cross validation and comparison with other climatologies. The temperature analysis is generally superior to the humidity analysis, because synthetic data (found by estimating monthly dew point from monthly mean minimum temperature) is added to the humidity analysis in regions of sparse data. Other climate variables included in the CRU05 data set include precipitation, radiation, temperature, cloud cover, frost frequency, vapor pressure and wet day frequency.

[23] The ECMWF data set has been derived from their 45-year reanalysis, usually known as ERA40 [Uppala *et al.*, 2005], which covers the period September 1957 to August 2002. A recent version of the ECMWF Numerical Weather Prediction system (cycle 23r4) is used for the entire analysis period. The advantage of reanalysis over operational analysis is that no system changes occur that might affect the analysis products, although there are significant changes in the observations. The ECMWF data, which span the common ISLSCP Initiative II period from 1986 to 1995, have been interpolated from the slightly larger model grid to the Initiative II uniform 1° global grid, as much as possible consistent with the land-sea mask definitions [see Betts *et al.*, 2006].

[24] The NCEP2 near-surface data set for ISLSCP-II was derived by COLA from the NCEP/DOE reanalysis covering the years 1979–2003 [Kanamitsu *et al.*, 2002]. The purpose of the reanalysis was to provide an improved version of the original NCEP/NCAR reanalysis [Kalnay *et al.*, 1996; Kistler *et al.*, 2001] for use by the Atmospheric Model Intercomparison Project (AMIP) II for GCM validation. The NCEP/DOE reanalysis uses a very similar analysis system to the NCEP/NCAR reanalysis and an upgraded version of the same general circulation model, with known errors fixed and assimilation of a more complete stream of observational data after 1993. To coregister the NCEP/DOE reanalysis on the ISLSCP 1° grid, data was regridded from its native T62 Gaussian grid (192×94 grid boxes globally) to $1 \times 1^\circ$ resolution. The fields that are used for Initiative II are near-surface meteorological fields, fluxes of heat, moisture and

momentum, radiation at the Earth's surface, and land surface state variables. There are five temporal categories of data; time invariant and monthly mean annual cycle fields (together referred to as "fixed" fields); monthly mean fields; monthly 3-hourly (mean diurnal cycle) fields, and 3-hourly fields. Two types of variables exist in this data; instantaneous fields (primarily state variables), and average fields (primarily flux fields expressed as a rate).

[25] These ERA40 and NCEP2 reanalysis data sets also include surface and top-of-atmosphere shortwave wave (SW) and longwave wave (LW) radiation fluxes (see next section), precipitation, including convective precipitation and snowfall, snow depth and runoff, and the surface sensible and latent heat fluxes. In addition, the ERA40 data set for ISLSCP Initiative II includes a set of boundary layer variables, about 100 m above the surface, to drive land surface models in stand alone modes.

2.4. Radiation and Clouds

[26] These data series (26 and 27) contain monthly, daily, 3-hourly, and monthly/3-hourly (diurnally resolved monthly averages) surface and top-of-atmosphere (TOA) radiation budget and monthly averaged cloud parameters over the globe at 1° spatial resolution. The SRB parameters are derived using radiative transfer based algorithms applied to the cloud data provided by the International Satellite Cloud Climatology Project (ISCCP) [Rossow *et al.*, 1996; Rossow and Schiffer, 1999]. The Initiative II SRB data differ from a similar set of radiative flux parameters derived from ISCCP, called "ISCCP-FD" [Zhang *et al.*, 2004].

[27] The monthly cloud parameters include total cloud amount, and several cloud optical and thermodynamic parameters including cloud optical depth, cloud top pressure, and temperature. Monthly column water vapor, total column ozone and surface skin temperature are also included in the monthly fields. All monthly and monthly 3-hourly parameters except TOA insolation include files with a monthly mean value, a monthly standard deviation, and monthly minimum and maximum values. Radiation parameters include downward, upward, and net SW and LW surface radiative fluxes. The data are intended for use in evaluation of climate and data assimilation products and will provide long-term diagnostic information on regional changes of surface radiation. The data also have demonstrated usefulness in interdisciplinary studies of land surface, biological, oceanographic, and cryospheric processes.

[28] Radiative fluxes are computed from the ISCCP cloud data for all-sky and clear conditions enabling the estimates of cloud radiative forcing of the energy fluxes. Several estimates of the different components of the SW radiative flux including direct, diffuse and photosynthetically active radiation (PAR) are also provided.

[29] To generate the SRB fluxes, ISCCP cloud properties are input into the algorithms documented by Pinker and Laszlo [1992] for SW and Gupta *et al.* [1992] for LW. Meteorological profile information is developed from the NASA Data Assimilation Office (DAO) Goddard Earth Observing System version 1 (GEOS-1) reanalysis. Ozone abundance is provided from Total Ozone Mapping Spectrometer (TOMS) and TIROS Operational Vertical Sounder measurements (TOVS) via the ISCCP data sets. Aerosol information is crudely included in the SW algorithm by

assuming aerosol properties on the basis of three surface types. Surface albedo is retrieved from clear-sky radiance information from ISCCP in the Pinker and Laszlo SW model assuming spectral variation on the basis of the land cover information from Matthews [1985]. Surface emissivity maps for LW calculations have been created from the IGBP Discover land surface data set contained in Initiative II.

[30] The ERA40 and NCEP2 reanalysis data sets include surface and top-of-atmosphere SW and LW radiation fluxes. The comparison of these with the SRB data [Betts *et al.*, 2006] is informative. Differences in the SW fluxes for many regions result from errors in the reanalysis of cloud fields; although there are some regions such as the Tibetan plateau where the SRB SW data has known biases and the SW fluxes from ERA40 may be better. For the surface LW fluxes, the SRB values depend on near-surface temperatures from the GEOS-1 reanalysis, and for some regions, such as high latitudes in winter, these have significant cold biases; so that the LW fluxes in ERA40 are probably superior in some regions [Betts *et al.*, 2006].

2.5. Hydrology, Topography, and Soils

[31] This category contains five data types aimed at quantifying the vertical transport of water between the atmosphere and terrestrial watersheds and the water movement within a watershed: (1) precipitation, (2) topography and elevation-based derivatives, (3) a soils data set with 18 variables including soil texture, carbon and hydraulic/thermal properties, (4) river routing and runoff and (5) global soil water storage in the rooting zone data. These data are useful for model validation as well as model development and diagnostic studies.

2.5.1. Precipitation

[32] Four gauge and/or satellite-based precipitation data sets spanning 1986 to 1995 are included; 13 monthly gauge-based daily precipitation, 14 monthly gauge-based precipitation from the Global Precipitation Climatology Center (GPCC), and 15 a satellite and gauge-based pentad precipitation series from the Global Precipitation Climatology Project (GPCP). As described in section 2.3 Initiative II also contains precipitation data within its monthly climate series (24) as well as within 22 and 23 its reanalysis-based near-surface meteorology data series. These data sets are described and discussed in more detail in section 7.

2.5.2. Topography

[33] Initiative II includes an aggregated version of HYDRO1k, developed at the U.S. Geological Survey's (USGS) National Center for EROS using their 30 arc-second digital elevation model (DEM) of the world, GTOPO30 [Gesch *et al.*, 1999] at spatial resolutions of 1/2 and 1°. The Initiative II version of HYDRO1k (12) provides statistical information (mean, standard deviation, skewness and kurtosis) in elevation, slope, aspect and a compound topographic index for each grid cell on the basis of the HYDRO1k data at their native 1 km resolution. The HYDRO1k data sets have been developed on a continent-by-continent basis for all landmasses of the globe, with the exception of Antarctica, Greenland and, for data quality reasons, the continent of Australia. A preliminary Antarctica, Greenland and mainland Australia portion of the data set were produced and are contained in Initiative II. However, the data layers for these three landmasses have not been

subjected to the same quality assessment as the other continents.

2.5.3. River Routing

[34] The gridded river networks for Initiative II (19) are based on the Simulated Topological Network, or STN-30p [Vörösmarty *et al.*, 2000], which was developed to provide the large-scale hydrological modeling community with an accurate representation of the global river system. STN-30p was developed prior to HYDRO1k, therefore its river network topology and all the derived information from the network such as basin delineation, upstream area, distance to oceans, etc. are not completely consistent with the USGS version of HYDRO1k. The elevation field provided with the STN-30p Initiative II data set combines HYDRO1k aggregated elevation at 30-min resolution with STN-30p, where the inconsistencies between the elevation and the flow direction data sets (i.e., increasing elevation along downstream flowpath) were eliminated. A $1/2^\circ$ and a 1° version of the STN-30p network are provided in the Initiative II collection and contain the corrected elevation data as well as multiple gridded data layers with associated basin and cell attribute tables with ancillary information on river basins and upstream cells.

2.5.4. Runoff

[35] Gridded monthly runoff fields were generated [Fekete *et al.*, 2002] by combining model-generated runoff estimates with *observed* river discharge data from the Global Runoff Data Center (GRDC). Initiative II contains both the estimated gridded monthly runoff fields (21) and the GRDC river discharge data (18). When using the gridded monthly runoff fields special note should be taken that the fields combine both model results and observations. To generate the monthly runoff fields, GRDC stations were first coregistered to STN-30p. Then the ratios of the observed versus modeled average annual runoff were applied as correction coefficients to the monthly modeled estimates to generate the monthly runoff fields (the mean annual water balance model runoff estimates were computed by averaging the modeled monthly water balance model runoff estimates from the interstation regions of the discharge monitoring stations). The resulting data set is intended to demonstrate the value of combining river discharge observations with spatially distributed runoff estimates from water balance calculations.

2.5.5. Soils

[36] The Initiative II soils data set (17), 1° gridded global maps of 18 selected soil parameters, including soil texture, are provided for two soil depths (0–30 cm and 0–150 cm). This data set was produced by the ISLSCP staff using a bootstrapping approach to link the soil units of the FAO/UNESCO Digital Soil Map of the World [Food and Agriculture Organization, 1995] to the pedon records (e.g., depth, particle size distribution, bulk density and extractable nutrient composition, etc.) in the International Soil Reference and Information Centre (ISRIC) Global Pedon Database. This extensive suite of pedosphere properties was assembled by the Data and Information System framework activity of the International Geosphere-Biosphere Programme (IGBP-DIS) from many disparate data collections held by the United States Department of Agriculture (USDA), the Food and Agriculture Organization (FAO) of the United Nations, and ISRIC, as well as

national soils institutes, individual soil scientists, and users of soil data. The original IGBP-DIS data collection is accessible at <http://daac.ornl.gov/>. The provision in this data set of multiple depth layers, additional texture classes, and numerous soil hydraulic parameters based on realistic data and robust methods provides a significant advance over the previous soils data sets.

