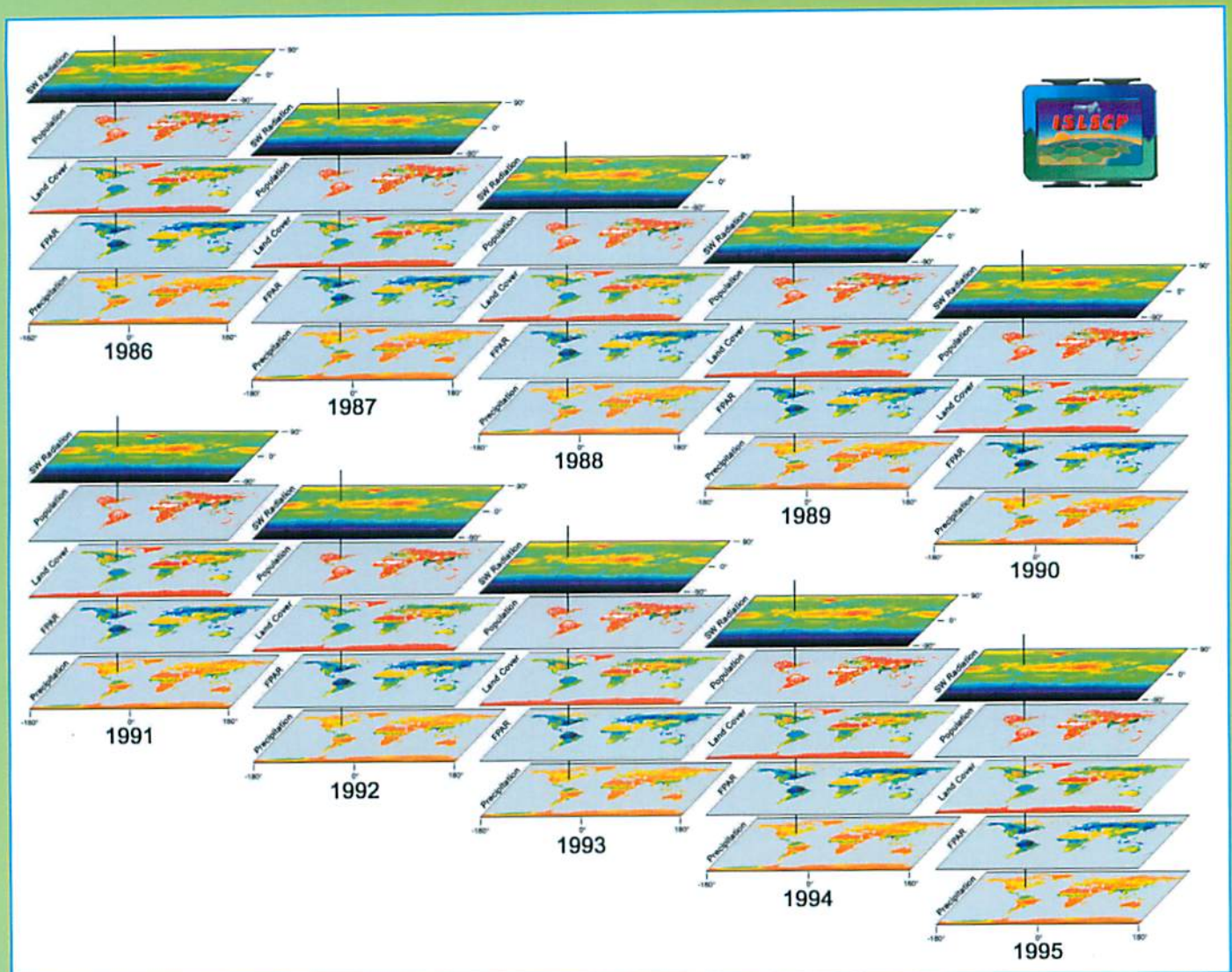


# Carbon, Water, and Energy Cycle Investigations Using a Common Global Data Set: The International Land Surface Climatology Project (ISLSCP) Initiative II



rep. 00262

# **Carbon, Water, and Energy Cycle Investigations Using a Common Global Data Set: The International Land Surface Climatology Project (ISLSCP) Initiative II**



Reprinted from  
*Journal of Geophysical Research*  
Published by AGU

### **Front cover**

The ISLSCP Initiative II data sets 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 focused on quantifying how the Earth is changing and the consequences of those changes for life on Earth. The record consists of 52 data sets, with a common series in the 10-year period from 1985 to 1996. Selected data extend well beyond this period. A small staff at NASA Goddard Space Flight Center coordinated the production, acquisition, peer review, evaluation, rework, and co-registration of each data set to a common grid and land-sea mask.

### **Back cover**

The communities that drove the definition of the ISLSCP Initiative II collection were investigators within the international Global Energy and Water Cycle Experiment (GEWEX), the International Geosphere-Biosphere Programme (IGBP), and the U.S. Global Change Research Program (USGCRP). The ISLSCP Initiative II data collection was the product of over 50 scientists supported by 33 different international organizations. ISLSCP collaborators consisted of both future users of the data, including hydrologists, meteorologists, and ecologists, and the data providers covering a broad range of Earth science disciplines.

Copyright 2007 by AGU  
2000 Florida Avenue, N.W., Washington DC 20009, USA

Figures, tables, and short excerpts may be reprinted in scientific books and journals if the source is properly cited; all other rights reserved.

## **Carbon, Water, and Energy Cycle Investigations Using a Common Global Data Set: The International Land Surface Climatology Project (ISLSCP) Initiative II**

- D22S01** *Forrest G. Hall, Eric Brown de Colstoun, George J. Collatz, David Landis, Paul Dirmeyer, Alan Betts, George J. Huffman, Lahouari Bounoua, and Blanche Meeson*  
ISLSCP Initiative II global data sets: Surface boundary conditions and atmospheric forcings for land-atmosphere studies (DOI 10.1029/2006JD007366)
- D22S02** *Zhichang Guo and Paul A. Dirmeyer*  
Evaluation of the Second Global Soil Wetness Project soil moisture simulations:  
1. Intermodel comparison (DOI 10.1029/2006JD007233)
- D22S03** *Zhichang Guo, Paul A. Dirmeyer, Zeng-Zhen Hu, Xiang Gao, and Mei Zhao*  
Evaluation of the Second Global Soil Wetness Project soil moisture simulations:  
2. Sensitivity to external meteorological forcing (DOI 10.1029/2006JD007845)
- D22S04** *Alan K. Betts, Mei Zhao, P. A. Dirmeyer, and A. C. M. Beljaars*  
Comparison of ERA40 and NCEP/DOE near-surface data sets with other ISLSCP-II data sets (DOI 10.1029/2006JD007174)
- D22S05** *Randal D. Koster, Balázs M. Fekete, George J. Huffman, and Paul W. Stackhouse Jr*  
Revisiting a hydrological analysis framework with International Satellite Land Surface Climatology Project Initiative 2 rainfall, net radiation, and runoff fields (DOI 10.1029/2006JD007182)
- D22S07** *Eric C. Brown de Colstoun, Ruth S. DeFries, and John R. G. Townshend*  
Evaluation of ISLSCP Initiative II satellite-based land cover data sets and assessment of progress in land cover data for global modeling (DOI 10.1029/2006JD007453)
- D22S08** *Forrest Hall, Jeffrey G. Masek, and G. James Collatz*  
Evaluation of ISLSCP Initiative II FASIR and GIMMS NDVI products and implications for carbon cycle science (DOI 10.1029/2006JD007438)
- D22S09** *Lahouari Bounoua, Jeffrey Masek, and Yves M. Turre*  
Sensitivity of surface climate to land surface parameters: A case study using the simple biosphere model SiB2 (DOI 10.1029/2006JD007309)
- D22S10** *F. Aires and C. Prigent*  
Toward a new generation of satellite surface products? (DOI 10.1029/2006JD007362)
- D22S12** *Marc L. Imhoff and Lahouari Bounoua*  
Exploring global patterns of net primary production carbon supply and demand using satellite observations and statistical data (DOI 10.1029/2006JD007377)
- D02S90** *Li Dan, Jinjun Ji, and Yong He*  
Use of ISLSCP II data to intercompare and validate the terrestrial net primary production in a land surface model coupled to a general circulation model (DOI 10.1029/2006JD007721)



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

Forrest G. Hall,<sup>1</sup> Eric Brown de Colstoun,<sup>2</sup> George J. Collatz,<sup>3</sup> David Landis,<sup>2</sup> Paul Dirmeyer,<sup>4</sup> Alan Betts,<sup>5</sup> George J. Huffman,<sup>2</sup> Lahouari Bounoua,<sup>3</sup> and Blanche Meeson<sup>6</sup>

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/>),

<sup>1</sup>University of Maryland Baltimore County, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.

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

<sup>3</sup>NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.

<sup>4</sup>Center for Ocean-Land-Atmosphere Studies, Calverton, Maryland, USA.

<sup>5</sup>Atmospheric Research, Pittsford, Vermont, USA.

<sup>6</sup>Ocean 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

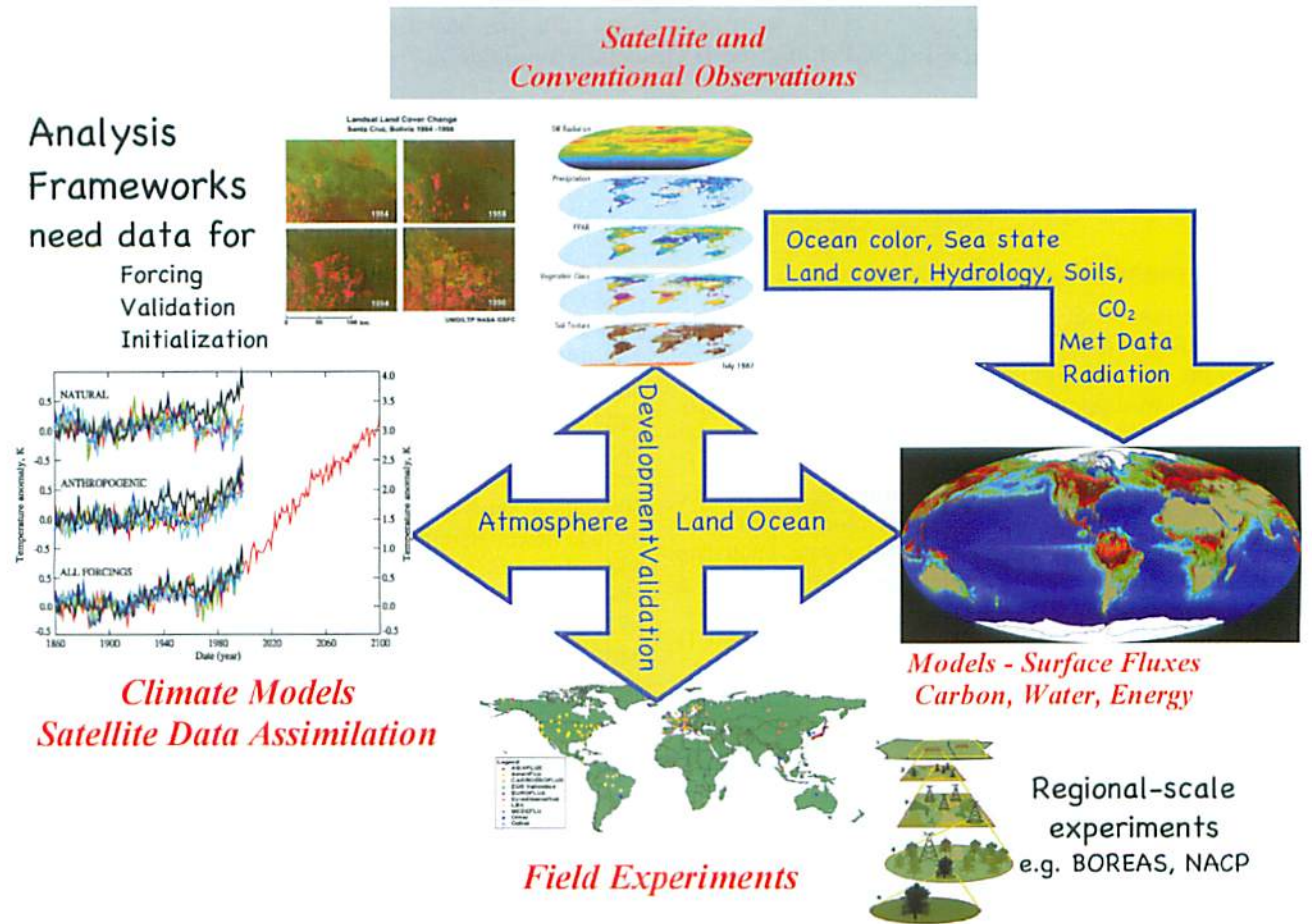
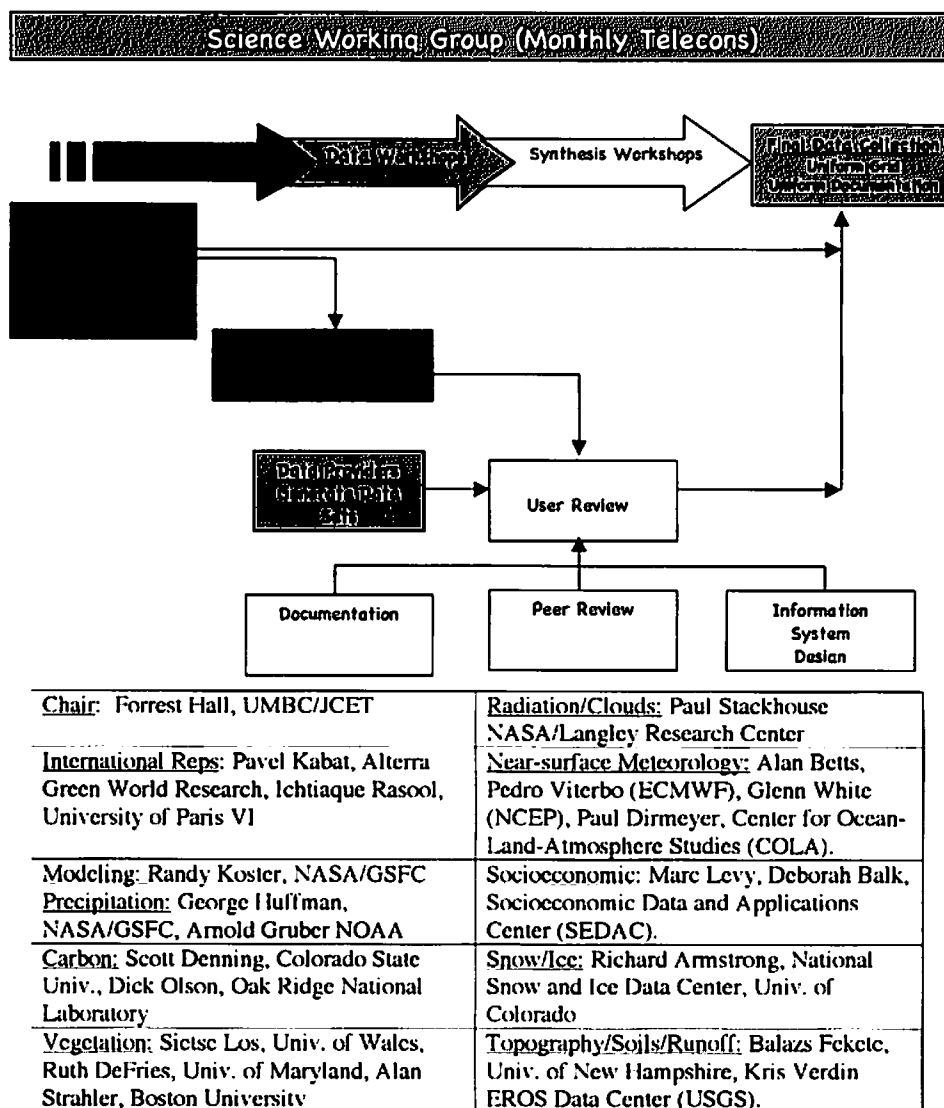