2.5.6. Root Zone Soil Water Storage Capacity

[37] The Global Soil Water Storage Capacity of the Rooting Zone data set (20) provides a method to describe potential vegetation rooting characteristics. Two inverse methods were employed to describe the extent of the rooting zone water storage size. The first method is based on the assumption that vegetation has adapted to the environment such that it makes optimum use of water [Kleidon and Heimann, 1998]. Using a simulation model of the land surface-vegetative cover, this method was implemented by maximizing absorption of Photosynthetically Active Radiation (PAR), leading to a maximization of evapotranspiration. The second method is based on the assumption that green vegetation indicates sufficient available water for transpiration. Rooting zone water storage size was inferred by minimizing the discrepancy of model simulated PAR absorption to satellite-derived PAR absorption. Satellite-derived absorbed PAR was calculated using the fraction of absorbed PAR and solar radiation data from the ISLSCP Initiative I data collection. This data set is derived independently from the Initiative II Rooting Depth data set (41) [Schenk and Jackson, 2002]; the relationship of the values of these two data sets has yet to be explored.

2.6. Snow, Sea Ice, and Oceans Data

[38] Although the focus of ISLSCP Initiative II was land, sea ice and oceans data sets are critical because they are an indicator of the state of the Earth's climate system. Because snow and ice surfaces represent exceptionally high albedo, with associated effects on surface energy exchange, a snow cover over land data set provided by the National Snow and Ice Data Center is included in Initiative II (29). This time series is also important because fluctuations in snow and ice extent are considered important indicators of climate change [Cavalieri *et al.*, 1997]. As well, snow/sea ice and SST are key variables in the coupling between the atmosphere and the ocean. Accurate knowledge of these variables is essential for climate monitoring, prediction and research. They are also key surface boundary conditions for numerical weather and climate prediction and for other atmospheric simulations using atmospheric general circulation models and regional models.

[39] The ISLSCP Initiative II snow and sea ice data are a subset of the NSIDC Northern Hemisphere EASE-Grid Weekly Snow Cover and Sea Ice Extent product [Armstrong and Brodzik, 2001] that combines snow cover and sea ice extent at weekly intervals for October 1978 through June 2001, and snow cover alone from 1966 through June 2001 (Sea ice data were not available prior to 23 October 1978.) The original data set was the first representation of combined snow and sea ice measurements derived from satellite observations for the period of record (October 1978 to June 2001). Designed to facilitate study of Northern Hemisphere seasonal fluctuations of snow cover and sea ice extent, the original NSIDC data set also includes monthly climatolo-

gies describing average extent, probability of occurrence, and variance. The Initiative II data set shows the extent of snow on the land at a variety of scales (1° , $1/2^\circ$ and $1/4^\circ$).

[40] Global sea ice extent (28) is based on the GSFC Sea Ice Concentrations from Nimbus-7 Scanning Multichannel Microwave Radiometer (SMMR) and the Defense Meteorological Satellites Program (DMSP) Special Sensor Microwave/Imager (SSM/I) Passive Microwave Data. These original data were regridded by NSIDC for ISLSCP Initiative II from their original 25 km spatial resolution and EASE-Grid into equal angle Earth grids with 1° , $1/2^\circ$ and $1/4^\circ$ spatial resolutions.

[41] In addition to its importance to climate modeling, the sea surface temperature data set (30) is also important in gas exchange between the ocean and atmosphere, including the air-sea flux of carbon. Gridded SST products have been developed to satisfy these needs. Gridded monthly and weekly sea surface temperature (SST) and long-term SST monthly climatology for the period 1971–2000 are provided in the Initiative II collection. Weekly normalized error variance fields are also provided. The data are derived using the National Oceanic and Atmospheric Administration (NOAA) Optimum Interpolation (OI) Version 2 (OIv2) global sea surface temperature analyses that use 7 days of in situ (ship and buoy) and satellite SST observations and SST values derived from sea ice concentration [Reynolds *et al.*, 2002]. These analyses are produced weekly using optimum interpolation on a 1° grid and are widely used for many climate modeling and weather forecasting studies.

3. Initiative II Multiple Data Series: Evaluation and Comparison

[42] As discussed above, there are multiple data series available for some variables, for example, there are five different products for albedo (33 through 38) and land cover (40, 45, 46, 48 and 49), three products for vegetation biophysical parameters and NDVI (42,43 and 44), two products for near-surface meteorology fields (22 and 23), and five products for precipitation fields (13, 14, 15, 16 and 24). In this section we will briefly describe the individual data sets, the rationale for their inclusion, contrast their generation methodologies and compare the multiple data series and provide guidance in their selection and use.

4. Land Cover Type

[43] This section provides a brief overview of the various Initiative II land cover data sets and provides insight into and compares their individual characteristics. For more in depth comparisons of the satellite-derived data sets see Brown *de Colstoun et al.* [2006] and Hansen and Reed [2000].

[44] The producers of the Initiative II historical landcover data sets (45 and 46) noted improving agreement between them in later time periods (1850 to 1990) as a result of improved input data [Klein Goldewijk and Ramankutty, 2004]. They also found improving agreement with aggregation of their data to coarser resolution (e.g., 2, 4 and 6°) as a result of spatial smoothing. They attributed differences in their products to differences in the input data used as well as differences in classification methods (i.e., fractional cropland cover versus discrete croplands/pasture classes).

[45] The satellite-based land cover type data sets for Initiative II (48, 50 and 51) and their characteristics are contained in Table 2. The various data sets have different input, processing techniques and classification algorithms [see also Loveland and Belward, 1997; Hansen *et al.*, 2000; Friedl *et al.*, 2002; DeFries *et al.*, 2000]. In comparison to the Initiative I 2-year land cover data, the Initiative II 10-year series is generated with improved classification algorithms, input data and spatial resolution. In addition, the Initiative II land cover product now provides the user with subgrid variability statistics and permits scaling among different spatial resolutions.

[46] The land cover type taxonomy also differs among the products. Both the IGBP-DIScover and MODIS land cover products use the 17-type taxonomy proposed by the IGBP, while the University of Maryland (UMD) data sets uses a 14-type version of the IGBP scheme (see Table 3). In contrast to the more general purposes motivating the IGBP data set, both the Simple Biosphere (SiB2) and Biosphere Atmosphere Transfer Scheme (BATS) taxonomies were generated to provide surface boundary conditions for surface-vegetation-atmosphere models within and out of GCM's [Sellers *et al.*, 1996a, 1996b, 1996c; Dickinson *et al.*, 1986; Dorman and Sellers, 1989].

[47] The AVHRR-based data sets (UMD land cover and continuous fields, IGBP-DIScover) were produced at a native 1 km spatial resolution from a 1 km global AVHRR data set for 1992–1993 [Eidenshink and Faundeen, 1994] from which the Initiative II aggregated products at $1/4$, $1/2$ and 1° were produced. Only the label of the dominant area fraction for each grid cell is provided in the IGBP-DIScover and MODIS aggregated products, while the UMD land cover product takes into account the woody-cover composition of each cell as well as the fraction of each class in the cell.

[48] The IGBP-DIScover product is the only available global product that has been validated against an independent set of high-resolution data [Scepan, 1999], obtaining an overall accuracy of $\sim 70\%$. A more limited evaluation of the UMD and MODIS land cover products has been conducted using training data only [Hansen *et al.*, 2000] showing a spectral separability of $\sim 70\%$. Classification accuracies are likely to be lower when computed using independent test data. In addition to mean accuracies, the MODIS land cover product provides gridded estimates of classifier confidence for each cell, to provide users uncertainty estimates. The IGBP-DIScover data set provides three different taxonomies [Townshend *et al.*, 1994] to address the different input requirements of the IGBP, SiB and BATS. In addition, Initiative II also includes a new product (40), continuous fields of vegetation cover, i.e., the % tree, grass and bare cover of each cell, and the % leaf type and/or leaf longevity for tree canopies. Each land cover data set was processed to a common land/water mask by the ISLSCP staff.

[49] Hansen and Reed [2000] and Brown *de Colstoun et al.* [2006] have compared the UMD and IGBP-DIScover classifications, which are derived from the common 1992–1993 AVHRR data set. Hansen and Reed [2000] found that for broad classes such as forest/woodland, grass/shrubs, crops, etc, the per pixel agreement at 1 km resolution was 74%, decreasing to 48% when all common classes listed in

Table 2. Primary Characteristics of the ISLSCP Initiative II Land Cover Products

Product Characteristics	IGBP-DIScover (51)	UMD (50)	MODIS (48)
Sensor	AVHRR	AVHRR	MODIS
Native resolution	1 km	1 km	1 km
Coverage period	Apr 1992 to Mar 1993	Apr 1992 to Mar 1993	Oct 2000 to Oct 2001
Input data	12 monthly NDVI composites	41 annual metrics from NDVI and AVHRR ch 1–5	monthly values for seven MODIS land bands and EVI; annual minimum, maximum mean for each of the above
Classification scheme	IGBP (17 classes), SiB (15 cl.) and BATS (20 cl.)	modified IGBP (14 classes)	IGBP (17 classes)
Classification technique	unsupervised clustering (ISOCLUS)	supervised classification tree (SPLUS)	supervised decision tree (modified C4.5)
Processing sequence	by continent	global	by continent
Validation	yes (~67%)	only internal (~70%)	being produced
Aggregation method	dominant type	modified dominant type	dominant type

Table 3 were included. They found that in general the IGBP-DIScover had more areas of all forest types while the UMD data set showed more areas with intermediate tree cover such as woody savannas and savannas (i.e., woodlands/wooded grasslands). They also found that the overall agreement between IGBP-DIScover and UMD was much greater (~80% to ~68%) at 1/2° resolution than the agreement of two well-known nonsatellite land cover maps [Olson *et al.*, 1983; Matthews, 1985] which have been extensively used in the past for modeling studies. On the other hand, Brown de Colstoun *et al.* [2006] analyzed the effects of the aggregation methods on the agreement of the two data sets and used the new data layers available in Initiative II to assess the areas of disagreement. They noted that when using a strictly dominant class criteria to label a pixel, the agreement between the two data sets increased with coarser resolution from 48% at 1 km to ~52% at 1°. In contrast, when comparing the IGBP-DIScover data set with the UMD data set using a modified aggregation scheme, they found that the agreement actually decreased with increasing spatial resolution, from 48% to 45.6%, clearly indicating the dependence of the product on the algorithm used. They note that again the areas of disagreement are between similar classes such as the various forest types, open and closed shrublands, etc., and not between large core classes. While areal proportions of cover types showed the same trends as Hansen and Reed [2000], Brown de

Colstoun *et al.* [2006] found that the principal areas of disagreement were specifically related to the IGBP-DIScover Mixed Forest class and the UMD Woodlands/Wooded Grasslands classes, particularly in Africa and boreal forest areas. Using the per-class proportions for each cell as well as the UMD continuous fields data, they show that this disagreement is amplified because of the discrete nature of the classes and is in fact not as large when considering the percent tree cover of each of the classes. Nonetheless, the UMD land cover and continuous fields products do show some inconsistencies in tree cover across certain classes such as Needleleaf forests, pointing to the need for a consistent approach to be applied in the production of both data sets in the future.