Figure 1. Analysis framework for GEWEX and IGBP/BAHC.

## 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.

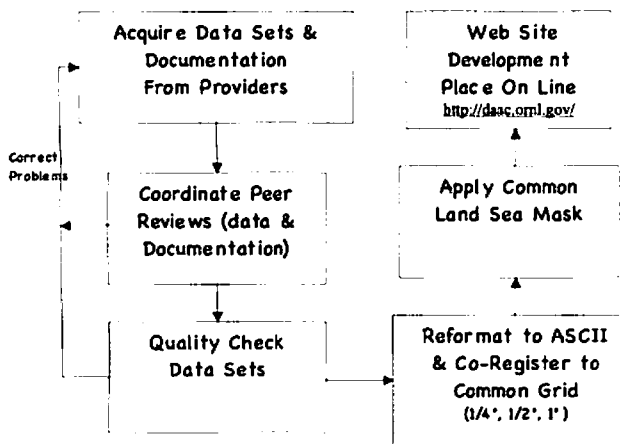
[4] The workshops were attended on average by about 50 scientists, consisting of the future users of the data including hydrologists, meteorologists, ecologists and the data providers covering expertise from remote sensing to meteorology and Earth and soil sciences. Without these meetings, some of the 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  $1^\circ$ ) 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  $2.5^\circ$  latitude and longitude. For compatibility, data sets with native resolutions greater than  $1^\circ$  were regridded to  $1^\circ$ . Except for data sets based on point measurements, all parameters in the entire data collection are available in common  $1^\circ$  versions.

## 2. ISLSCP Initiative II: Data

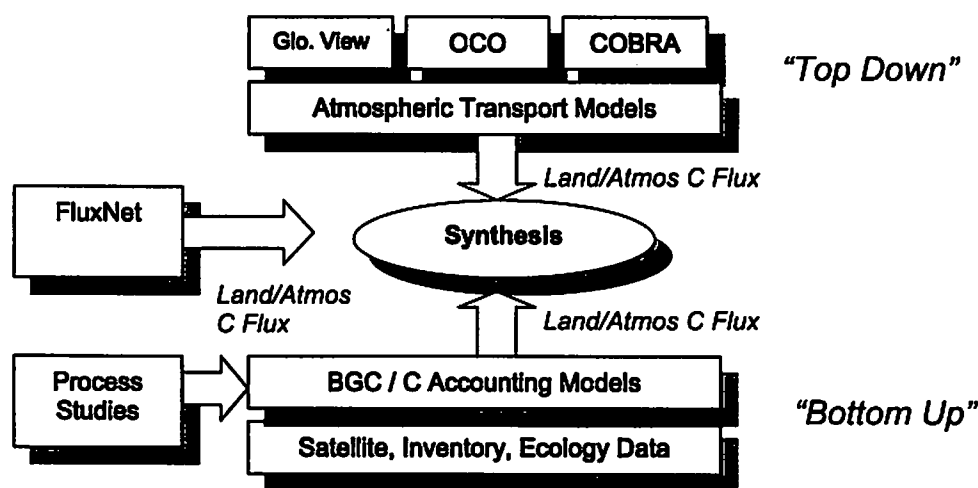
[10] In this section we describe the individual elements of the data collection and the rationale for their inclusion. In the Comparison and Evaluation section to follow, we contrast, compare and evaluate those data sets that contain common parameter sets (land cover, NDVI, albedo, precipitation and 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 CO<sub>2</sub> 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, CO<sub>2</sub>, 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  $1^\circ$  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  $1^\circ$  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  $1^\circ$  grid, data was regridded from its native T62 Gaussian grid ( $192 \times 94$  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° 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 ( $1^\circ$ ,  $1/2^\circ$  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  $1^\circ$ ,  $1/2^\circ$  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  $1^\circ$  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  $6^\circ$ ) 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  $1^\circ$  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 (ISOCCLUS)	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 <sup>a</sup>	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

<sup>a</sup>This 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 Fpar 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<sub>2</sub> record. Biologic CO<sub>2</sub> 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<sub>2</sub> 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  $\mu\text{m}$ ) 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	Vermote 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; Vermote 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 Guitman, 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 [Guitman 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-

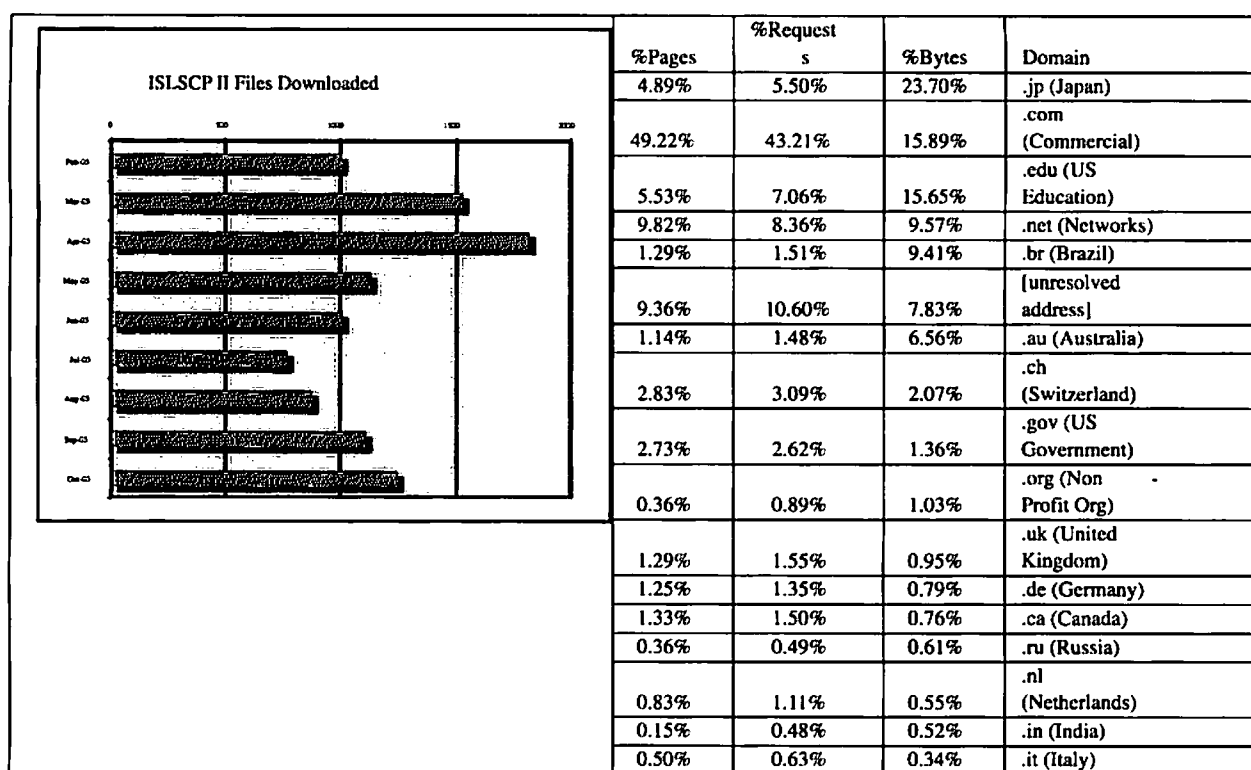
pared with the observationally based GPCP data set [Betts *et al.*, 2006]. Over the tropical oceans both reanalyses have more rainfall than GPCP, with ERA40 greater than NCEP2. Roads [2003] discusses the high bias of the NCEP2 reanalysis with respect to the Tropical Rainfall Measuring Mission (TRMM) 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
Carbon					
1	air-sea CO <sub>2</sub> gas exchange	Taro Takahashi, Lamont Doherty Earth Observatory, Columbia University	1995	5 × 4°, 1°	net sea-air CO <sub>2</sub> flux and sea-air CO <sub>2</sub> partial pressure (pCO <sub>2</sub> ) difference
2	atmospheric CO <sub>2</sub> consumption by continental erosion	Wolfgang Ludwig, University of Perpignan, France	N/A	1 and 0.5°	related to riverine flux data set
3	CO <sub>2</sub> emissions from fossil fuel burning	Gregg Marland and Antoinette Brenkert, 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 <sub>2</sub> , CH <sub>4</sub> , N <sub>2</sub> O) and tropospheric ozone precursor gas emissions (CO, NO <sub>x</sub> , NMVOC, SO <sub>2</sub> )
5	FLUXNET CO <sub>2</sub> (compiled from regional networks)	ORNL Distributed Active Archive Center (DAAC)	1992–1995	point data	example gap-filled CO <sub>2</sub> , 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
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	2363 point measurements and associated ancillary data
8	GPPDI gridded NPP data	Stephen Prince, University of Maryland; Daolan Zheng, University of Toledo	various	1 and 0.5°	not complete global coverage
9	GlobalView: atmospheric CO <sub>2</sub> concentrations	Ken Masarie, Climate Monitoring and Diagnostics Laboratory (CMDL), NOAA	varies by site	point data	GlobalView CO <sub>2</sub> 2003 data; smoothed, interpolated and extrapolated data
10	GlobalView: atmospheric methane concentrations	Ken Masarie, CMDL, NOAA	varies by site	point data	GlobalView CH <sub>4</sub> 2001 data; smoothed, interpolated and extrapolated data
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°	mean of 17 NPP models from Potsdam NPP model intercomparison
Hydrology, soils, and topography					
12	digital elevation and elevation-based derivatives from HYDRO1k	Kristine Verdin, National Center for EROS (USGS)	N/A	1 and 0.5°	subcell statistics for elevation, slope, aspect and compound topo. index
13	gauge-based daily precipitation	Pingping Xie and John Janowiak Climate Prediction Center, NOAA	1986–1995	1 and 0.5°	daily data for global land areas from Global Telecommunication Network (GTS)
14	precipitation monthly	Bruno Rudolf, Global Precipitation Climatology Centre (GPCC), Germany	1986–1995	1 and 0.5°	GPCC version 2 global gridded monthly gauge product
15	precipitation monthly (satellite and gauge)	George Huffman and David Bolvin, Global Precipitation Climatology Project (GPCP), NASA	1986–1995	2.5, 1°	monthly data, error fields and 1979–1999 climatology
16	precipitation pentad (satellite and gauge)	Pingping Xie, Climate Prediction Center, NOAA	1986–1995	2.5, 1°	GPCP pentad (5-day) data
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°	data set based on a CD-ROM developed by the Global Soil Data Task of the IGBP Data and Information Service
18	river discharge point data	Thomas Maurer, Global Runoff Data Centre (GRDC), Germany	1986–1995	point data	temporal coverage varies by station
19	river routing data (STN-30p)	Charles Vörösmarty and Balázs Fekete, University of New Hampshire (UNH)	N/A	1 and 0.5°	simulated topological network with gridded river basin data and attribute tables
20	global soil water storage capacity of the rooting zone	Axel Kleidon, University of Maryland	1987–1988 averaged 1°		data set used ISLSCP I forcing data

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)	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 Dirneyer 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	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	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	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)	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	Ancillary Tom Logan, Jet Propulsion Laboratory; ISLSCP II staff	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 CO<sub>2</sub> 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., 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. Beuscher, 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. Rittenberg (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. Köllberg (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\\_PRS13.pdf](http://www.ecmwf.int/publications/library/ecpublications_pdf/era40/ERA40_PRS13.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@agnes.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.



# Evaluation of the Second Global Soil Wetness Project soil moisture simulations:

## 1. Intermodel comparison

Zhichang Guo<sup>1</sup> and Paul A. Dirmeyer<sup>1</sup>

Received 27 February 2006; revised 19 September 2006; accepted 9 October 2006; published 22 November 2006.

[1] Driven with the meteorological data sets based on the reanalyses and gridded observational data archived by the International Satellite Land-Surface Climatology Project (ISLSCP) Initiative II, eleven different land surface models generated global soil moisture data sets for the 10-year period (1986–1995) for the Second Global Soil Wetness Project (GSWP-2). We evaluate these model simulations against in situ observations over grasslands and agricultural regions in the former Soviet Union, United States (Illinois), China, and Mongolia from the Global Soil Moisture Data Bank in terms of their ability to estimate the actual column plant-available soil moisture in the top 1-m soil layer, to simulate the phasing of the annual cycle, and to represent observed interannual variability. Results from these 11 land surface models show that they reproduce reasonably well the observed interannual variability and phasing of the annual cycle. Statistical analysis also shows that the median root mean square of errors among these models ranges from 4 to 8 cm of soil moisture. Similar to what has been found in soil moisture simulations for GSWP-1, the absolute values of soil moisture are poorly simulated by most models. However, the models do a good job of reproducing the soil moisture anomalies. This suggests that the global soil wetness data set produced by GSWP-2 can be used for analyzing climate variability and initializing GCMs by using transform strategies. This also has relevance to subseasonal to seasonal forecasts since soil moisture anomalies may potentially have impact on precipitation.

**Citation:** Guo, Z., and P. A. Dirmeyer (2006), Evaluation of the Second Global Soil Wetness Project soil moisture simulations: 1. Intermodel comparison, *J. Geophys. Res.*, **111**, D22S02, doi:10.1029/2006JD007233.

## 1. Introduction

[2] As an important component of the earth system, the land surface plays a vital role in the global climate system through interactions with the atmosphere [e.g., Shukla and Mintz, 1982; Sellers *et al.*, 1997; Chen *et al.*, 2001; Betts, 2004]. Accurate simulation of land surface states is critical to the skill of weather and climate forecasts and research on the terrestrial hydrological cycle. However, land surface modeling efforts suffer from sparse land surface observation networks for use in model initialization and data assimilation. This is particularly true for soil moisture, which is the definitive land surface state variable; key for model initial conditions from which the global weather and climate forecasts begin integrations, and a vital factor affecting surface heat fluxes and land surface temperature.

[3] Currently there is neither a global in situ observational network nor good estimates from satellite remote sensing for soil moisture. One approach for generating the long-term global estimates of the state of soil moisture is to integrate

uncoupled land surface models driven by externally specified near-surface meteorological forcing. The sources of global gridded analyses of the near-surface state of the atmosphere can be either operational analyses or retrospective analyses (reanalyses) from the major national meteorological centers. Besides the meteorological forcing, the offline simulation approach also needs information on soil and vegetation distributions. One complete set of such products comes from the International Satellite Land Surface Climatology Project Initiative II (II2) data set [Hall *et al.*, 2006], which contains soil, vegetation and other parameters required for complex land surface models. II2 also includes a version of near-surface meteorology over the whole globe with a common land-sea mask to the land surface parameters [Betts *et al.*, 2006]. This data set includes the latest reanalysis products from the National Centers for Environmental Prediction/Department of Energy (NCEP/DOE) reanalysis [Kanamitsu *et al.*, 2002] and the European Centre for Medium-Range Weather Forecasts (ECMWF) 40-year re-analysis (ERA-40) [Simmons and Gibson, 2000; Betts and Beljaars, 2003] for the 10-year period 1986–1995.

[4] The Second Global Soil Wetness Project (GSWP-2) is a multiinstitutional modeling research activity to produce a complete multimodel set of land surface state variables and fluxes by using current state-of-the-art land surface models

<sup>1</sup>Center for Ocean-Land-Atmosphere Studies, Calverton, Maryland, USA.

**Table 1.** Land Surface Schemes That Derived Soil Moisture Products for Evaluation<sup>a</sup>

Name	LSS	Group	LSS Reference	Soil, Vegetation
IIS	Bucket	U. Tokyo	<i>Manabe</i> [1969]; <i>Robock et al.</i> [1995]	n/a
ISBA	ISBA	Meteo France/CNRM	<i>Etchevers et al.</i> [2001]	soil: default; vegetation: IGBP
SSiB	SSiB	COLA	<i>Sellers et al.</i> [1986]; <i>Xue et al.</i> [1991, 1996]; <i>Dirmeyer and Zeng</i> [1997, 1999]	soil: GSWP-2; vegetation: SiB
LaD	LaD	USGS and NOAA/GFDL	<i>Milly and Shmakin</i> [2002a, 2002b], <i>Shmakin et al.</i> [2002]; <i>Swenson et al.</i> [2003]; <i>Milly et al.</i> [2003]	soil: default; vegetation: SiB
CLM	CLM2-TOP	U. Texas at Austin	<i>Bonan et al.</i> [2002]; <i>Niu et al.</i> [2005]	soil: default; vegetation: default
VISA	VISA	U. Texas at Austin	<i>Yang and Niu</i> [2003]; <i>Niu et al.</i> [2005]	soil: default; vegetation: default
NSIPP	Catchment	NASA GSFC/HSB	<i>Koster et al.</i> [2000]; <i>Ducharne et al.</i> [2000]; <i>Stieglitz et al.</i> [2001]	soil: GSWP-2; vegetation: SiB
Noah	Noah	NCEP/EMC	<i>Ek et al.</i> [2003]	soil: GSWP-2; vegetation: default
SWAP	SWAP	Institute of Water Problems, Russia	<i>Gusev and Nasonova</i> [1998, 2000, 2002, 2003]	soil: GSWP-2; vegetation: IGBP
MOSES	MOSES 2	Met Office, UK	<i>Cox et al.</i> [1999]; <i>Essery et al.</i> [2003]	soil: default; vegetation: default
MOSC	Mosaic	NASA/GSFC/HSB	<i>Koster and Suarez</i> [1992]	soil: GSWP-2; vegetation: SiB

<sup>a</sup>Information on soil and vegetation data sets used for the baseline experiment is also included. Data sets are either supplied by GSWP-2 or the model's default. For vegetation distributions, GSWP-2 supplied data sets include IGBP and SiB categories.

driven with the core 10-year period of data provided by I12. In order to produce the best available meteorological forcing for the GSWP-2 participating models, gridded observational data in the I12 data set are employed to correct some systematic errors found in the reanalysis products using the hybrid procedures described by *Zhao and Dirmeyer* [2003] and *Dirmeyer et al.* [2006]. The resulting GSWP-2 data are one of the best multiyear global estimates of soil moisture simulations available. In this paper, we evaluate GSWP-2 soil moisture simulations against available in situ observations to provide an estimate of uncertainties in land surface models and an indirect validation and quality check of the I12 data sets. In a companion paper [*Guo et al.*, 2006], the sensitivity of soil moisture simulations to different external forcing is further evaluated.

[5] Since soil moisture is a key factor affecting land surface temperature and latent heat flux, and hence the partitioning of outgoing convective fluxes between sensible and latent heat, it is given a great deal of attention when considering the performance of atmospheric general circulation models (AGCMs) as well as land surface models. *Robock et al.* [1998] found serious shortcomings in the Atmospheric Model Intercomparison Project (AMIP) soil moisture fields. *Li et al.* [2005] and *Lu et al.* [2005] evaluated reanalysis soil moisture simulations over China and the United States respectively. *Dirmeyer et al.* [2004] assessed eight multi-decade global soil moisture products with observed data from around the globe. These studies provide an assessment of the characteristics of various soil moisture estimates. Since soil moisture products used for validation in these studies are produced with various land-atmosphere coupled models or offline simulations driven with different meteorological forcing, the comparison also includes impacts of errors in the host AGCM or differences in the external meteorological forcing in the case of offline simulations. This makes it difficult to draw overall conclusions from such disparate studies. As all land surface models in GSWP-2 are forced with the same meteorological forcing, the evaluation in this study provides a fair comparison to sample the uncertainties in the land surface models specifically. This paper is comparable to the work of *Entin et al.* [1999] who validated the simulated soil moisture from the models in the pilot Global Soil Wetness Project (GSWP-1). The GSWP-2 not only extends the core period in GSWP-1 from 2-year simulations (1987–1988) to a

10-year period (1986–1995), but also includes updates in both the land surface models and meteorological forcing. Section 2 describes the various model estimates and in situ observations used for evaluation. Evaluation of soil moisture simulations against in situ observations is presented in section 3. Conclusions are given in section 4.

## 2. Land Surface Models and Preprocessing of the Data Sets

### 2.1. Land Surface Models

[6] Table 1 lists the models that we included in the evaluation. All of these models are built around physically based parameterizations of the behavior of the land surface water balance and a full representation of the surface energy balance. Not all participating GSWP-2 models are included in this analysis: some are missing owing to a lack of availability of key model output at the time of data processing and technical difficulties in data analysis due to model soil column definitions. These models represent a broad cross section of schemes in use today. IIS and LaD stem from the simple so-called “first generation” bucket model. MOSES2 and CLM2-TOP are examples of the latest “third generation” land surface models which include a sophisticated representation of the carbon cycle, biogeochemistry and dynamic vegetation. The remainder of the models belong to the “second generation” of land surface models, which include a realistic treatment of the surface energy balance and intermediate complexity of effect of vegetation. This classification approach is standard among land surface models, but does not accurately rank the sophistication of the soil hydrology among schemes, which is an important aspect of this study. Most of the second-generation models trace some lineage in their design to either the Simple Biosphere (SiB) model of *Sellers et al.* [1986] or to the Biosphere Atmosphere Transfer Scheme (BATS) [*Dickinson et al.*, 1986]. More detailed descriptions of these models can be found in the references listed in Table 1 and given by *Dirmeyer et al.* [2006].

### 2.2. Model Estimates

[7] The GSWP-2 experiment was carried out with each land model set up to run globally (excluding Antarctica) on a regular  $1^\circ \times 1^\circ$  latitude-longitude grid. All models that

**Table 2.** Meteorological Forcing and Vegetation Parameter Data Used for the Baseline Simulation of GSWP-2

Name	Description	Main Products	Hybrid Products
Tair	near-surface air temperature at 2 m	NCEP/DOE	CRU [New et al., 1999, 2000]
Qair	near-surface specific humidity at 2 m	NCEP/DOE	CRU [New et al., 1999, 2000]
Wind	near-surface wind speed at 10 m	NCEP/DOE	
SWdown	surface incident shortwave radiation	SRB [Stackhouse et al., 2000]	
LWdown	surface incident longwave radiation	SRB [Stackhouse et al., 2000]	
Psurf	surface pressure	NCEP/DOE	
Rainf	rainfall rate	NCEP/DOE	GPCC, GPCP [Rudolf et al., 1994; Huffman et al., 1997]
Rainf_C	convective rainfall rate	NCEP/DOE	GPCC, GPCP [Rudolf et al., 1994; Huffman et al., 1997]
Snowf	snowfall rate	NCEP/DOE	GPCC, GPCP [Rudolf et al., 1994; Huffman et al., 1997]
VegClass	vegetation class	IGBP/SIB/BATS	
LAI	leaf area index	Univ. of Wales	
vegFrac	fraction of vegetation cover	UK	
FPAR	fraction of photosynthetically active radiation	UK	
gmFrac	greenness fraction	calculated from Univ. of Wales' data	

participated in the GSWP-2 experiments were integrated in an offline mode, driven by hybrid meteorological forcing from global reanalysis, rescaled by gridded observations and satellite products where available. Standardized soil and vegetation distributions from II2 were provided for each modeling group. Each participating model generated a complete set of land surface state variables and fluxes for the 10-year period 1986–1995, which include daily soil moisture fields. Model estimates used for evaluation come from the GSWP-2 baseline experiment. Table 2 lists the default II2 data sets used in the baseline simulations. Deviations from the recommended specifications of land surface parameter data sets listed in Table 2 are indicated for each model in Table 1. The time series for each model is retrieved from its global products at the grid box that contains a particular station (as described in the next section). In this paper we use column plant-available soil moisture for the evaluation, which is calculated as a summation of soil moisture at each soil layer while subtracting the wilting point at that layer. The data set for each model are converted to a columnar plant-available soil moisture within the top 1-m according to the specific layer structure in the model. If the model soil depth is shallower than 1 m, the deep layer plant-available soil moisture amount is weighted to fill in the additional depth. If the model soil depth is deeper than 1 m, only the column soil moisture summation for the top 1 m is calculated. For those single-layer bucket-type models, column plant-available soil moisture is calculated according to their bucket capacity. The reported values in BUCKET/II2 and LaD data sets are used without any further processing because the 150-mm bucket depth originally prescribed by Manabe [1969] represents the active middle third of the soil moisture range of a 1-m column of soil with porosity of 0.45. It is noted that this factor only affects the statistics of the root mean square of errors, and does not affect temporal correlations between the model estimates and observations of soil moisture. To fairly compare the simulated soil moisture to observations, monthly mean soil moisture values are averaged from daily mean model outputs when the observed soil moisture is available. Some models are more highly calibrated than others, sometimes using the same in situ soil moisture

observations we use for validation. No attempt is made to account for this, or otherwise to “level the playing field.”

### 2.3. In Situ Observations

[8] The most complete collection of soil moisture measurements with long temporal and global coverage is the Global Soil Moisture Data Bank (GSMDB) of Robock et al. [2000]. This collection of station measurements covers regions of North America, Europe and Asia. Most measurements are gravimetric and taken in agricultural areas. We focus on station data in four regions (Illinois, USA; China; Mongolia and the former Soviet Union). The former Soviet Union (FSU) data are further divided into two categories, representing winter and spring cereal fields. These data sets span decades, although individual stations may have a much shorter record of observations.

[9] Measurement conventions vary from region to region for the GSMDDB data sets. The column plant-available soil moisture for the top 1-m soil layer preprocessed by Robock et al. [2000] are used for the evaluation, and the instantaneous values are aggregated to get the monthly mean observations. No interpolation is applied to fill the missing data, and only the data when and where they are available are used. Those stations where more than 75% of the observations are missing during the evaluation period are excluded from the evaluation exercise.