[50] The ISLSCP Initiative II collection provides a suite of land cover data sets that represent a significant improvement to the data available in Initiative I. Users need to be aware, however, of the following important recommendations regarding this land cover suite:

[51] 1. In general, users should not difference the MODIS (48) and the AVHRR data sets (50 or 51) to derive land cover change. There may simply be too many methodological-based differences in the data sets.

[52] 2. Users requiring classes such as Permanent Ice and/or Wetlands should use the IGBP-DIScover data set. In fact they may be able to apply the IGBP-DIScover ice class to the UMD data set (50) if desired.

Table 3. Global Land Cover Type Taxonomies Provided in ISLSCP Initiative II

IGBP	UMD	SiB	BATS
evergreen needleleaf forests	evergreen needleleaf forests	evergreen needleleaf trees	evergreen needleleaf trees
evergreen broadleaf forests	evergreen broadleaf forests	broadleaf evergreen trees	evergreen broadleaf tree
deciduous needleleaf forests	deciduous needleleaf forests	needleleaf deciduous trees	deciduous needleleaf trees
deciduous broadleaf forests	deciduous broadleaf forests	broadleaf deciduous trees	deciduous broadleaf trees
mixed forests	mixed forests	broadleaf and needleleaf trees	mixed forest
closed shrublands	closed bushlands or shrublands	broadleaf shrubs with perennial ground cover	evergreen shrubs, deciduous shrubs
open shrublands	open shrublands	broadleaf shrubs with bare soil	evergreen shrubs, deciduous shrubs
woody savannas	woodlands	broadleaf trees with groundcover	interrupted woodlands
savannas	wooded grasslands/shrublands	broadleaf trees with groundcover	interrupted woodlands
grasslands	grasslands	groundcover only	short grass, tall grass
permanent wetlands		persistent wetlands ^a	bog and marshes
croplands	croplands	winter wheat and broadleaf trees	irrigated crops, crops, mixed farming
urban and built-up	urban and built-up	urban and built-up	
cropland/natural vegetation mosaic		dry coastal regions	forest/field mosaic
snow/ice		perpetual ice	ice caps/glaciers
barren/sparsely vegetated	barren	bare soil	semidesert, desert
water bodies	water bodies	water bodies	inland water, ocean
		dwarf trees and shrubs with groundcover	tundra
			water and land mixtures

^aThis class is included in ISLSCP II data sets but is not part of original SiB scheme.

[53] 3. The aggregation method used for the UMD data set is in all likelihood more robust than a strictly dominant type. Users should always consult the subcell makeup of the dominant type and are encouraged to use these data layers to create products which may better suit their needs.

[54] 4. Users needing multiple classification schemes are encouraged to use the IGBP-DIScover data set as it is provided in three different schemes. Again users may be able to create their own scheme through the use of the subcell characteristics layers.

[55] 5. Finally, users should consider integrating the continuous fields information in data set 40 into their analyses as an independent source of information.

5. NDVI and Biophysical Parameters

[56] Two Normalized Difference Vegetation Index (NDVI) data sets were provided to ISLSCP Initiative II: (1) data set 42 and 43, the Fourier-Adjusted, Sensor and Solar zenith angle corrected, Interpolated, Reconstructed (FASIR) monthly time series 1981–1998 (described by *Hall et al.* [2006]), and (2) data set 44, the Global Inventory Modeling and Mapping Studies (GIMMS) monthly time series spanning the 1981 to 2002 period [*Pinzon et al.*, 2006]. Biophysical parameters are also derived from FASIR NDVI and are included as part of the Initiative II collection. Key aspects of each algorithm and their differences are highlighted below in Table 4, are summarized immediately below, and are discussed in depth by *Hall et al.* [2006].

[57] The AVHRR raw data used for GIMMS and FASIR are somewhat different. Both used maximum NDVI composited data to reduce atmospheric and cloud contamination. However, FASIR used the cloud-screened Pathfinder AVHRR bands 1 and 2 series of *James and Kalluri* [1994], whereas GIMMS began with the NOAA/NCAR top of atmosphere (TOA) 15-day data series. GIMMS used NOAA 9 data to fill a 4-month NOAA 11 gap (September 1994 to January 1995) while FASIR extrapolated the NDVI record to fill the gap. The processing approaches differ considerably. To produce surface reflectance data corrected for orbital drift over the years, FASIR applied calibration, bidirectional reflectance function (BRF) and atmospheric corrections (no water vapor) individually to bands 1 and 2 of the cloud-screened Pathfinder AVHRR series of *James and Kalluri* [1994]. To further reduce snow and cloud contamination, Fourier filtering was applied to the NDVI time series and in the tropics spatial aggregation to further mitigate water vapor and cloud contamination. The GIMMS processing approach did not utilize atmospheric correction, except for volcanic stratospheric aerosol following the El Chichon and Mt. Pinatubo eruptions [*Rosen and Kjome*, 1994], and applied corrections to NDVI directly (i.e., did not attempt to correct individual bands). GIMMS used the NOAA thermal band for cloud screening, did not use Fourier filtering to reduce snow and cloud effects and did not use spatial aggregation in the tropics. Hence tropical cloud contamination may be more problematic. GIMMS adjusted the NDVI record for the effects of varying solar illumination angle utilizing the empirical mode decomposition technique [*Huang et al.*, 1998, 1999].

[58] *Hall et al.* [2006] compare and evaluate the FASIR and GIMMS products and reach four important conclusions,

summarized below, that could impact their use in carbon, water and energy cycle analyses.

[59] 1. Neither FASIR nor GIMMS NDVI can be considered absolutely calibrated or completely atmospherically corrected. Both GIMMS and FASIR use vicarious calibrations; however GIMMS NDVI is a top of the atmosphere product, corrected only for stratospheric aerosols from Pinatubo and El Chichon. GIMMS is empirically corrected for variations in the time of NOAA satellite overpass (hence solar illumination angle) over the 22-year interval. FASIR NDVI, AVHRR bands 1 and 2 are more nearly corrected to nadir-looking surface reflectance, however there are no explicit corrections for water vapor or tropospheric aerosols, although the Fourier filtering used in FASIR may mitigate water vapor effects on NDVI.

[60] 2. Because neither FASIR nor GIMMS products are completely corrected to surface reflectance, neither NDVI record should be used in an absolute sense for carbon, water, energy or climate analyses. Rather, NDVI anomalies (i.e., monthly or annual average NDVI subtracted out) should be used for comparisons. Absolute NDVI differences between the FASIR and GIMMS records are large in magnitude and geographically widespread. However, the NDVI data sets can be compared within the context of process models when the NDVI is scaled to F_{par} on the basis of observed minimum and maximum NDVI values within biomes for each vegetation type.

[61] 3. FASIR and GIMMS NDVI anomaly records generally agree, particularly for the last decade. However, significant exceptions exist. In 1984, for example, the two records differ in their global NDVI anomaly by as much as 0.02 or 20% of their range. These differences are likely to be significant in terms of their implied impacts on global anomalies in carbon, water and energy budgets.

[62] 4. Neither NDVI record explains the interannual and spatial variability in the observed atmospheric CO_2 record. Biologic CO_2 fluxes predicted using FASIR or GIMMS NDVI as inputs to a biogeochemical model show no correlation with fluxes derived from atmospheric inversion studies. This suggests that factors other than $fPAR$ drive the larger interannual variations in CO_2 flux on a global basis.

6. Albedo Products

[63] The Initiative II data collection currently contains five albedo data sets (33 through 38) containing several types of albedo parameters, snow-free albedo, broadband albedo, clear-sky albedo, and white-sky albedo. Albedo is in the simplest terms the ratio of energy reflected by a surface to that incident upon it at a given wavelength. Broadband albedo is the average albedo across a wavelength interval or band, typically the solar spectrum (0.3–5 μm) for example. Clear-sky albedo is the fraction of incident direct sunlight reflected by a surface, while white-sky albedo is the fraction of reflected incident diffuse radiation. Snow-free albedo is the albedo of a surface free from snow cover. Snow-free albedo is often used in GCM models to compute snow-on albedo by modifying the snow-free albedo to account for changes due to forecast snow cover. Satellite-measured snow-on albedo including snow cover can be used for validation of the GCM estimated snow-on albedo.