### 3. Results

[10] To be comparable to Entin et al. [1999], we have chosen plant-available water rather than the total soil moisture for the evaluation. The total numbers of the stations over the domains are: China (40 stations), Illinois (19 stations), Mongolia (42 stations) and FSU (171 stations total among the two categories) (see Figure 1 for station locations). Even applying the threshold of no more than 75% missing data, many stations are not useable for the comparisons. The percentages of stations used are as follows: China (90%), Illinois (100%), Mongolia (65%), FSU (46%, spring cereal; 47%, winter cereal). There are few or no data during winter over FSU and Mongolia in particular because of frozen ground.

[11] Koster and Milly [1997] showed that different models have very different ranges of soil moisture variations for the same locations. Soil moisture is a reservoir term that comes to

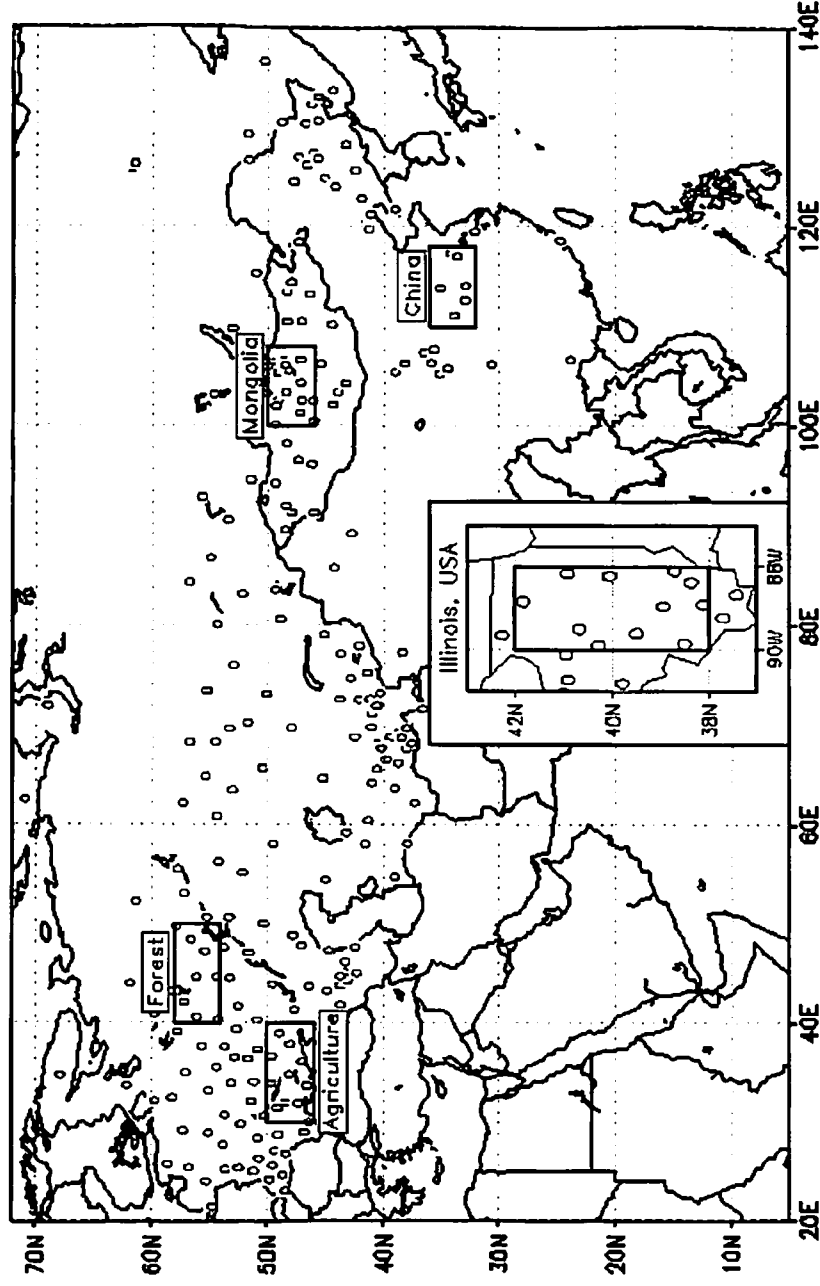


Figure 1. Soil moisture observation stations for Eurasia and Illinois. Boxes used for spatial averages in Figures 7 and 8 are indicated.

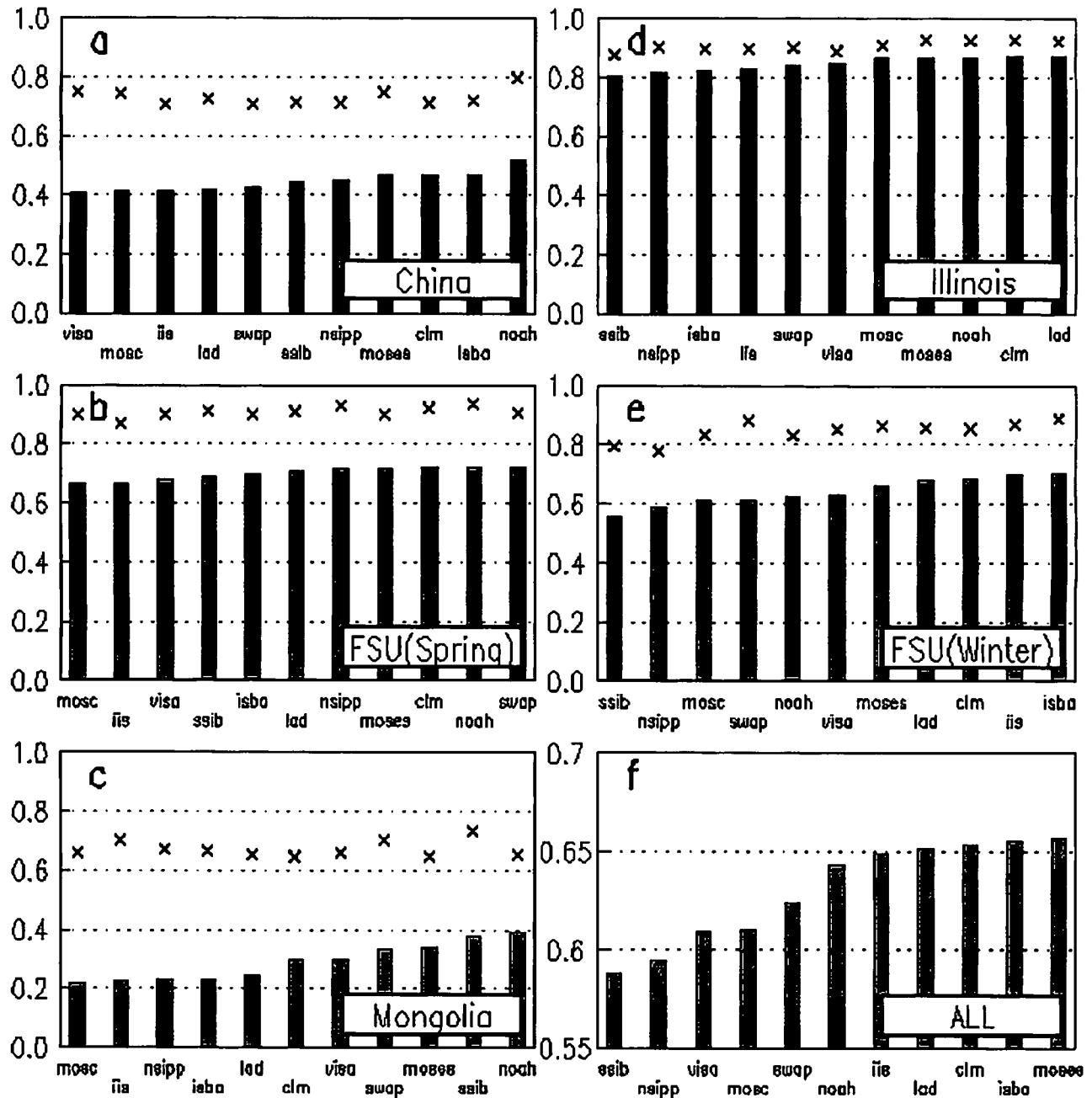
equilibrium based on the balance of precipitation, evapotranspiration and runoff. Precipitation is the same for all models in GSWP-2. Thus it is the balance between evapotranspiration and runoff in each model that determines the climatology of soil moisture. Even small differences in parameterizations can have a drastic effect on this balance. Temporal correlations are a particularly appropriate measure of the models' performance in simulating observed soil moisture time series in this regard. Over some regions, the hydrologic cycle is dominated by a strong annual cycle with well-defined wet and dry seasons. The correlation for the total field may largely reflect model's performance in simulating the phasing of the annual cycle of soil moisture in such areas, with interannual variations contributing only a small portion of the signal. To evaluate model performance in representing interannual variability, we also examine the simulation of anomalies by removing the mean annual cycle from the total series. For ease of comparison in the following figures we use the median correlation, defined as the median of a group of correlations between model simulations and observations over a number of stations in a region, as the measure for evaluating the model's overall performance in simulating observed temporal variability of soil moisture over these regions. We also include extrema in these figures to give an idea of the range of skill within each model.

[12] Figures 2 and 3 compare median correlations between the model estimates and observations of soil moisture in terms of the total series (Figure 2) and the anomalies (Figure 3) for stations over individual domains (Figures 2a–2e and 3a–3e) and totaled over all domains

(Figures 2f and 3f). It is obvious that differences in skill exist between regions, and they are much larger than the differences among land surface models. The relatively small intermodel differences reflect the fact that the same meteorology forcing is used to drive these models, and it suggests that the quality of soil moisture simulations is largely controlled by the quality of external forcing data or land surface parameters used for the offline experiment. Illinois shows some of the highest overall median correlations for both the total field and anomalies. It is not the quality of the soil moisture measurements, but rather the quality of the meteorological data used to force the models, which is likely of highest quality over Illinois. As an example, the GSWP-2 Science and Implementation Plan shows that the rain gauge density is considerably higher over the eastern U.S. than other areas with in situ soil moisture data [International GEWEX Project Office, 2002]. Oki *et al.* [1999] showed for GSWP-1 that the quality of the model simulations of surface hydrology was directly proportional to the rain gauge density.

[13] Correlations between the observations and model estimates are relatively low over Mongolia. This might be a net result of relatively poorer quality of meteorological forcing, soil moisture observations used for the validation, and the failure of the land surface models to simulate key aspects of the hydrological cycle over that region. Also, correlations for all models tend to be higher for the total field than for the anomalies over Illinois and FSU while they have the opposite tendency over Mongolia and China. In Mongolia and China there is much less of a seasonal cycle of soil moisture than for the other locations. Precip-

## Median Correlation Monthly Means (1986–1995)



**Figure 2.** Median temporal correlations between monthly mean simulated columnar plant-available soil moisture and observations covering the period 1986–1995 for stations (a–e) over individual domains and (f) over all domains for each model. The products are shown in ranked order, from lowest to highest correlations, on the basis of the median values. The highest correlations for each model are also indicated with cross marks.



## Median Correlation

Anomalies (1986–1995)