Table 4. FASIR and GIMMS Processing Similarities/Differences

	Data Input	Calibration	View/Illumination Angle Corrections	Atmospheric Corrections	Aerosol Corrections	Other
FASIR monthly band 1, band 2, NDVI	NOAA Pathfinder* data set [James and Kalluri, 1994]; 10-day composites; corrected for mol scattering; NOAA 11 gap Sep 1994 to Jan 1995 filled by interpolation	*after James and Kalluri [1994]; additional adjustments using desert calibration targets	Li sparse and Ross thick kernel [Wanner et al., 1995]	molecular and ozone as per James and Kalluri [1994]; no water vapor correction; however, the Fourier adjustment provides some mitigation	volcanic aerosol corrected El Chichon Apr 1982 to Dec 1984; Pinatubo Jun 1991 to Dec 1994 [Sato et al., 1993]	CLAVR cloud screen; gap-filled cloudy and missing data
GIMMS monthly NDVI only	NOAA/NCAR top of atmosphere 15-day composites! NOAA 11 gap Sep 1994 to Jan 1995 filled with NOAA 9	Verote and Kaufman [1995], desert calibration; AVHRR NDVI is adjusted higher to match 1997–2003 SPOT NDVI	empirical mode decomposition	no explicit correction; indirect correction results from compositing and scaling to match SPOT NDVI values	volcanic aerosol corrected Pinatubo Apr 1982 to Dec 1984; El Chichon Jun 1991 to Dec 1994 [Sato et al., 1993; Verote et al., 1997]	cloud screen using AVHRR thermal band

[64] Data set 33 [Sellers et al., 1996a, 1996b] is the Initiative II standard snow-free albedo product and was generated to be compatible with the other Initiative II data sets. The remaining four albedo data sets (34 through 38) are included in the collection mainly for comparison and validation.

[65] Data set 33 is a monthly mean snow-free surface albedo spanning 1982 to 1999. It is derived from the FASIR biophysical parameters fields of data set 42. The monthly mean albedo is an average over time of the instantaneous albedo, a function of the properties of the land surface and the solar zenith angle, weighted by the incident radiation; the incident radiation was provided by running the Colorado State University (CSU) General Circulation Model (GCM) [Randall et al., 1996] using the atmospheric radiation parameterization of Harshvardhan et al. [1987].

[66] Data set 34 is an Earth Radiation Budget Experiment (ERBE) clear-sky albedo based on the analysis of scanning radiometer instruments on ERBE [Barkstrom, 1984]. It contains global, top of atmosphere, clear sky albedo from January 1986 to February 1990. It was generated at 2.5° spatial resolution, but for compatibility was subsequently regridded to a 1° spatial resolution by the Initiative II staff. Both the original data at 2.5° resolution and the 1° data set are provided.

[67] Set 36, derived from AVHRR channel 1 and channel 2 reflectance is a 5-year (April 1985 to December 1987 and January 1989 to March 1991) NOAA snow-free albedo data set [Csiszar and Gutman, 1999]. It contains average monthly data and was generated as a monthly climatology for use in GCMs at the National Centers for Environmental Prediction (NCEP). The data set is compatible in temporal coverage and spatial resolution with a monthly climatology of green vegetation fraction [Gutman and Ignatov, 1998] currently in use at NCEP. The monthly means of clear-sky, surface, broadband, snow-free albedo correspond to an overhead sun illumination angle.

[68] Data set 37, also derived from AVHRR channel 1 and channel 2 reflectance [Strugnell et al., 2001; Strugnell and Lucht, 2001], provides clear sky surface albedo and BRDF model parameters for two months in 1995 (representing the Northern Hemisphere winter and summer). Three parameters, BRDF, white-sky albedo and black-sky albedo at local solar noon are generated for three broad bands. These parameters can be linearly combined as a function of the fraction of diffuse skylight (itself a function of optical depth) to provide an actual or instantaneous albedo at local solar noon.

[69] Data set 38, the MODIS BRDF/Albedo Product (MOD43B), provides measures of clear sky surface albedo every 16 days [Lucht et al., 2000; Schaaf et al., 2002]. Both white-sky albedo and black-sky albedo at local solar noon are provided for seven spectral bands and three broad bands. Data set 35, a gap-filled version of data set 38 [Moody et al., 2004] is to be provided in the final Initiative II online collection.

7. Precipitation Products

[70] Precipitation is a discontinuous atmospheric variable that can be generated at a large range of geographic and temporal scales and has significant spectral power at all these scales, from instantaneous and local to decadal and global. Precipitation has only nonnegative values and there-

fore its statistics are different from those of other atmospheric variables.

[71] ISLSCP Initiative II has collected a number of precipitation data sets that draw on very different data sources, analysis techniques and spatial and temporal coverage. Many of these data sets have a period of record that extends well beyond the Initiative II decade, and users can obtain longer series from the original providers (see documentation for individual data sets available at <http://www.daac.ornl.gov>).

[72] For many nonexpert users, the GPCP Satellite Gauge (SG) [Adler *et al.*, 2003] provides the all-around best single monthly precipitation data set. Over land, which is the primary focus of ISLSCP Initiative II, the GPCP SG [Xie *et al.*, 2003; Huffman *et al.*, 2001] consists of a standard gauge analysis with climatological bias correction, in combination with a community-based satellite-only product to improve estimates where gauges are sparse. Furthermore, the GPCP SG provides a seamless transition to that satellite-only product alone over the oceans. Finally the GPCP SG is globally complete, albeit with reduced confidence at high latitudes.

[73] Users are urged to consult the ancillary data for the various data sets to help determine the applicability of any particular data set to their needs. In general, fewer samples in a grid box indicate higher uncertainty. Note that the “error” ancillary field, when available, is “random error.” None of the data sets contain bias error estimates. Validation is extremely challenging because few independent, sufficiently dense collections of gauges exist to provide the necessary ground truth. Comparisons to alternative data, such as stream flow, can provide insight into the consistency of the precipitation with other parts of the hydrologic cycle.

7.1. Gauge Analyses

[74] Point observations of accumulated precipitation clearly define “precipitation at the Earth’s surface.” In addition, gauge data provide the longest period of record, at least at certain locations. However, gauge measurements suffer a number of technical issues, which generally result in a low bias because of “undercatch.” This bias primarily depends on the aerodynamics of hydrometeors falling in a wind field in the vicinity of the gauge’s orifice. Higher winds and more slowly falling hydrometeors, such as snow and drizzle, induce a worse bias, sometimes causing a shortfall that is more than 50% of the true precipitation amount. Sevruk [1989] provides one review of these issues.

[75] When the point gauge measurements are transformed to gridded area averages, such as in the ISLSCP II data sets, additional problems arise. First, for many global areas there are not enough gauges available to accurately represent the true area average. Worse, the gauge sites are biased toward developed areas. Data are almost totally lacking over oceans, but sampling is also poor for mountains, deserts, and areas suffering societal upheavals. In mountainous areas precipitation amounts are typically greater at higher elevations, so straight interpolation among the available gauges, which are mostly located in valleys, will usually result in systematic underestimates. Neither gauges data set in the Initiative II collection accounts for this problem.

[76] The instrumentation and reporting methods used in recording gauge data are highly inhomogeneous, making it

difficult to ensure uniform quality. One facet of this issue is that researchers generally cannot completely correct the undercatch bias. As a first step, users are advised to apply the Legates [1987] climatological undercatch corrections that are supplied as part of the Initiative II data, particularly if they are working in areas that experience snow. Note that the GPCP SG incorporates these climatological bias corrections.

7.2. GPCC

[77] The GPCC [Rudolf, 1993] has developed extensive bilateral agreements with data providers around the world to obtain data not usually transmitted to public archives, many of which fill holes and thin spots in the publicly archived network of stations. Furthermore, the GPCC has developed and applied a rigorous quality control system. There is a first, automated step, then a manual inspection that includes integration of qualitative reports of extreme events. As a by-product, this technique has allowed the GPCC to discover and correct numerous metadata errors. The month-to-month GPCC analysis is based on the total precipitation at each station, and there is some concern that this may smooth strong climatic gradients in precipitation in data-sparse regions. At a minimum, interpolation across long distances can yield suspect values in data-sparse regions.

7.3. CRU

[78] The CRU precipitation data are part of a coordinated analysis of several variables spanning the 20th century. As such, it is easy to compare different kinds of data well beyond the Initiative II decade. The CRU analysis scheme separately analyzes the mean (monthly) climate for the period 1961–1990 and the monthly anomalies from that climatology expressed as percentage departures. It is argued that this approach may well preserve mean climatic gradients better than GPCC’s approach in data-sparse regions. In contrast to the GPCC, CRU inserts synthetic zero anomalies in large data voids, forcing the analysis to converge to the (separately determined) climatology. Tests indicate that this step may unrealistically restrict variability in regions that are persistently data-void. CRU screens the precipitation data with an automated quality control for reasonableness, but does not quality control the metadata. The CRU database shows a strong decline in number of stations over the Initiative II decade. Much of this decrease is in areas of initially dense data, but there are also drop outs in data-sparse regions.

7.4. Satellite-Gauge Combinations

[79] ISLSCP Initiative II provides a temporal hierarchy of three data sets that contain combinations of input satellite and gauge data sets. These have the advantage over gauge analyses of providing quasi-global coverage at relatively fine space/time resolution. The combination schemes provided here are designed to minimize known biases. In particular, gauge analyses are incorporated where available, which is mostly over land. A final advantage is that remote sensing data are intrinsically area-averaged, unlike gauges.

[80] On the other hand, the errors in the remote sensing algorithms are only partially characterized, particularly the biases. This arises from the general lack of independent validation data across the full range of climate zones for which we must make estimates. As a result, the combination

estimates will continue to suffer some effects of heterogeneity in space and time in the complement of remote sensing instruments. In particular, the higher-quality data from low-Earth-orbit passive microwave sensors are relatively sparse, while the lower-quality infrared (IR) estimates are plentiful. Microwave estimates are unavailable over frozen surfaces, so wintertime land and polar combination estimates will systematically have lower reliability. Note that gauges also have lower reliability in those regions because of the undercatch bias that affects both the gauge analyses (above) and the satellite gauge combinations.

7.5. GPCP SG (Monthly)

[81] The GPCP SG employs a third-generation, community-based combination algorithm to generate a globally complete monthly estimate. Users should be aware that the successive calibration of the IR by microwave and then gauge, which is designed to take advantage of the bias characteristics of each, also has the effect of forcing the bias to resemble the bias of the last available calibrator. Over land, the GPCP SG bias is typically close to the gauge's, and otherwise at low and midlatitudes it is close to the microwave bias. There is a major data source boundary in 1987: microwave data are not available before July 1987 and for December 1987. The bias characteristics between microwave and nonmicrowave months should be similar (by construction), but the small-scale spatial variance is smaller in the nonmicrowave months.