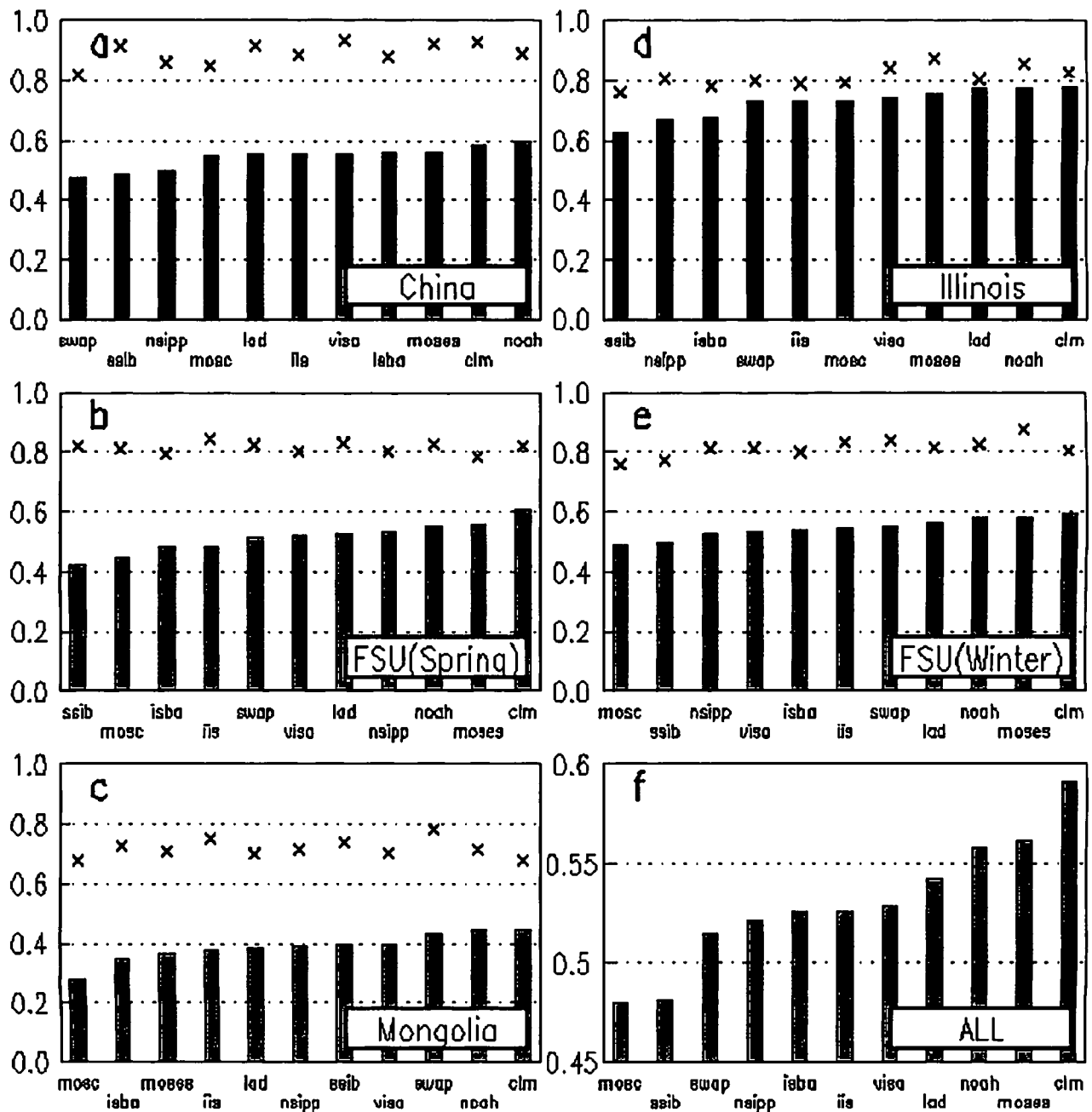
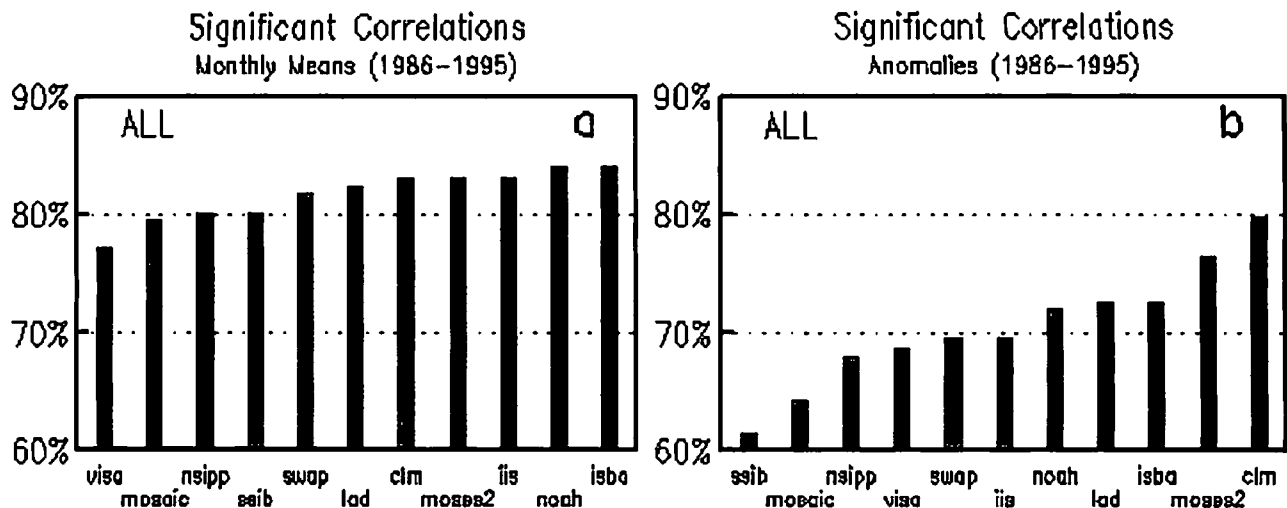


Figure 3. Same as Figure 2 except for the anomalies (annual cycle removed).



**Figure 4.** Fraction of stations where temporal correlations between the model estimates and observations of columnar plant-available soil moisture for (a) the monthly mean and (b) the anomalies are significantly correlated at 95% confidence level for the period 1986–1995 when stations in all five regions are considered. The products are shown in ranked order from lowest to highest significance counts.

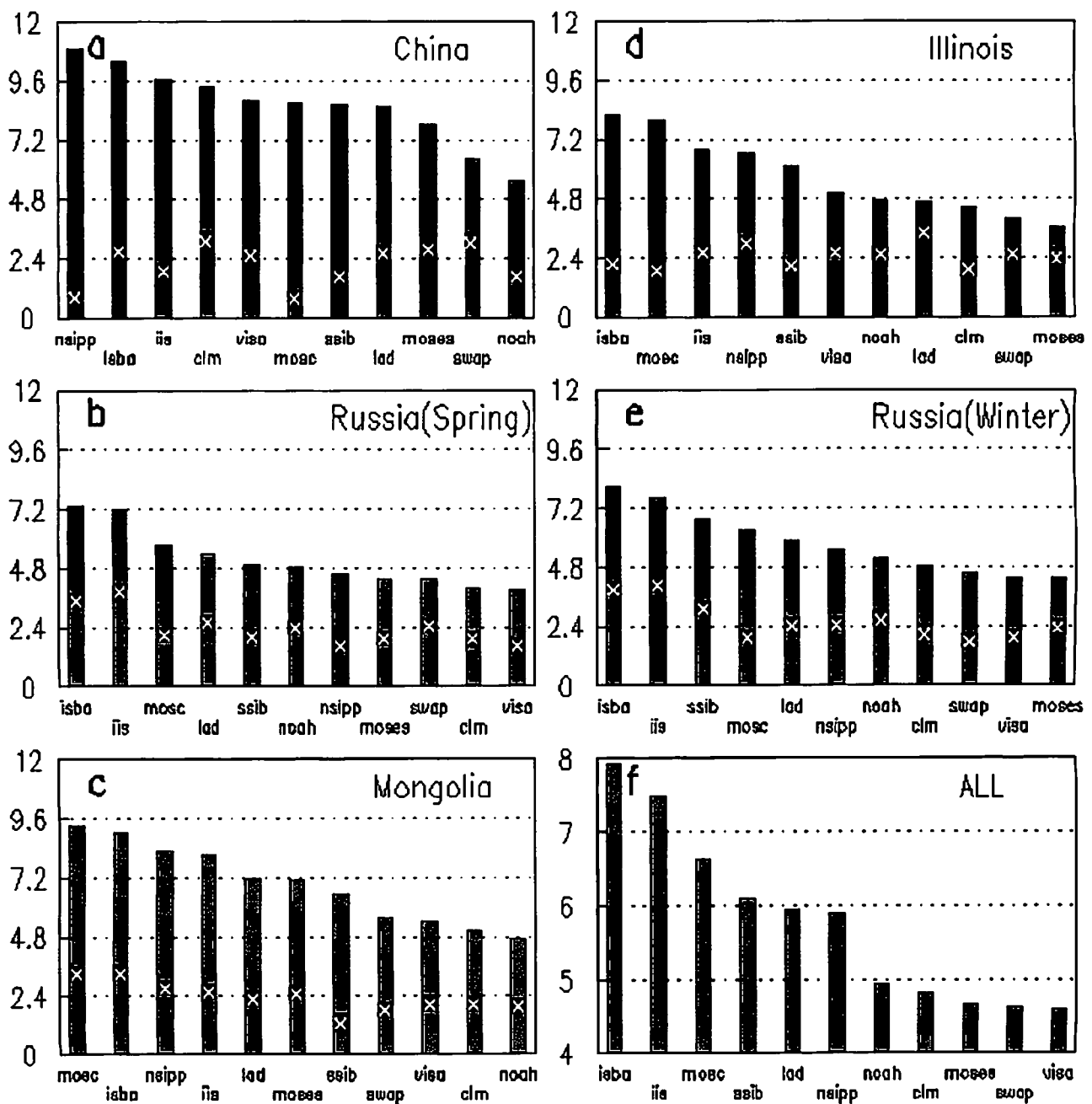
itation and evapotranspiration have the same seasonal cycle, both peaking in the summer owing to the influence of the summer monsoon *Entin et al.* [1999]. Thus it is more challenging for the models to reproduce seasonal variability in soil moisture over these two regions. Figures 2 and 3 also show that there are certain differences in skill among land surface models. Validation with long-term in situ data indicates that some models are generally superior to others, though they are not always the best in all regions and for both the total fields and anomalies. In general, CLM, Noah, and MOSES2 have higher temporal correlations than the other models. It is noted that the NSIPP model reports the topographically averaged root zone moisture which reflects the spatial average across several hydrological regimes. This makes the NSIPP soil moisture outputs unsuitable to be compared to in situ measurements. The highest correlations over individual domains for each model are also indicated with cross marks in Figures 2 and 3. In fact, some models show a very tight grouping of correlations across the stations, while others show a large spread and even occasionally extreme outliers, especially on the low side (not shown). As a result, two models that have comparable median skill may have very different abilities to simulate the soil moisture at individual stations, and the model with the largest median correlation is not necessarily the model that achieves the highest correlations over individual locations.

[14] Another way to evaluate the model performance is to compare the fraction of stations where the significance of temporal correlations between simulated soil moisture and observations is above a certain level. Figure 4 shows the fraction of stations whose significance level is higher than 95% for each model for the total series (Figure 4, left) and the anomalies (Figure 4, right) when all stations over five domains are considered together. Missing data is a common problem for soil moisture observations, and it makes the number of available observation data records vary from station to station. The degrees of freedom, on which the significance level is dependent, is taken to be half of the

number of available monthly observation data records in each station. Figure 4 indicates that every model can simulate the time series with significant correlations at more than 75% of the stations. The fraction of significant time series is lower for the anomalies. However, it is still above 60% for all models with half of the models higher than 70%. It is also seen in Figure 4 that the intermodel differences in skill of simulating temporal variability for the anomalies are larger than for the monthly mean fields, a range of 18% versus 7% for the total field.

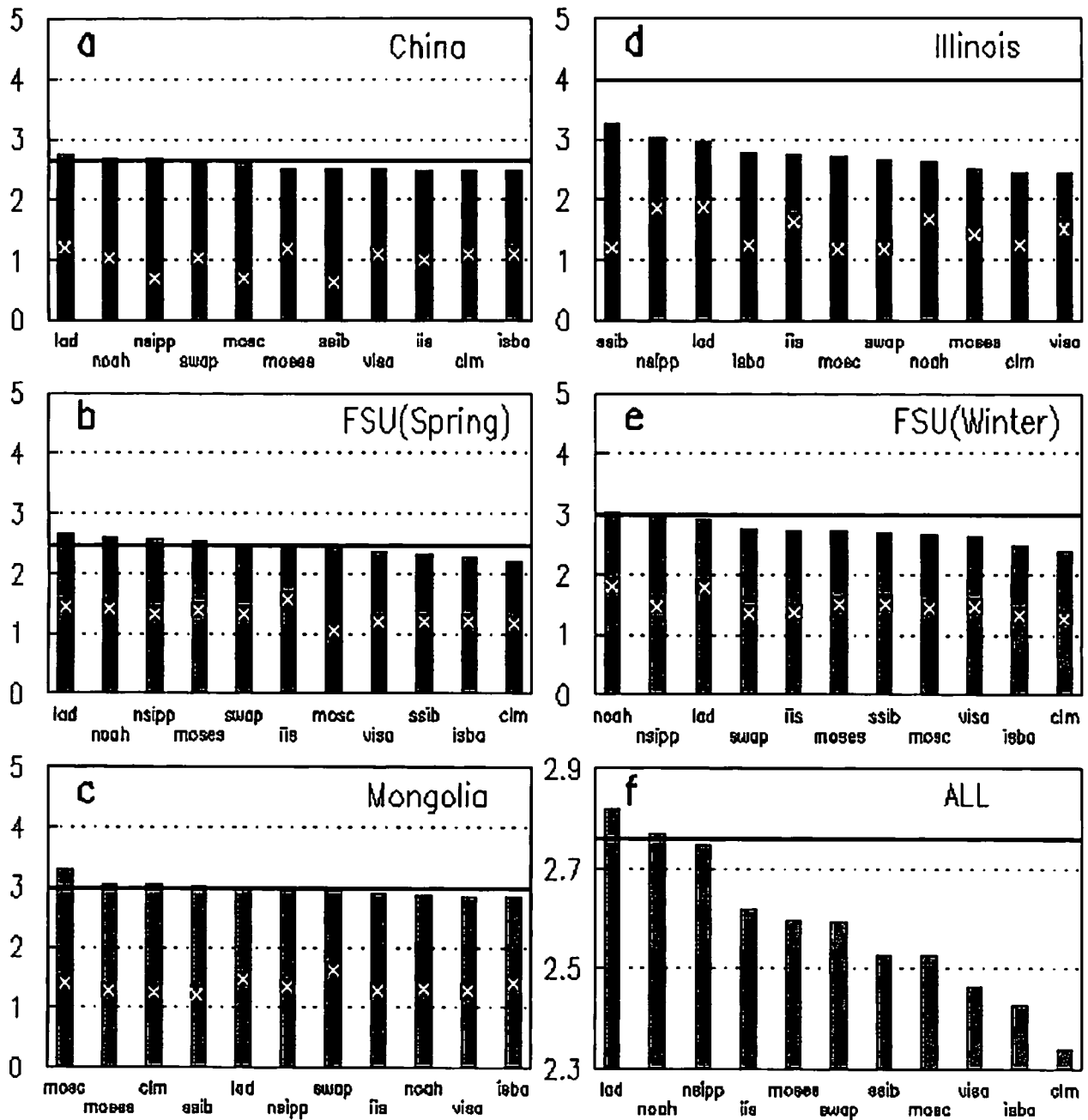
[15] Not only is temporal variability of soil moisture valuable for climate studies, but also the actual mass of soil moisture is important to many applications [*Robock et al.*, 1998]. For assessment of accuracy, the root mean square of errors (hereafter referred to as RMSE) is used over a temporal domain. Figures 5 and 6 compare median RMSE between the model estimates and observations of soil moisture for both the total series (Figure 5) and the anomalies (Figure 6) for stations over individual domain (Figures 5a–5e and 6a–6e) and stations over all domains (Figures 5f and 6f). The lowest RMSE over each domain for each model is also indicated with cross marks in these figures, indicating the best simulation in each region. RMSE of the total series highlight biases in soil moisture. Similar to what has been found by *Entin et al.* [1999] in the evaluation of the pilot GSWP soil moisture simulations, estimation of the actual soil moisture is still a problem for the models in GSWP-2. The median RMSE between the model estimates and observations of soil moisture in the total fields over China and Mongolia tends to be larger than that over the rest of domains. The median RMSE for the total series ranges between 4 and 11 cm (Figure 5). If the active soil moisture range of a 1-m column is assumed to be 15 cm according to the value prescribed by *Manabe* [1969] for his bucket model, the median RMSE represents 25–70% errors in the simulated monthly soil moisture values. We also see a large spread in the simulated plant-available soil moisture values among models in Figure 5. Over all domains, this uncertainty in the

# Median RMSE Monthly Means (1986–1995)

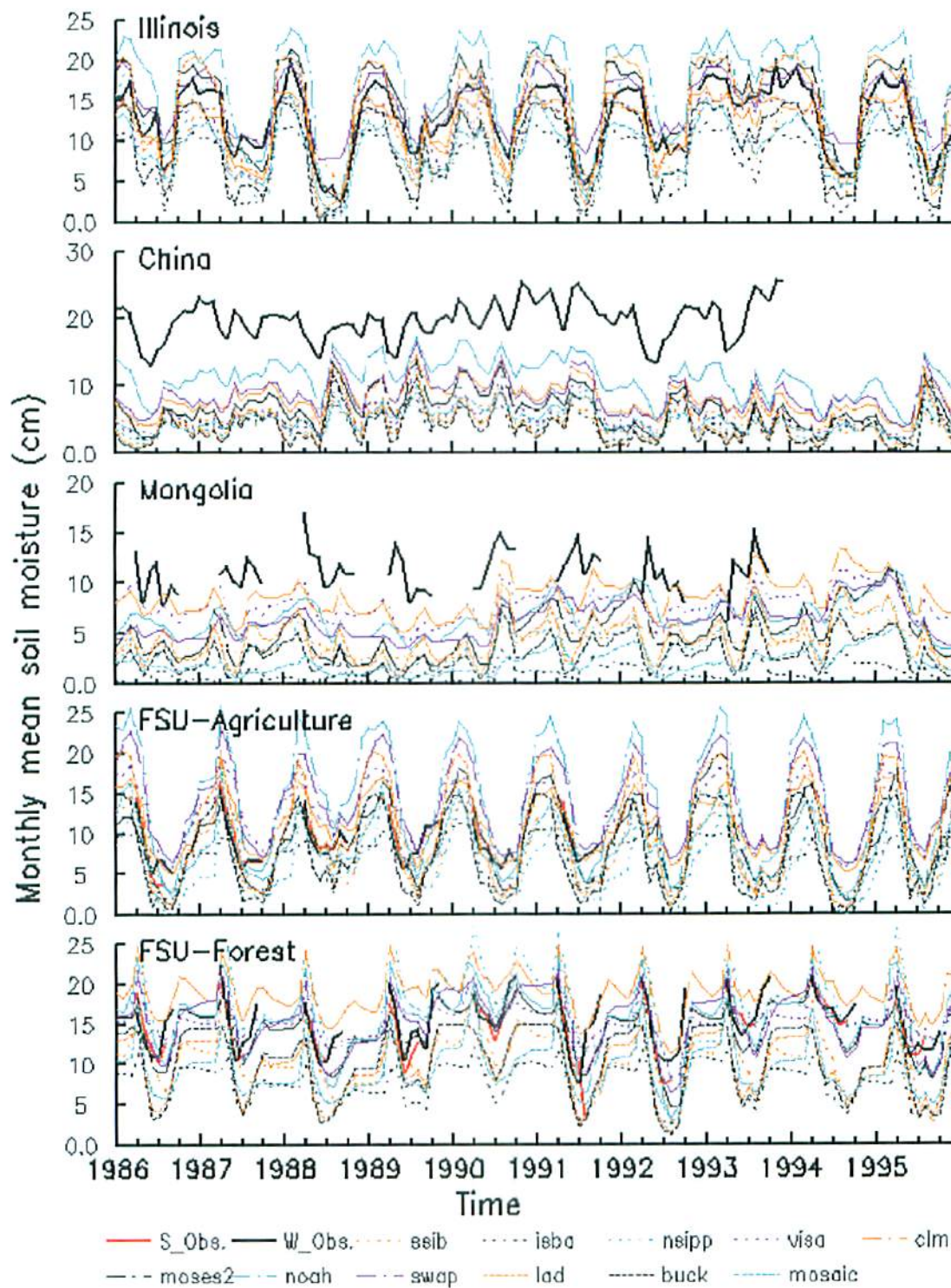


**Figure 5.** Median root mean square of errors (RMSE) between the monthly mean simulated columnar plant-available soil moisture and observations (1986–1995) for stations (a–e) over individual domains and (f) over all domains. The products are shown in ranked order, from highest to lowest RMSE, on the basis of the median values. Units are centimeters of water in the soil. The smallest RMSE for each model are also indicated with cross marks.

# Median RMSE Anomalies (1986–1995)



**Figure 6.** Same as Figure 5 except for the anomalies (annual cycle is removed). Units are centimeters of water in the soil. The median standard deviation of soil moisture anomaly for stations (a–e) over individual domain and (f) over all domains is represented by the black solid line.



**Figure 7.** Time series of columnar plant-available soil moisture in the top 1 m from model estimates and observations for the specific areas indicated in Figure 1. The boxes are at 30°E–40°E, 46°N–50°N for FSU (former Soviet Union) agriculture; 40°E–50°E, 54°N–58°N for FSU forest; 90°W–88°W, 38°N–42°N for Illinois; 110°E–118°E, 32°N–36°N for China; and 100°E–108°E, 46°N–50°N for Mongolia. S\_Obs. represents the observations for the FSU spring cereal fields, while W\_Obs. represents the observations for the FSU winter cereal fields and the standard observations over the other regions.



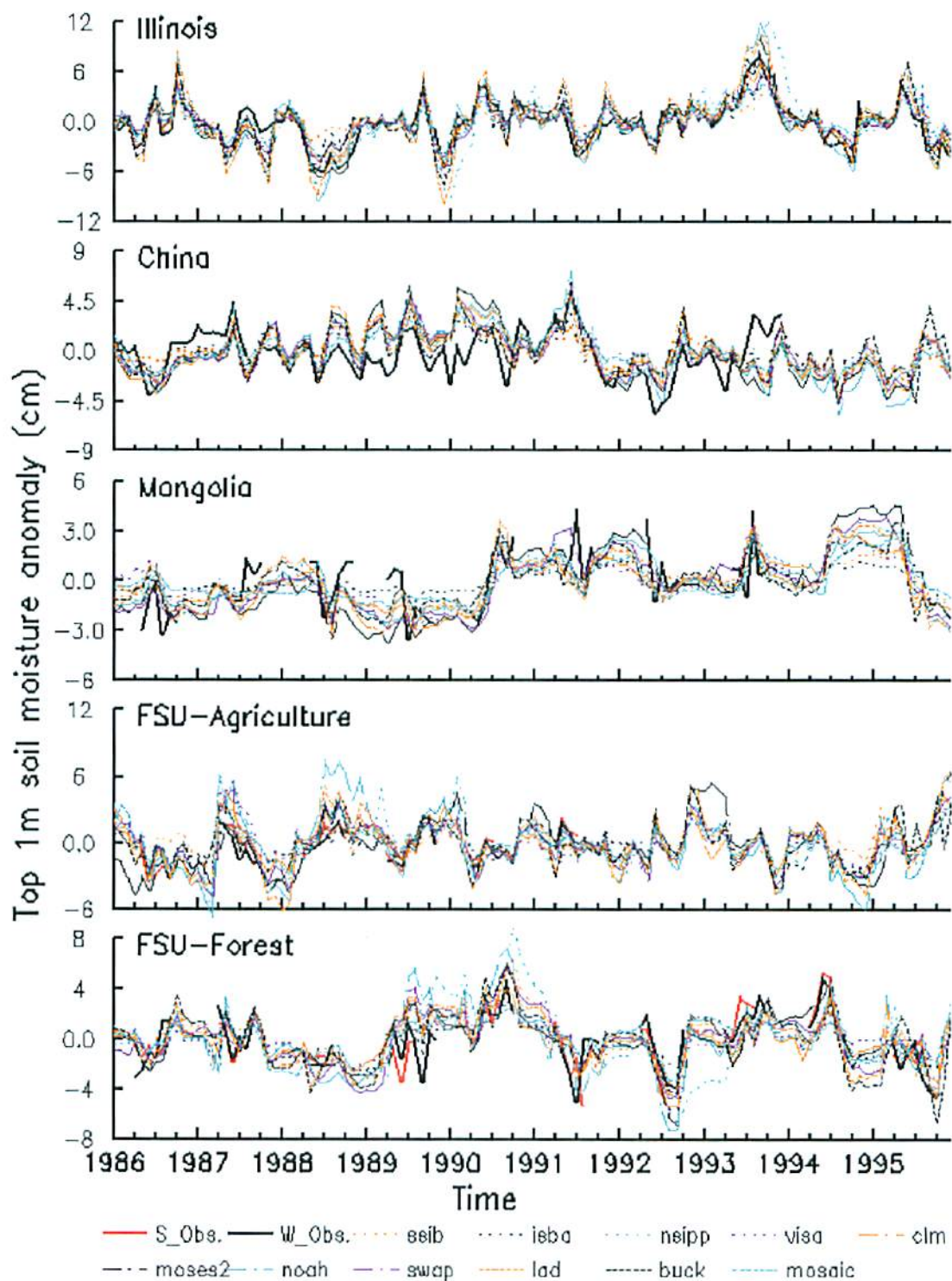


Figure 8. Same as Figure 7 except for the anomalies (annual cycle removed).

land surface models is almost as large as the average median RMSE across the models. This suggests that part of the problem in the estimate of actual soil moisture can be attributed to the uncertainty in the land surface models and land surface parameters supplied to the models. It is likely a coincidence that the model with the highest total field correlation in Figure 2f (ISBA) has the largest total RMSE (Figure 5f), and vice versa (VISA) when all stations over five domains are considered.

[16] RMSE of the anomalies reveal no systematic bias, but may include other errors such as spurious trends and accumulated errors of fluxes (evapotranspiration and runoff) across the annual cycle and from year to year. The median standard deviation of soil moisture anomaly for stations over individual domains (Figures 6a–6e) and stations over all domains (Figure 6f) is represented by the black solid line. This represents the error that would result from a forecast of climatology. This can be considered as an additional model of soil wetness; a minimum threshold or benchmark that any land surface scheme should be able to beat. For almost all models in each region, the median RMSE of soil moisture anomaly is smaller than the median standard deviation of soil moisture. Over Illinois, the median RMSE is much smaller than the median standard deviation of soil moisture anomaly for all models. This suggests that the soil moisture anomalies over the region can be simulated with high accuracy.

[17] In order to better simulate the plant-available soil moisture, the model needs an accurate spatial distribution of wilting point and other soil properties, and should accurately simulate the changes in moisture content in the soil. Owing to the different strategies used for simulating the terrestrial hydrology and different global data sets of soil properties used by some GSWP-2 land surface models, different models have different climatologies or “operating ranges” for plant-available soil moisture [Koster and Milly, 1997]. The large spread in the median RMSE across models can be attributed to the different “operating ranges” that different land surface models have. However, such spread in RMSE of plant-available soil moisture among models is largely reduced for the anomaly fields in Figure 5f. The median RMSE in these models range from 2.3 cm to 2.8 cm. This small intermodel spread in the median RMSE of anomalies reflects the removal of uncertainties resulting from the different “operating ranges” in the land surface models.

[18] Similar to Entin *et al.* [1999], spatial averages over the five domains are also used to evaluate the GSWP-2 soil moisture simulations against the observations in this study. Both the averaging method and the five regions selected for averaging are same as that used by Entin *et al.* [1999]. The regions selected are either natural grassland or agricultural cultivation to avoid signals associated with heterogeneous vegetation cover within the regions. The optimal averaging technique developed by Kagan [1979] is used for the spatial average since it has the advantage of taking into account the spatial scale of soil moisture and gives less weight to stations where time series of soil moisture are correlated and more weight to stations where soil moisture tends to evolve independently. This method is very helpful for producing a spatial average when there is intermittent missing information, causing a changing distribution of stations.

[19] Figure 7 shows the plant-available soil moisture in the top 1 m produced by the eleven models and the observations

for each of the regions in Figure 1. As expected from Figure 5, a large spread appears in the monthly mean soil moisture simulated by these models. For all domains, the spread of model estimates is about 10 cm from the driest to wettest model. The time series of observations in Figure 7 show that there are pronounced wet and dry seasons for Illinois and FSU whereas there is much less of a seasonal cycle for China and Mongolia. All models do an excellent job of reproducing the seasonal cycle for the regions of Illinois and FSU. Similar to what has been found by Entin *et al.* [1999], a large dry bias is obvious for all models in the regions of China and Mongolia. This might be caused by errors in the land surface parameters supplied to the models or biases in the external forcing in these regions. The only soil property information that is available for the Chinese or Mongolian stations is wilting point.

[20] We also explore the interannual variability of soil moisture in Figure 8, where the seasonal cycle has been removed from the total fields. After removing the seasonal cycle, both the intermodel spread and biases of simulated soil moisture are largely reduced. Figure 8 shows that land surface models can do a relatively good job of reproducing regional soil moisture anomalies. The interannual variability of soil moisture is well simulated by all models. In Illinois, the models show their ability to simulate the drier than normal summer season in 1988 and the wetter than normal summer in 1993. However, we also see that intermodel spread is greatest during such extreme events. This is also true for the other regions except Mongolia, where long periods of frozen soil during winter can lock in a large intermodel spread regardless of the magnitude of the anomaly. It is not obvious from the area-averaged monthly data shown in Figures 7 and 8, but when point-wise daily output is examined, the models clearly converge during significant rain events (when they all tend toward saturation) and then spread during dry-down as each has a different rate of water loss by runoff plus evapotranspiration [Dirmeyer *et al.*, 2006]. A couple of weeks after a large rain event there is again a large spread among models in total soil moisture. The models also tend to asymptotically approach different minima of soil moisture during droughts (another manifestation of the different “operating ranges” among models). The spread for extremes in anomalies is also an artifact of the different levels of mean soil moisture subtracted from each model, especially for the spread during pluvial regimes. From daily output, the models can be seen to converge during droughts for the total soil moisture, but they do not for anomalies as we have computed them.