7.6. GPCP Pentad

[82] The GPCP provides a 5-day (pentad) product over the latitude band 40°N-S. The pentad-to-pentad precipitation values are primarily driven by the IR estimates, with some input from microwave and gauge data. The pentad values are scaled to approximately add up to the GPCP SG for each month at each grid box separately. To the extent that short-period IR estimates strongly contribute to this product, the accuracy will be systematically less than for the monthly SG.

7.7. Numerical-Model-Based Estimates

[83] There are two global precipitation products from the two reanalyses, ERA40 and NCEP2. These data have the advantage of providing global coverage at relatively fine space and time resolution and improved consistency with the other reanalysis fields, both dynamic and thermodynamic, which are constrained by the input observations. Data-sparse regions benefit from information that was inserted into the system at an earlier time "upstream" of the given region. Precipitation and the other surface fluxes are computed from short-term integrations of the model. At middle and high latitudes, where synoptic-scale forcing dominates, the sequences of precipitation events estimated by the reanalyses tend to have reasonable skill, while at lower latitudes the convectively driven regimes show significant departures from most of the observational data sets. The validity of the diurnal cycle in tropical land regions is particularly open to question. Precipitation in the reanalyses is most affected by the spin-up of the dynamic fields, and the 24–36 hour forecast precipitation fields for both ERA40 and NCEP2 are believed to be the best available (W. Ebisuzaki pers. comm.). The reanalyses have significant biases when com-

pared with the observationally based GPCP data set [Betts *et al.*, 2006]. Over the tropical oceans both reanalyses have more rainfall than GPCP, with ERA40 greater than NCEP2. Roads [2003] discusses the high bias of the NCEP2 reanalysis with respect to the Tropical Rainfall Measuring Mission (TRMM) satellite precipitation. The high bias of tropical precipitation in the ERA40 reanalysis stems from a problem with the use of satellite radiances in the analysis of humidity [Troccoli and Källberg, 2004]. ERA40 also has a negative bias over the Amazon in the boreal winter. For NCEP2, the biases over the tropics are smaller than in ERA40. In midlatitudes, the NCEP2 biases from GPCP are generally positive over the oceans in the winter hemisphere, negative over the oceans in the summer hemisphere, and positive over the summer continents. The corresponding midlatitude biases of ERA40 from GPCP are generally smaller. The difference fields between NCEP2 and ERA40 show that NCEP2 has generally more precipitation over the summer continents, and less over the tropical oceans; where there are also differences in the location and width of the convergence zones in the two reanalyses. Over Africa in the boreal summer, the ITCZ precipitation in both reanalyses does not extend as far north as in the GPCP analysis. ERA40 has a known error in the diurnal cycle of precipitation over land (a bias toward precipitation too early in the day) that is larger in the tropics [Betts and Jakob, 2002] than the midlatitudes. Despite the differences in their means, the seasonal anomaly patterns for both reanalyses and GPCP are remarkably similar. Generally the anomalies for the higher-resolution ERA40 are a little closer to the GPCP analysis than for NCEP2, which has generally slightly larger anomalies. Precipitation in the reanalyses is entirely a computed field, while the GPCP analysis is derived from a observations. The reanalyses show coherent anomaly patterns in the summer hemispheres with high precipitation associated with cool-wet anomalies and the converse. This suggests that reanalyses have a good representation of the major circulation changes in the atmosphere. Examples are given by Betts *et al.* [2006].

8. ISLSCP Initiative II Usage

[84] The ISLSCP Initiative II data are available at <http://www.daac.ornl.gov>. Over 300 gigabytes of data are immediately available to scientific users and the public.

[85] As can be seen from Figure 5, there was a steady increase in the number of data files downloaded from April 2004 through an Initiative II science and evaluation workshop in May of 2005. The May 2005 workshop, an open workshop, was held to review ISLSCP science investigations by the user community and to obtain a thorough data evaluation prior to publication. The workshop hosted about 50 users of the data who reported their scientific findings as well as problems and issues in accessing or using the data or documentation. These presentations serve as the basis for the articles in this JGR special issue. The problems uncovered in both the data and its documentation were relatively minor and those identified have been corrected.

[86] Web site usage statistics (Figure 5) show a diverse group of users from many domains and from many countries, with the largest number of users from Japan, corporations, and US education. Future plans for the data collection are for the majority of the data sets to be released

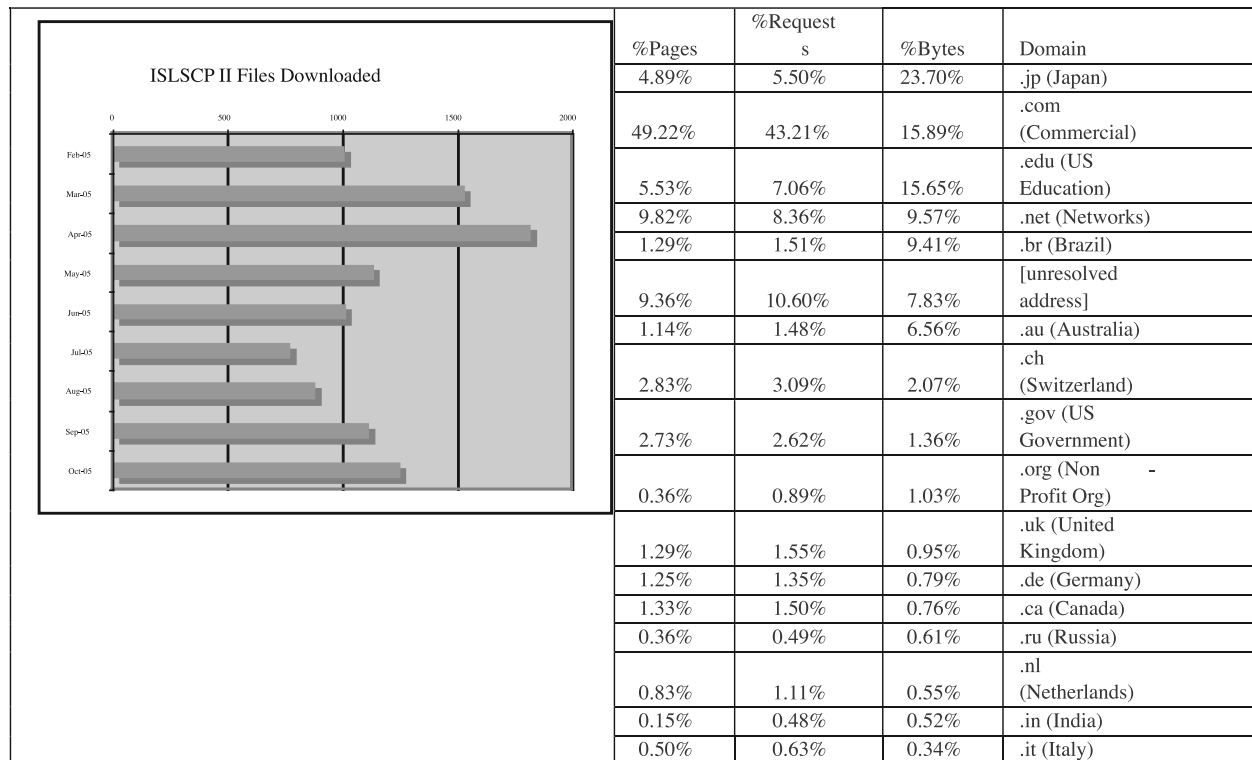


Figure 5. (left) Number of ISLSCP II files downloaded from the combined Initiative II servers. The collection was first made available in April of 2004, and in the first 9 months over 4500 data files were downloaded. The largest spike in usage was leading up to the Initiative II Workshop in May 2005, with over 1800 data files downloaded. In total (as of October 2005), over 15,000 data files have been downloaded. (right) Table showing the various groups that have accessed the Initiative II data collection, users from Japan, commercial groups, and the U.S. educational area account for much of the traffic.

on a set of 4 DVD-ROMs (holding about 16 gigabytes of data) to the public. The largest data sets (which will not fit onto the current DVD media) will continue to be made available through the ORNL DAAC.

9. Conclusions and Future Directions

[87] The ISLSCP Initiative II was an undertaking involving the efforts of 50 to 100 cooperating scientists through teleconferences and twice-yearly meetings. The glue for this diverse community was a science working group (Figure 2) and a low level of funding supporting a small GSFC staff. Given the level of effort from the greater science community (time and travel), it is reasonable to ask, does the Initiative II product justify the effort expended? Secondly, given that there is still a strong need and a demand for integrated, interdisciplinary data collections, how should follow-on efforts be changed to meet those needs?

[88] The first question, the cost-benefit question, has several components: (1) the quality of the data collection, (2) the utility and usability of its implementation and (3) the value of the science that comes from its use. Regarding the quality of the Initiative II collection, while many of the collection's data types would have been produced anyway, new data sets and a significantly improved definition of data sets already in production resulted from the extensive discussions and interactions between the analysis community and data providers in the twice-yearly 3-day ISLSCP

workshops. Secondly, while some of the individual data sets are currently available from the data producers, those in the Initiative II collection have been placed on a common grid at 1/4, 1/2 and 1°, have undergone two peer reviews followed by a careful staff review. This review uncovered a number of problems and issues with many of the data sets that the GSFC-based staff corrected in collaboration with the data providers. This step significantly improved the quality and usability of the data. Third, a common land-sea mask was applied to all data sets and missing grid cells were gap-filled. This common procedure applied to all data sets in the collection makes it significantly more amenable to intercomparison work, with increased consistency among analysis results. In the May 2005 workshop, very few problems were reported, and data accessibility and usability were highly rated.

[89] Regarding the value of the science resulting from the use of the Initiative I and II collections, several projects sponsored by a number of international agencies are leveraged on ISLSCP Initiative II, including the Global Soil Wetness Project (GSWP 2), the Global Land Data Assimilation System (GLDAS), the Global Carbon Observing System (GCOS), NASA Interdisciplinary Science (IDS) projects, funded efforts in NASA's hydrology program, and the NASA seasonal to interannual prediction project (NSIPP). Even in this early stage publications resulting from the recent release of Initiative II are finding their

Table A1. ISLSCP Initiative II Data Collection Summary Table

Data Category	Data Set Title	Author(s) and Originating Institution	Temporal Coverage	Spatial Scale	Data Set Comments
1	air-sea CO ₂ gas exchange	Taro Takahashi, Lamont Doherty Earth Observatory, Columbia University	1995	5 × 4°, 1°	net sea-air CO ₂ flux and sea-air CO ₂ partial pressure (pCO ₂) difference
2	atmospheric CO ₂ consumption by continental erosion	Wolfgang Ludwig, University of Perpignan, France	N/A	1 and 0.5°	related to riverine flux data set
3	CO ₂ emissions from fossil fuel burning	Gregg Marland and Antoinette Brekert, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory (ORNL)	1950–1995	1°	decadal resolution
4	Emission Database for Global Atmospheric Research (EDGAR 3)	Jos Olivier, National Institute of Public Health and the Environment (RIVM), Netherlands	1970–1995	1°	greenhouse gas (CO ₂ , CH ₄ , N ₂ O) and tropospheric ozone precursor gas emissions (CO, NO _x , NMVOC, SO ₂)
5	FLUXNET CO ₂ (compiled from regional networks)	ORNL Distributed Active Archive Center (DAAC)	1992–1995	point data	example gap-filled CO ₂ , water and energy data for Harvard Forest and N. BOREAS old black spruce sites only
6	global riverine fluxes of carbon and sediments to the oceans	Wolfgang Ludwig, University of Perpignan, France	N/A	2.5 × 2°, 1 and 0.5° tabular and ASCII map data	2363 point measurements and associated ancillary data
7	Global Primary Production Data Initiative (GPPDI) net primary production (NPP) class B point data	Richard Olson and Jonathan Scurlock, ORNL; Tom Gower, University of Wisconsin	varies by site	Point data	not complete global coverage
8	GPPDI gridded NPP data	Stephen Prince, University of Maryland; Daolan Zheng, University of Toledo	various	1 and 0.5°	GlobalView CO ₂ 2003 data; smoothed, interpolated and extrapolated data
9	GlobalView: atmospheric CO ₂ concentrations	Ken Masarie, Climate Monitoring and Diagnostics Laboratory (CMDL), NOAA	varies by site	point data	GlobalView CH ₄ 2001 data; smoothed, interpolated and extrapolated data
10	GlobalView: atmospheric methane concentrations	Ken Masarie, CMDL, NOAA	varies by site	point data	mean of 17 NPP models from Potsdam NPP model intercomparison
11	International Geosphere Biosphere Programme (IGBP) global NPP model intercomparison data	Wolfgang Cramer, Potsdam Institute for Climate Impact Research (PIK), Germany	N/A	1 and 0.5°	subcell statistics for elevation, slope, aspect and compound topo. index
12	digital elevation and elevation-based derivatives from HYDRO1k	Kristine Verdin, National Center for EROS (USGS)	N/A	1 and 0.5°	daily data for global land areas from Global Telecommunication Network (GTS)
13	gauge-based daily precipitation	Pingping Xie and John Janowiak Climate Prediction Center, NOAA	1986–1995	1 and 0.5°	GPCC version 2 global gridded monthly gauge product
14	precipitation monthly	Bruno Rudolf, Global Precipitation Climatology Centre (GPCC), Germany	1986–1995	1 and 0.5°	monthly data, error fields and 1979–1999 climatology
15	precipitation monthly (satellite and gauge)	George Huffman and David Bolvin, Global Precipitation Climatology Project (GPCP), NASA	1986–1995	2.5, 1°	GPCP pentad (5-day) data
16	precipitation pentad (satellite and gauge)	Pingping Xie, Climate Prediction Center, NOAA	1986–1995	2.5, 1°	data set based on a CD-ROM developed by the Global Soil Data Task of the IGBP Data and Information Service
17	ISLSCP II global gridded soil characteristics	ORNL DAAC; Robert Scholes, Council for Scientific and Industrial Research, South Africa; Eric Brown de Colstoun (ISLSCP II Staff)	N/A	1°	temporal coverage varies by station
18	river discharge point data	Thomas Maurer, Global Runoff Data Centre (GRDC), Germany	1986–1995	point data	simulated topological network with gridded river basin data and attribute tables
19	river routing data (STN-30p)	Charles Vörösmarty and Balázs Fekete, University of New Hampshire (UNH)	N/A	1 and 0.5°	data set used ISLSCP I forcing data
20	global soil water storage capacity of the rooting zone	Axel Kleidon, University of Maryland	1987–1988 averaged ¹		

Table A1. (continued)

Data Category	Data Set Title	Author(s) and Originating Institution	Temporal Coverage	Spatial Scale	Data Set Comments
21	UNH/GRDC composite monthly runoff	Balázs Fekete and Charles Vörösmarty, UNH	1986–1995	1 and 0.5°	monthly gridded runoff averages
22	European Centre for Medium-range Weather Forecasts (ECMWF) near surface meteorology parameters from the ECWMF reanalysis (ERA40)	Near-surface meteorology Anton Beljaars (ECMWF), Alan Betts, and Eric Brown de Colstoun	1986–1995	1°	subset of monthly average, diurnal, and 3-hourly data from ERA40
23	National Centers for Environmental Predictions (NCEP) reanalysis II meteorology data	Paul Dimmeyer and Mei Zhao, Center for Ocean-Land-Atmosphere Studies (COLA), Glenn White, NCEP	1986–1995	1°	COLA version of the NCEP II reanalysis; monthly average, diurnal, and 3-hourly data
24	Climate Research Unit (CRU) monthly climate time series	Mark New, CRU University of East Anglia, United Kingdom	1986–1995	1 and 0.5°	monthly averages of various climate variables
25	CRU monthly mean climatology (1961–1990)	Mark New, CRU University of East Anglia, United Kingdom	1961–1990 averages	1 and 0.5°	monthly data averaged over the 1961–1990 period
26	International Satellite Cloud Climatology Project (ISCCP) clouds	Radiation and clouds Paul Stackhouse, NASA Langley Research Center	1986–1995	1°	monthly mean, standard deviation, maximum, minimum
27	surface radiation budget (SRB) radiation fields	Paul Stackhouse, NASA Langley Research Center	1986–1995	1°	monthly average, diurnal, and 3-hourly data
28	global sea ice extent	Snow, sea ice, and oceans Richard Armstrong and Ken Knowles, National Snow and Ice Data Center (NSIDC), University of Colorado	1986–1995	1, 0.5, and 0.25°	Tabular data and 1° ASCII maps
29	Northern Hemisphere snow cover extent	Richard Armstrong and Ken Knowles, NSIDC, University of Colorado	1986–1995	1, 0.5, and 0.25°	tabular data and 1° ASCII maps
30	optimally interpolated sea surface temperature (SST)	Richard Reynolds, National Climatic Data Center; Diane Stokes, NCEP, NOAA	1986–1995	1°	monthly and weekly SST analyses and 1971–2000 monthly climatology
31	gridded population of the world	Socioeconomic Gregg Yetman and Deborah Balk, Center for International Earth Science Information Network (CIESIN), Columbia University	1990, 1995	1, 0.5, and 0.25°	gridded population counts and density; Socioeconomic Data and Application Center (SEDAC) data set
32	Global Gridded Gross Domestic Product (GDP)	Gregg Yetman and Deborah Balk, CIESIN, Columbia University	1990	1, 0.5, and 0.25°	SEDAC data set
33	albedo (snow-free)	Vegetation Don Dazlich, Colorado State University	1982–1998	1°	monthly data calculated from FASIR NDVI data set
34	albedo (snow-inclusive)	David Young and Takmeng Wong, NASA Langley Research Center	1986–1990	2.5, 1°	monthly data from Earth Radiation Budget Experiment (ERBE)
36	albedo (snow-free 5-year monthly climatology)	Ivan Csiszar, University of Maryland	1985–1991	Average 1, 0.5, and 0.25°	monthly averages for 1985–1991 period from the AVHRR
37	AVHRR Albedo and Bidirectional Reflectance Distribution Function (BRDF) parameters for 1995	Alan Strahler and Crystal Schaaf, Boston University	1995	1, 0.5, and 0.25°	monthly data for February and July 1995
38	MODIS albedo for 2001	Crystal Schaaf and Alan Strahler, Boston University	2001	1, 0.5, and 0.25°	multispectral, broadband albedo for 16-day periods with quality information
39	C4 vegetation percentage	Chris Still, University of California at Santa Barbara	1996–1998	1°	% of each cell which possesses the C4 photosynthetic pathway
40	continuous fields of vegetation cover	Ruth DeFries, University of Maryland; Matt Hansen, South Dakota State University	1992–1993	1, 0.5, and 0.25°	% tree, grass and bare cover and % needleleaf, broadleaf, deciduous, evergreen for tree cover

Table A1. (continued)