#### 4. Discussion and Conclusions

[21] Model estimates of soil moisture by eleven land surface models participating in GSWP-2 are validated against the observations. All models do a reasonably good job in reproducing the phasing of seasonal cycle and the interannual variation over the regions examined. Within regions there can be tremendous variations of any model to simulate the time series of soil moisture at different stations. All models seem to have comparable upper bounds on skill, even though the median values of correlations and RMSE vary substantially from model to model. Simulating the actual values of observed soil moisture is still a challenging task for all models. Both the root mean square of errors (RMSE) and

the spread of RMSE across models are large. However, all models reproduce the anomalies (with seasonal cycle removed) with reasonable accuracy. Also, the uncertainty in land surface models for reproducing anomalies is small in comparison to the absolute biases. This suggests that the global soil wetness data set produced by GSWP-2 can be used for analyzing climate variability and initializing GCMs by using certain strategies, such as the transform technique proposed by Dirmeyer *et al.* [2004]. This also has relevance to subseasonal to seasonal forecasts since simulation of soil moisture anomalies might be more important than that for actual monthly mean values owing to the persistence of soil moisture anomalies and its potential impact on future precipitation.

[22] The evaluation also shows that the quality of external meteorological forcing plays an important role in the offline simulations of soil moisture. Thus we are motivated in a companion paper to evaluate the sensitivity of soil moisture simulations to the different meteorological data sets provided by the ISLSCP Initiative II, and provide an indirect validation and comparison of the II2 data sets. Also, if the differences among meteorological data sets represent the uncertainty in our knowledge of these drivers of the land surface climate, the sensitivity analysis provides an assessment of impact of that uncertainty on land surface hydrology.

[23] **Acknowledgments.** We would like to thank all of the modeling groups for their participation and patience. We give special thanks to Jennifer Miletta Adams for making the forcing data available online to all the modeling groups. The authors thank Alan Robock and Haibin Li for providing soil moisture observations for validation. The authors also thank Masaru Kitsuregawa, Eiji Ikoma and Masaki Yasukawa for their generous contribution to the development of the GSWP-2 ICC web site. We are indebted to Guo-Yue Niu and an anonymous reviewer for their constructive comments on the earlier version of the manuscript. This study was supported by National Aeronautics and Space Administration grant NAG5-11579.

## References

- Betts, A. K. (2004), Understanding hydrometeorology using global models, *Bull. Am. Meteorol. Soc.*, **85**, 1673–1688.
- Betts, A. K., and A. C. M. Beljaars (2003), ECMWF ISLSCP-II near-surface dataset from ERA-40, *ERA-40 Proj. Rep. Ser. 8*, Eur. Cent. for Medium-Range Weather Forecasts, Reading, U.K. (Available at [http://www.ecmwf.int/publications/library/ecpublications/\\_pdf/ERA40\\_PRS\\_8.pdf](http://www.ecmwf.int/publications/library/ecpublications/_pdf/ERA40_PRS_8.pdf))
- Betts, A. K., M. Zhao, P. A. Dirmeyer, and A. C. M. Beljaars (2006), Comparison of ERA40 and NCEP/DOE near-surface datasets with other ISLSCP-II datasets, *J. Geophys. Res.*, **111**, D22S04, doi:10.1029/2006JD007174.
- Bonan, G. B., K. W. Oleson, M. Vertenstein, S. Levis, X. Zeng, Y. Dai, R. E. Dickinson, and Z.-L. Yang (2002), The land surface climatology of the Community Land Model coupled to the NCAR Community Climate Model, *J. Clim.*, **15**, 3123–3149.
- Chen, F., R. A. Pielke, and K. Mitchell (2001), Development and application of land surface models for mesoscale atmospheric models: Problems and promises, in *Land Surface Hydrology, Meteorology and Climate: Observations and Modeling, Water Sci. Appl.*, vol. 3, edited by V. Lakshmi, J. Albertson, and J. Schaake, pp. 107–135, AGU, Washington, D. C.
- Cox, P. M., et al. (1999), The impact of new land surface physics on the GCM simulation of climate and climate sensitivity, *Clim. Dyn.*, **15**, 183–203.
- 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 TN-275+STR*, 69 pp., Natl. Cent. for Atmos. Res., Boulder, Colo.
- Dirmeyer, P. A., and F. J. Zeng (1997), A two dimensional implementation of the Simple Biosphere (SiB) model, *COLA Tech. Rep. 48*, 30 pp., Cent. for Ocean-Land-Atmos. Stud., Calverton, Md.
- Dirmeyer, P. A., and F. J. Zeng (1999), An update to the distribution and treatment of vegetation and soil properties in SiB, *COLA Tech. Rep. 78*, 25 pp., Cent. for Ocean-Land-Atmos. Stud., Calverton, Md.
- Dirmeyer, P. A., Z. Guo, and X. Gao (2004), Comparison, validation and transferability of eight multi-year global soil wetness products, *J. Hydrometeorol.*, **5**, 1011–1033.
- Dirmeyer, P. A., X. Gao, M. Zhao, Z. Guo, T. Oki, and N. Hanasaki (2006), The Second Global Soil Wetness Project (GSWP-2): Multi-model analysis and implications for our perception of the land surface, *Bull. Am. Meteorol. Soc.*, in press.
- Ducharne, A., R. D. Koster, M. J. Suarez, M. Stieglitz, and P. Kumar (2000), A catchment-based approach to modeling land surface processes in a general circulation model: 1. Parameter estimation and model demonstration, *J. Geophys. Res.*, **105**, 24,823–24,838.
- EK, M. B., K. E. Mitchell, Y. Lin, P. Grunmann, E. Rogers, G. Gayno, V. Koren, and J. D. Tarpley (2003), Implementation of the upgraded Noah land-surface model in the National Centers for Environmental Prediction operational mesoscale Eta model, *J. Geophys. Res.*, **108**(D22), 8851, doi:10.1029/2002JD003296.
- Entin, J. K., A. Robock, K. Y. Vinnikov, V. Zabelin, S. Liu, and A. Namkhay (1999), Evaluation of Global Soil Wetness Project soil moisture simulations, *J. Meteorol. Soc. Jpn.*, **77**, 183–198.
- Essery, R. L. H., et al. (2003), Explicit representation of subgrid heterogeneity in a GCM land surface scheme, *J. Hydrometeorol.*, **4**, 530–543.
- Etchevers, P., C. Colaz, and F. Habets (2001), Simulation of the water budget and the river flows of the Rhone basin from 1981 to 1994, *J. Hydrol.*, **244**, 60–85.
- Guo, Z.-C., P. A. Dirmeyer, Z.-Z. Hu, X. Gao, and M. Zhao (2006), Evaluation of the Second Global Soil Wetness Project soil moisture simulations: 2: Sensitivity to different meteorological forcing and implications for data quality, *J. Geophys. Res.*, **111**, D22S03, doi:10.1029/2006JD007845.
- Gusev, Y. M., and O. N. Nasonova (1998), The land surface parameterization scheme Swap: Description and partial validation, *Global Planet. Change*, **19**, 63–86.
- Gusev, Y. M., and O. N. Nasonova (2000), An experience of modelling heat and water exchange at the land surface on a large river basin scale, *J. Hydrol.*, **233**, 1–18.
- Gusev, Y. M., and O. N. Nasonova (2002), The simulation of heat and water exchange at the land-atmosphere interface for the boreal grassland by the land-surface model SWAP, *Hydrol. Processes*, **16**, 1893–1919.
- Gusev, Y. M., and O. N. Nasonova (2003), The simulation of heat and water exchange in the boreal spruce forest by the land-surface model SWAP, *J. Hydrol.*, **280**, 162–191.
- Hall, F. G., E. Brown de Colstoun, G. J. Collatz, D. Landis, P. Dirmeyer, A. Betts, G. Huffman, L. Bounoua, M. Bosilovich, and B. Meeson (2006), The ISLSCP Initiative II global data sets: Surface boundary conditions and atmospheric forcings for land-atmosphere studies, *J. Geophys. Res.*, doi:10.1029/2006JD007366, in press.
- Huffman, G. J., R. F. Adler, P. Arkin, A. Chang, R. Ferraro, A. Gruber, J. Janowiak, A. McNab, B. Rudolf, and U. Schneider (1997), The Global Precipitation Climatology Project (GPCP) combined precipitation dataset, *Bull. Am. Meteorol. Soc.*, **78**, 5–20.
- International GEWEX Project Office (2002), The Second Global Soil Wetness Project science and implementation plan, *IGPO Publ. Ser. 37*, 69 pp., Silver Spring, Md.
- Kagan, R. (1979), *Averaging of Meteorological Fields* (in Russian), Gidrometeoizdat, St. Petersburg, Russia. (English translation, Springer, New York, 1997.)
- 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**, 1631–1648.
- Koster, R. D., and P. C. D. Milly (1997), The interplay between transpiration and runoff formulations in land surface schemes used with atmospheric models, *J. Clim.*, **10**, 1578–1591.
- Koster, R. D., and M. J. Suarez (1992), Modeling the land surface boundary in climate models as a composite of independent vegetation stands, *J. Geophys. Res.*, **97**, 2697–2716.
- Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar (2000), A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure, *J. Geophys. Res.*, **105**, 24,809–24,822.
- Li, H., A. Robock, S. Liu, X. Mo, and P. Viterbo (2005), Evaluation of reanalysis soil moisture simulations using updated Chinese soil moisture observations, *J. Hydrometeorol.*, **6**, 180–193.
- Lu, C.-H., M. Kanamitsu, J. O. Roads, W. Ebisuzaki, and K. E. Mitchell (2005), Evaluation of soil moisture in the NCEP-NCAR and NCEP-DOE global reanalyses, *J. Hydrometeorol.*, **6**, 391–408.
- Manabe, S. (1969), Climate and the ocean circulation: 1. The atmospheric circulation and the hydrology of the Earth's surface, *Mon. Weather Rev.*, **97**, 739–774.
- Milly, P. C. D., and A. B. Shmakin (2002a), Global modeling of land water and energy balances: Part I. The land dynamics (LaD) model, *J. Hydrometeorol.*, **3**, 283–299.

- Milly, P. C. D., and A. B. Shmakin (2002b), Global modeling of land water and energy balances: Part II: Land-characteristic contributions to spatial variability, *J. Hydrometeorol.*, **3**, 301–310.
- Milly, P. C. D., A. Cazenave, and C. Gennero (2003), Contribution of climate-driven change in continental water storage to recent sea-level rise, *Proc. Natl. Acad. Sci. U. S. A.*, **100**(23), 13,151–13,161.
- New, M., M. Hulme, and P. Jones (1999), Representing twentieth-century space-time climate variability: Part I. Development of a 1901–96 mean monthly terrestrial climatology, *J. Clim.*, **12**, 829–856.
- New, M., M. Hulme, and P. Jones (2000), Representing twentieth-century space-time climate variability: Part II. Development of 1901–96 monthly grids of terrestrial surface climate, *J. Clim.*, **13**, 2217–2238.
- Niu, G.-Y., Z.-L. Yang, R. E. Dickinson, and L. E. Gulden (2005), A simple TOPMODEL-based runoff parameterization (SIMTOP) for use in global climate models, *J. Geophys. Res.*, **110**, D21106, doi:10.1029/2005JD006111.
- Oki, T., T. Nishimura, and P. Dirmeyer (1999), Assessment of annual runoff from land surface models using Total Runoff Integrating Pathways (TRIP), *J. Meteorol. Soc. Jpn.*, **77**, 235–255.
- Robock, A., K. Y. Vinikov, C. A. Schlosser, N. A. Speranskaya, and Y. Xue (1995), Use of midlatitude soil moisture and meteorological observations to validate soil moisture simulations with biosphere and bucket models, *J. Clim.*, **8**, 15–35.
- Robock, A., C. A. Schlosser, K. Y. Vinnikov, N. A. Speranskaya, J. K. Entin, and S. Qiu (1998), Evaluation of the AMIP soil moisture simulations, *Global Planet. Change*, **19**, 181–208.
- Robock, A., K. Y. Vinnikov, G. Srinivasan, J. K. Entin, S. E. Hollinger, N. A. Speranskaya, S. Liu, and A. Namkhai (2000), The global soil moisture data bank, *Bull. Am. Meteorol. Soc.*, **81**, 1281–1299.
- Rudolf, B., H. Hauschild, W. Reuth, and U. Schneider (1994), Terrestrial precipitation analysis: Operational method and required density of point measurements, in *Global Precipitation and Climate Change*, NATO ASI Ser. I, vol. 26, edited by M. Desbois and F. Desalmond, pp. 173–186, Springer, New York.
- Sellers, P. J., Y. Mintz, Y. C. Sud, and A. Dalcher (1986), A simple biosphere model (SiB) for use within general circulation models, *J. Atmos. Sci.*, **43**, 505–531.
- Sellers, P. J., et al. (1997), Modeling the exchanges of energy, water, and carbon between continents and the atmosphere, *Science*, **275**, 502–509.
- Shmakin, A. B., P. C. D. Milly, and K. A. Dunne (2002), Global modeling of land water and energy balances: Part III. Interannual variability, *J. Hydrometeorol.*, **3**, 311–321.
- Shukla, J., and Y. Mintz (1982), Influence of land-surface evapotranspiration on the Earth's climate, *Science*, **215**, 1498–1501.
- Simmons, A. J., and J. K. Gibson (2000), The ERA-40 Project Plan, *ERA-40 Proj. Rep. Ser. 1*, 63 pp., Eur. Cent. for Medium-Range Weather Forecasts, Reading, U.K. (Available at [http://www.ecmwf.int/publications/library/ecpublications/pdf/ERA40\\_PRS\\_1.pdf](http://www.ecmwf.int/publications/library/ecpublications/pdf/ERA40_PRS_1.pdf))
- Stackhouse, P. W., Jr., S. K. Gupta, S. J. Cox, M. Chiacchio, and J. C. Mikovitz (2000), The SRB Project Release 2 data set: An update, *WCRP GEWEX News*, **10**(3), 4.
- Stieglitz, M., A. Ducharme, R. D. Koster, and M. J. Suarez (2001), The impact of detailed snow physics on the simulation of snow cover and subsurface thermodynamics at continental scales, *J. Hydrometeorol.*, **2**, 228–242.
- Swenson, S., J. Wahr, and P. C. D. Milly (2003), Estimated accuracies of regional water storage variations inferred from the Gravity Recovery and Climate Experiment (GRACE), *Water Resour. Res.*, **39**(8), 1223, doi:10.1029/2002WR001808.
- Xue, Y., P. J. Sellers, J. L. Kinter, and J. Shukla (1991), A simplified biosphere model for global climate studies, *J. Clim.*, **4**, 345–364.
- Xue, Y., F. J. Zeng, and C. A. Schlosser (1996), SSiB and its sensitivity to soil properties: A case study using HAPEX-Mobilhy data, *Global Planet. Change*, **13**, 183–194.
- Yang, Z.-L., and G.-Y. Niu (2003), The versatile integrator of surface and atmosphere processes (VISA): Part I: Model description, *Global Planet. Change*, **38**, 175–189.
- Zhao, M., and P. Dirmeyer (2003), Production and analysis of GSWP-2 near-surface meteorology data sets, *COLA Tech. Rep. 159*, 22 pp., Cent. for Ocean-Land-Atmos. Stud., Calverton, Md. (Available at [ftp://grads.iges.org/pub/ctr/ctr\\_159.pdf](ftp://grads.iges.org/pub/ctr/ctr_159.pdf))