Data Category	Data Set Title	Author(s) and Originating Institution	Temporal Coverage	Spatial Scale	Data Set Comments
41	ecosystem rooting depths	Rob Jackson, Duke University; H. Jochen Schenk, California State University Fullerton	N/A	1°	global maps of mean 50% and 95% rooting depths
42	FASIR biophysical parameter fields	Sietse Los, University of Wales at Swansea, United Kingdom	1982–1998	1, 0.5, and 0.25°	derived from the FASIR-NDVI data set
43	FASIR Normalized Difference Vegetation Index (NDVI) monthly	Sietse Los, University of Wales at Swansea, United Kingdom	1982–1998	1, 0.5, and 0.25°	uses Pathfinder AVHRR DATA
44	Global Inventory Modeling and Mapping Studies (GIMMS) NDVI	Compton Tucker, Jorge Pinzon, and Molly Brown, NASA Goddard Space Flight Center	1981–2002	1, 0.5, and 0.25°	
45	historical croplands fractional cover	Navin Ramankutty and Jonathan Foley, University of Wisconsin	1700–1992	1 and 0.5°	every 50 years (1700–1850); every 10 years (1850–1980); every year (1986–1992)
46	historical land cover and land use	Kees Klein Goldewijk, RIVM, Netherlands	1700–1990	1 and 0.5°	every 50 years (1700–1950); every 10 years (1950–1990)
47	leaf area index (LAI) from field measurements	Jonathan Scurlock (ORNL), ORNL DAAC	1932–2000	point data	1008 worldwide point measurements compiled from the literature
48	MODIS land cover product	Mark Friedl, Alan Strahler, John Hodges, Boston University	2000	1, 0.5, and 0.25°	dominant land cover type, fraction of each cover type and classifier confidence for each cell
49	potential vegetation	Navin Ramankutty and Jonathan Foley, University of Wisconsin	N/A	1 and 0.5°	represents natural vegetation before human alteration
50	UMD land cover classification	Matt Hansen, South Dakota State University; Ruth Defries, University of Maryland	1992–1993	1, 0.5, and 0.25°	dominant land cover type and fraction of each cover type in each cell
51	vegetation classification (IGBP-DIScover)	Tom Loveland and Stephen Howard, National Center for EROS (USGS)	1992–1993	1, 0.5, and 0.25°	dominant type and fraction of each cover type; three classification schemes (IGBP, SIB, BATS)
52	land/water masks, land outline overlays, latitude and longitude grids	Tom Logan, Jet Propulsion Laboratory; ISLSCP II staff	Ancillary N/A	1, 0.5, and 0.25°	binary water masks and fractional water/land cover in each cell

way into the refereed literature from the GSWP and others, including those in this special section. However, if the scientific utilization of the Initiative I collection is any measure, over 13,000 CDs have been ordered from the Goddard DAAC, over 300,000 files downloaded, with over 500 citations in the scientific literature. These articles support a variety of uses, including weather forecast improvements, hydrological applications, macroscale basin modeling and biogeochemical and carbon tracer models. Already as can be seen from Figure 5, actual data downloads from the Initiative II collection are averaging more than 1000 each month by a diverse group of users from many domains and from many countries.

[90] There is no reason to doubt that the production of integrated, interdisciplinary data collections is worthwhile. Without such a collective effort on the part of the science community, in the end each segment of the community has to expend their own resources to produce subsets of these data collections, which then are not only not easily available to the larger community, but suffer in terms of uneven data quality and format, incomplete or missing documentation.

[91] Where do we go from here? The Initiative II 1986–1995 10-year period only begins to span the period of record needed to observe climate trends, seasonal to interannual variations in carbon, water, and energy cycling rates, and to understand and quantify interactions and feedbacks among the land, oceans, and atmosphere. All these are necessary to address the science questions posed in Table 1. In addition, several new data sets coming online need to be captured. Sensors aboard TERRA, AQUA, and TRMM (U.S.), ENVISAT (E.U.) will provide improved information on vegetation, clouds, aerosols, and precipitation. The ISLSCP I and II Initiatives have built a community of modelers and data providers that are working well together. ISLSCP bridges the carbon and water communities, bringing them together frequently. ISLSCP brings participants from major projects together on a frequent basis, for example, from Global LDAS and the Global Soil Wetness Project. Maintaining that momentum is extremely important.

Appendix A: ISLSCP Initiative II Data Collection Summary Table

[92] Table A1 provides a listing of 52 interdisciplinary data sets provided in the International Satellite Land Surface Climatology Project (ISLSCP) Initiative II Data Collection. The entire data collection can be accessed at <http://daac.ornl.gov/>.

[93] **Acknowledgments.** ISLSCP Initiative II was funded in part by the NASA Hydrology and Terrestrial Ecology Program. Eric Wood, Dennis Lettenmaier and Jared Entin were the NASA Hydrology Program Managers. Diane Wickland is the NASA Terrestrial Ecology Program Manager. The authors would also like to thank the ISLSCP Science Working Group (Figure 2) for their generous contributions of time and talent to the monthly teleconferences and the biannual meetings. Finally, thanks to the many data providers in Appendix A and data users who through these meetings defined the detailed data requirements for ISLSCP Initiative II collection and produced the data within the collection.

References

- Adler, R. F., et al. (2003), The Version 2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present), *J. Hydrometeorol.*, 4(6), 1147–1167.
- Armstrong, R. L., and M. J. Brodzik (2001), Recent Northern Hemisphere snow extent: A comparison of data derived from visible and microwave satellite sensors, *Geophys. Res. Lett.*, 28(19), 3673–3676.
- Barkstrom, B. R. (1984), The Earth Radiation Budget Experiment (ERBE), *Bull. Am. Meteorol. Soc.*, 65, 1170–1185.
- Betts, A. K., and C. Jakob (2002), Evaluation of the diurnal cycle of precipitation, surface thermodynamics, and surface fluxes in the ECMWF model using LBA data, *J. Geophys. Res.*, 107(D20), 8045, doi:10.1029/2001JD000427.
- Betts, A. K., M. Zhao, P. A. Dirmeyer, and A. C. M. Beljaars (2006), Comparison of ERA40 and NCEP/DOE near-surface data sets with other ISLSCP-II data sets, *J. Geophys. Res.*, 111, D22S04, doi:10.1029/2006JD007174.
- Brown de Colstoun, E. C., R. S. DeFries, and J. R. G. Townshend (2006), Evaluation of ISLSCP Initiative II satellite-based land cover data sets and assessment of progress in land cover data for global modeling, *J. Geophys. Res.*, doi:10.1029/2006JD007453, in press.
- Cavalieri, D. J., C. L. Parkinson, P. Gloersen, and H. J. Zwally (1997), Arctic and Antarctic sea ice concentrations from multichannel passive-microwave satellite data sets: October 1978 to December 1996, user's guide, *NASA Tech. Memo. 104647*, 17 pp.
- Collatz, G. J., J. A. Berry, and J. S. Clark (1998), Effects of climate and atmospheric CO₂ partial pressure on the global distribution of C4 grasses: Present, past, and future, *Oecologia*, 114, 441–454.
- Csiszar, I., and G. Gutman (1999), Mapping global land surface albedo from NOAA/AVHRR, *J. Geophys. Res.*, 104, 6215–6228.
- DeFries, R., M. Hansen, and J. Townshend (2000), Global continuous fields of vegetation characteristics: A linear mixture model applied to multiyear 8 km AVHRR data, *Int. J. Remote Sens.*, 21, 1389–1414.
- 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 Tech. Note NCAR/TN-275+STR*, Natl. Cent. for Atmos. Res., Boulder, Colo.
- Dorman, J. L., and P. J. Sellers (1989), A global climatology of albedo, roughness length and stomatal resistance for atmospheric general circulation models as represented by the simple biosphere model, *J. Appl. Meteorol.*, 28, 833–855.
- Eidenshink, J. C., and J. L. Faundeen (1994), The 1 km AVHRR global land data set: First stages in implementation, *Int. J. Remote Sens.*, 15, 3443–3462.
- Fekete, B. M., C. J. Vörösmarty, and W. Grabs (2002), High-resolution fields of global runoff combining observed river discharge and simulated water balances, *Global Biogeochem. Cycles*, 16(3), 1042, doi:10.1029/1999GB001254.
- Food and Agriculture Organization (1995), The digital soil map of the world, version 3.5., Rome.
- Friedl, M. A., et al. (2002), Global land cover mapping from MODIS: Algorithms and early results, *Remote Sens. Environ.*, 83, 287–302.
- Gesch, D. B., K. L. Verdin, and S. K. Greenlee (1999), New land surface digital elevation model covers the Earth, *Eos Trans. AGU*, 80(6), 69–70.
- Gupta, S. K., W. L. Darnell, and A. C. Wilber (1992), A parameterization of longwave surface radiation from satellite data: Recent improvements, *J. Appl. Meteorol.*, 31, 1361–1367.
- Gutman, G., and A. Ignatov (1998), Derivation of green vegetation fraction from NOAA/AVHRR for use in numerical weather prediction models, *Int. J. Remote Sens.*, 19, 1533–1543.
- Hall, F. G., J. Masek, and J. G. Collatz (2006), Evaluation of ISLSCP Initiative II FASIR and GIMMS NDVI products and implications for carbon cycle science, *J. Geophys. Res.*, doi:10.1029/2006JD007438, in press.
- Hansen, M., and B. Reed (2000), A comparison of the IGBP-DIScover and the University of Maryland 1 km global land cover products, *Int. J. Remote Sens.*, 21, 1365–1374.
- Hansen, M. C., R. S. DeFries, J. R. G. Townshend, and R. Sohlberg (2000), Global land cover classification at 1 km spatial resolution using a classification tree approach, *Int. J. Remote Sens.*, 21, 1331–1364.
- Harshvardhan, R. Davies, D. A. Randall, and T. G. Corsetti (1987), A fast radiation parameterization for general circulation models, *J. Geophys. Res.*, 92, 1009–1016.
- Huang, N. E., et al. (1998), The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. R. Soc. London*, 545, 903–995.
- Huang, N. E., et al. (1999), A new view of nonlinear water waves: the Hilbert spectrum, *Annu. Rev. Fluid Mech.*, 31, 417–457.
- Huffman, G. J., R. F. Adler, M. Morrissey, D. T. Bolvin, S. Curtis, R. Joyce, B. McGavock, and J. Susskind (2001), Global precipitation at one-degree daily resolution from multi-satellite observations, *J. Hydrometeorol.*, 2(1), 36–50.
- James, M. E., and S. N. V. Kalluri (1994), The Pathfinder Land Data Set: An improved coarse resolution data set for terrestrial monitoring, *Int. J. Remote Sens.*, 15(17), 3347–3363.

- Kalnay, E., et al. (1996), The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc.*, **77**, 437–471.
- Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter (2002), NCEP/DOE AMIP-II reanalysis (R-2), *Bull. Am. Meteorol. Soc.*, **83**, 1631–1643.
- Kistler, R., et al. (2001), The NCEP-NCAR 50 year reanalysis monthly means CD-ROM and documentation, *Bull. Am. Meteorol. Soc.*, **82**, 247–267.
- Kleidon, A., and M. Heimann (1998), A method of determining rooting depth from a terrestrial biosphere model and its impacts on the global water and carbon cycle, *Global Change Biol.*, **4**, 275–286.
- Klein Goldewijk, K. (2001), Estimating global land use change over the past 300 years: The HYDE database, *Global Biogeochem. Cycles*, **15**(2), 417–433.
- Klein Goldewijk, K., and N. Ramankutty (2004), Land cover change over the last three centuries due to human activities: The availability of new global data sets, *Geojournal*, **61**, 335–344.
- Legates, D. R. (1987), A climatology of global precipitation, *Publ. Climatol.*, **40**, 85 pp.
- Loveland, T. R., and A. S. Belward (1997), The IGBP-DIS global 1 km land cover data set, DISCover: first results, *Int. J. Remote Sens.*, **18**, 3289–3295.
- Lucht, W., C. B. Schaaf, and A. H. Strahler (2000), An algorithm for the retrieval of albedo from space using semiempirical BRDF models, *IEEE Trans. Geosci. Remote Sens.*, **38**, 977–998.
- Matthews, E. (1985), Atlas of archived vegetation and land use and seasonal albedo data sets, *NASA Tech. Memo.* 68199, 54 pp.
- Moody, E. G., M. D. King, S. Platnick, C. B. Schaaf, and F. Gao (2004), Spatially complete global spectral surface albedos: Value-added datasets derived from Terra MODIS land products, *IEEE Trans. Geosci. Remote Sens.*, **43**, 144–158.
- New, M., M. Hulme, and P. Jones (1999), Representing twentieth-century space-time climate variability. Part I: Development of a 1961–90 mean monthly terrestrial climatology, *J. Clim.*, **12**, 829–856.
- New, M., M. Hulme, and P. D. Jones (2000), Representing twentieth century space-time climate variability. Part 2: Development of a 1901–96 monthly grids of terrestrial surface climate, *J. Clim.*, **13**, 2217–2238.
- Olson, J. S., J. Watts, and L. Allison (1983), Carbon in live vegetation of major world ecosystems, *Rep. W-7405-ENG-26*, Oak Ridge Natl. Lab., U.S. Dep. of Energy, Oak Ridge, Tenn.
- Pinker, R. T., and I. Laszlo (1992), Modeling surface solar irradiance for satellite applications on a global scale, *J. Appl. Meteorol.*, **31**, 194–211.
- Pinzon, J., M. E. Brown, and C. J. Tucker (2006), Satellite time series correction of orbital drift artifacts using empirical mode decomposition, in *Hilbert-Huang Transform: Introduction and Applications*, edited by N. Huang, chap. 10, in press.
- Ramankutty, N., and J. A. Foley (1998), Characterizing patterns of global land use: An analysis of global croplands data, *Global Biogeochem. Cycles*, **12**(4), 667–686.
- Ramankutty, N., and J. A. Foley (1999), Estimating historical changes in global land cover: Croplands from 1700 to 1992, *Global Biogeochem. Cycles*, **13**(4), 997–1027.
- Randall, D. A., D. A. Dazlich, C. Zhang, A. S. Denning, P. J. Sellers, C. J. Tucker, L. Bounoua, S. O. Los, C. O. Justice, and I. Y. Fung (1996), A revised land surface parameterization (SiB2) for GCMs. Part 3. The greening of the Colorado State University general circulation model, *J. Clim.*, **9**(4), 738–763.
- Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang (2002), An improved in situ and satellite SST analysis for climate, *J. Clim.*, **15**, 1609–1625.
- Roads, J. (2003), The NCEP/NCAR, NCEP/DOE and TRMM tropical atmosphere hydrologic cycles, *J. Hydrometeorol.*, **4**, 826–840.
- Rosen, J. M., and N. T. Kjöme (1994), Decay of Mount Pinatubo aerosol at midlatitudes in the northern and southern hemispheres, *J. Geophys. Res.*, **99**(D12), 25,733–25,739.
- Rossow, W. B., and R. A. Schiffer (1999), Advances in understanding clouds from ISCCP, *Bull. Am. Meteorol. Soc.*, **80**, 2261–2287.
- Rossow, W. B., A. W. Walker, D. E. Beusichel, and M. D. Roiter (1996), International Satellite Cloud Climatology Project (ISCCP): Documentation of new cloud data sets, *WMO/TD-737*, 115 pp., World Meteorol. Organ., Geneva, Switzerland.
- Rudolf, B. (1993), Management and analysis of precipitation data on a routine basis, paper presented at International WMO/IAHS/ETH Symposium on Precipitation and Evaporation, Sloval Hydrometeorol. Inst., Bratislava, Slovakia.
- Sato, M., et al. (1993), Stratospheric aerosol optical depths, 1850–1990, *J. Geophys. Res.*, **98**(D12), 22,987–22,994.
- Scepan, J. (1999), Thematic validation of high-resolution global land-cover data sets, *Photogramm. Eng. Remote Sens.*, **65**, 1051–1060.
- Schaaf, C. B., et al. (2002), First operational BRDF, albedo and nadir reflectance products from MODIS, *Remote Sens. Environ.*, **83**, 135–148.
- Schenk, H. J., and R. B. Jackson (2002), The global biogeography of roots, *Ecol. Monogr.*, **72**, 311–328.
- Sellers, P. J., et al. (1996a), The ISLSCP Initiative I global data sets: Surface boundary Conditions and atmospheric forcings for land-atmosphere studies, *Bull. Am. Meteorol. Soc.*, **77**(9), 1987–2005.
- Sellers, P. J., D. A. Randall, C. J. Collatz, J. A. Berry, C. B. Field, D. A. Dazlich, C. Zhang, C. D. Collelo, and L. Bounoua (1996b), A revised land surface parameterization (SiB2) for atmospheric GCMs. Part 1: Model formulation, *J. Clim.*, **9**, 676–705.
- Sellers, P. J., S. O. Los, C. J. Tucker, C. O. Justice, D. A. Dazlich, G. J. Collatz, and D. A. Randall (1996c), A revised land surface parameterization (SiB2) for atmospheric GCMs. Part 2: The generation of global fields of terrestrial biophysical parameters from satellite data, *J. Clim.*, **9**, 706–737.
- Sevruk, B. (1989), Reliability of precipitation measurements, in *Precipitation Measurements (Proceedings of WMO/IAHS/ETH Workshop)*, edited by B. Sevruk, pp. 13–19, Swiss Fed. Inst. of Technol., Zurich, Switzerland.
- Still, C. J., J. A. Berry, G. J. Collatz, and R. S. DeFries (2003), Global distribution of C3 and C4 vegetation: Carbon cycle implications, *Global Biogeochem. Cycles*, **17**(1), 1006, doi:10.1029/2001GB001807.
- Strugnell, N., and W. Lucht (2001), Continental-scale albedo inferred from AVHRR data, land cover class and field observations of typical BRDFs, *J. Clim.*, **14**, 1360–1376.
- Strugnell, N., W. Lucht, and C. Schaaf (2001), A global albedo data set derived from AVHRR data for use in climate simulations, *Geophys. Res. Lett.*, **28**, 191–194.
- Townshend, J. R. G., C. O. Justice, D. Skole, J.-P. Malingreau, J. Cihlar, P. Teillet, F. Sadowski, and S. Ruttenberg (1994), The 1 km resolution global data set: Needs of the International Geosphere Biosphere Programme, *Int. J. Remote Sens.*, **15**, 3417–3442.
- Troccoli, A., and P. Källberg (2004), Precipitation correction in the ERA-40 reanalyses, *ERA-40 Proj. Rep.* 13, 6 pp., Eur. Cent. for Med.-Range Weather Forecasts, Reading, U. K. (Available at http://www.ecmwf.int/publications/library/ecpublications/pdf/era40/ERA40_PR13.pdf).
- Uppala, S. M., et al. (2005), The ERA-40 reanalysis, *Q. J. R. Meteorol. Soc.*, **131**, 2961–3012.
- Vermote, E., and Y. J. Kaufman (1995), Absolute calibration of AVHRR visible and near-infrared channels using ocean and cloud views, *Int. J. Remote Sens.*, **16**(13), 2317–2340.
- Vermote, E., et al. (1997), Data pre-processing: Stratospheric aerosol perturbing effect on the remote sensing of vegetation: Correction method for the composite NDVI after the Pinatubo Eruption, *Remote Sens. Rev.*, **15**, 7–21.
- Vörösmarty, C. J., B. M. Fekete, M. Meybeck, and R. B. Lammers (2000), Geomorphometric attributes of the global system of rivers at 30-minute spatial resolution, *J. Hydrol.*, **237**, 17–39.
- Wanner, W., X. Li, and A. H. Strahler (1995), On the derivation of kernel-driven models of bidirectional reflectance, *J. Geophys. Res.*, **100**(D10), 21,077–21,090.
- Xie, P. P., et al. (2003), GPCP Pentad precipitation analyses: An experimental dataset based on gauge observations and satellite estimates, *J. Clim.*, **16**(13), 2197–2214.
- Zhang, Y.-C., W. B. Rossow, A. A. Lacis, V. Oinas, and M. I. Mishchenko (2004), Calculation of radiative fluxes from the surface to top of atmosphere based on ISCCP and other global data sets: Refinements of the radiative transfer model and the input data, *J. Geophys. Res.*, **109**, D19105, doi:10.1029/2003JD004457.

A. Betts, Atmospheric Research, RR 2, Box 2602, Pittsford, VT 05763, USA. (akbetts@aol.com)

L. Bounoua and G. J. Collatz, NASA Goddard Space Flight Center, Code 614.4, Greenbelt, MD 20771, USA. (bounoua@ltpmail.gsfc.nasa.gov; jcollatz@ltpmail.gsfc.nasa.gov)

E. Brown de Colstoun, G. J. Huffman, and D. Landis, Science Systems and Applications Inc., NASA Goddard Space Flight Center, Code 614.4, Greenbelt, MD 20771, USA (ericbdc@ltpmail.gsfc.nasa.gov; huffman@agns.gsfc.nasa.gov)

P. Dirmeyer, Center for Ocean-Land-Atmosphere Studies, Powder Mill Road, Suite 302, Calverton, MD 20705, USA. (dirmeyer@cola.iges.org)

F. G. Hall, University of Maryland Baltimore County, NASA Goddard Space Flight Center, Code 614.4, Greenbelt, MD 20771, USA. (fghall@ltpmail.gsfc.nasa.gov)

B. Meeson, Ocean US, 2300 Clarendon Boulevard, Suite 1350, Arlington, VA 22201, USA.