---

P. A. Dirmeyer and Z. Guo, Center for Ocean-Land-Atmosphere Studies, 4041 Powder Mill Road, Calverton, MD 20705, USA. (guo@cola.iges.org)

# Evaluation of the Second Global Soil Wetness Project soil moisture simulations:

## 2. Sensitivity to external meteorological forcing

Zhichang Guo,<sup>1</sup> Paul A. Dirmeyer,<sup>1</sup> Zeng-Zhen Hu,<sup>1</sup> Xiang Gao,<sup>1</sup> and Mei Zhao<sup>1</sup>

Received 27 July 2006; accepted 20 September 2006; published 22 November 2006.

[1] The quality of meteorological forcing data has a strong impact on the simulation of the land surface component of the hydrological cycle. In this paper, the sensitivity of soil moisture simulations to combinations of different external meteorological forcing and vegetation parameters supplied by the International Satellite Land-Surface Climatology Project (ISLSCP) Initiative II (II2) as part of the Second Global Soil Wetness Project (GSWP-2) is evaluated by using the SSiB land surface model. The simulated plant-available soil moisture in the top 1 m of soil is compared against in situ observations over grasslands and agricultural regions in the former Soviet Union, United States (Illinois), China, and Mongolia from the Global Soil Moisture Data Bank. It is found that the skill of the simulations is very sensitive to the source of meteorological forcing and that hybridization of reanalysis products with observational data substantially improves the soil moisture simulations compared to reanalysis data alone. Sensitivity is highest among products of precipitation, radiation, and vegetation class. The range of changes in the skill of soil moisture simulations with the same land surface model in 13 different sensitivity studies is as large as that resulting from 11 different land surface models driven with the same meteorological forcing from the baseline integrations of GSWP-2. Assuming differences among versions of meteorological forcing fields are indicative of uncertainties in our knowledge of these drivers of the land surface climate, the impact of that uncertainty on land surface hydrology is as large as that from the variations among land surface models.

**Citation:** Guo, Z., P. A. Dirmeyer, Z.-Z. Hu, X. Gao, and M. Zhao (2006), Evaluation of the Second Global Soil Wetness Project soil moisture simulations: 2. Sensitivity to external meteorological forcing, *J. Geophys. Res.*, *111*, D22S03, doi:10.1029/2006JD007845.

## 1. Introduction

[2] Soil moisture is an important state variable for studying the role played by the hydrological cycle in the overall functioning of atmospheric and climate processes for many reasons [Robock *et al.*, 1998]. First, the soil moisture affects latent heat fluxes, and hence controls the partitioning of available surface energy between the sensible and latent heat fluxes, with a strong effect on the resulting surface temperature; second, it influences the horizontal fluxes of moisture, namely runoff. Excess of soil moisture is a precursor for flooding. Third, soil moisture availability is crucial for crop growth, and a deficiency of soil moisture is tantamount to drought and increasing possibility of forest fire occurrences. Finally, soil moisture provides a long “memory” of past states for the climate system over the land beyond the timescale of deterministic predictability for weather. All these features make soil moisture one of the most important land surface variables for weather and subseasonal prediction, climate modeling, water resource management, and agricultural applications.

[3] All state-of-the-art atmospheric general circulation models (AGCMs) are coupled with a land surface model to calculate vertical fluxes of moisture, energy, and momentum, and to represent the state of the land surface. Soil moisture becomes an indispensable land surface variable for model initial conditions from which global weather and climate models begin integrations, and a state variable of model boundary conditions for those dated AGCMs which specified soil moisture with pregenerated soil moisture data sets. Moreover, a number of modeling studies have elucidated the sensitivity of climate simulations and weather forecasts to soil moisture anomalies [e.g., Shukla and Mintz, 1982; Fennessy and Shukla, 1999; Chen *et al.*, 2001; Dirmeyer, 2001; Douville, 2003]. More recently, Koster *et al.* [2004] used a multimodel approach to identify hot spot regions where precipitation might be strongly affected by soil moisture anomalies during the boreal summer. Pal and Eltahir [2001] used observations over the midwestern United States to illustrate the pathways and mechanisms through which soil moisture affects precipitation. Taylor and Ellis [2006] demonstrate observational evidence of a positive feedback between soil moisture and rainfall in semiarid conditions. All these studies suggest that soil moisture plays an important role in the weather and climate system, and knowledge of the state of soil moisture may be essential for improving climate prediction on seasonal-

<sup>1</sup>Center for Ocean-Land-Atmosphere Studies, Calverton, Maryland, USA.

to-interannual timescales [AMS Council, 2001] when considering the performance of AGCMs and land surface models.

[4] There is no global in situ observational network, nor complete estimates from satellite remote sensing for soil moisture. Thus generation of long-term global estimates of soil moisture must be accomplished using land surface models to perform either offline integrations driven by externally specified near-surface meteorological forcing or coupled simulations with atmospheric input from the host AGCM. In either case, the accuracy of soil moisture simulations is affected by the quality of the atmospheric forcing and in the efficacy of the land surface models, and thus may vary significantly from time to time and from place to place in the model domain [Lu et al., 2005]. The second Global Soil Wetness Project (GSWP-2), a multi-institutional modeling activity to produce a complete multi-model set of land surface state variables and fluxes, provides a number of different versions of meteorological forcing and vegetation parameters generated from data sets archived by the International Satellite Land Surface Climatology Project II (IIS2 [Hall et al., 2006]). In a companion paper [Guo and Dirmeyer, 2006], we examined the uncertainty of soil moisture simulations among eleven different land surface models participating in GSWP-2. In this paper, we evaluate the response of soil moisture simulation to changes in forcing data and surface parameters. We also compare the magnitude of uncertainty of soil moisture simulations generated by different land surface models with that resulting from uncertainties in meteorological forcing.

[5] It is natural to assume that a land surface model that produces a better simulation of terrestrial processes also produces more accurate time series of soil moisture simulations. Similarly, one may assume that for the same land surface model, a set of meteorological forcing of improved quality will generally produce more accurate soil moisture simulations. Thus evaluation of sensitivity of soil moisture simulations to external meteorological forcing in this paper provides an indirect quality check and intercomparison of IIS2 data sets.

[6] Since long time series of operational analyses are full of discontinuities and other inconsistencies (as the underlying model changes over time), retrospective analyses (reanalyses) with a frozen model and assimilation framework are used as the main source of meteorological forcing for GSWP-2 models. Reanalyses have the added benefit of being able to assimilate additional observations that are not normally available in near-real-time. However, the reanalyses contain systematic errors compared to observations because they strongly reflect the biases and errors in the underlying AGCMs and assimilation schemes. Similar to many previous studies [Liston et al., 1993; Meeson et al., 1995; Schnur and Lettenmaier, 1997; Dirmeyer et al., 1999; Dirmeyer and Tan, 2001; Berg et al., 2003; Hirabayashi et al., 2005; Sheffield et al., 2006; Qian et al., 2006], GSWP-2 uses a hybridization of reanalyses and gridded observational data to minimize the systematic errors in the reanalyses [Zhao and Dirmeyer, 2003]. It is noted that the reanalyses also incorporate observational data with their assimilation systems, but they are basically still modeling products with strong dependency on the underlying AGCMs, particularly for flux terms. In addition, in situ precipitation observations are not directly assimilated into the reanalysis products. As

discussed by Dirmeyer et al. [2006] and Zhao and Dirmeyer [2003] in more detail, the hybridization used for GSWP-2 combines the high time resolution and complete spatial coverage of the reanalysis with the ground truth archived in the monthly observational data sets. This effectively constrains time-averaged precipitation and near-surface air temperature to observations but includes the short-timescale variability of these variables from the reanalysis. Furthermore, evaluation of the sensitivity of soil moisture simulations to external meteorological forcing in this paper provides an opportunity to assess the impact of hybridization used in meteorological forcing on soil moisture simulations. In section 2, we describe the GSWP-2 sensitivity experiments, the different versions of meteorological forcing used for model simulations and the SSiB land surface model. A brief introduction of the in situ observations used for evaluation is also included. An evaluation of the soil moisture simulations is presented in section 3. Conclusions are given in section 4.

## 2. Land Surface Model, Experiments, and Evaluation Data

### 2.1. GSWP-2 Sensitivity Experiments

[7] GSWP-2 provides several versions of meteorological forcing variables and land surface parameters to allow the participating groups to examine the role of uncertainty in the external forcing data and specification of land surface properties in model simulations of surface state variables and fluxes [International GEWEX Project Office, 2002]. A baseline control experiment and four categories of sensitivity experiments are designed to test the specific response of the models to changes in precipitation, radiation, near-surface meteorology, and vegetation data sets, respectively. Table 1 lists the meteorological forcing data used in each sensitivity experiment and its comparison with the baseline experiment.

[8] Precipitation is the major driver for the terrestrial hydrology. Precipitation data from the European Centre for Medium-Range Weather Forecasts (ECMWF) 40-year reanalysis (ERA-40) [Simmons and Gibson, 2000; Betts and Beljaars, 2003] are used in P1 and PE while data from the National Centers for Environmental Prediction/Department of Energy (NCEP/DOE) reanalysis [Kanamitsu et al., 2002] are used in P4 and the baseline experiment (B0). These integrations show the consequences of the choice of reanalysis rainfall on the simulation of the surface water balance, and the relative merits of precipitation statistics of NCEP/DOE and ERA-40 reanalysis also can be indirectly examined. In P1 and P4, the precipitation data are pure reanalysis products, without any observational data used for adjustment. Comparison to PE and B0 shows the impact of hybridization on the surface hydrologic simulation, and allows for an evaluation of reanalysis precipitation errors. In P2 and P3, precipitation is hybridized with the observational gauge data from the Global Precipitation Climatology Centre (GPCC [Rudolf et al., 1994]), but without the relaxation to the satellite-estimated precipitation from the Global Precipitation Climatology Project (GPCP [Huffman et al., 1997]) in regions of low gauge density as applied in the baseline integration. The impact of remote sensing data on estimates of global surface hydrologic cycle can be



**Table 1.** Description of GSWP-2 Sensitivity Experiments and Comparison With Baseline Experiment<sup>a</sup>

Name	Description	Comparable Data Used in B0
B0	baseline integration (the meteorological forcing and vegetation data sets used in the sensitivity integration are the same as the B0 baseline experiment except the variables indicated below)	see <i>Guo and Dirmeyer</i> [2006, Table 2] for meteorological forcing used in B0
M1	all NCEP/DOE meteorological data without any hybridization with observational data	all NCEP/DOE meteorological data hybridized with observational data
M2	all ECMWF meteorological data without any hybridization with observational data.	same as above
P1	precipitation from ERA-40 without hybridization with observational data	precipitation from NCEP/DOE hybridized with GPCC and GPCP [ <i>Rudolf et al.</i> , 1994; <i>Huffman et al.</i> , 1997] and corrected for wind-caused gauge undercatch
PE	ERA-40 precipitation hybridized with GPCC gridded gauge analysis	same as above
P2	NCEP/DOE precipitation hybridized with GPCC gauge analysis and corrected for wind-caused gauge undercatch	same as above
P3	NCEP/DOE precipitation hybridized with GPCC gauge analysis without adjustment for wind-caused gauge undercatch	same as above
P4	NCEP/DOE precipitation without hybridization with observational data	same as above
R1	radiation from NCEP/DOE reanalysis	radiation products from SRB [ <i>Stackhouse et al.</i> , 2000]
R2	radiation from ERA-40	same as above
R3	radiation from ISCCP [ <i>Rosow and Zhang</i> , 1995]	same as above
V1	University of Maryland vegetation class data [ <i>Hansen et al.</i> , 2000]	IGBP vegetation class data [ <i>Loveland et al.</i> , 2000]
I1	climatological annual cycle of vegetation parameters	observed monthly vegetation parameters

<sup>a</sup>The italicized integrations indicate the sensitivity experiments where precipitation or all the meteorological forcing data are not hybridized with observational data.

evaluated between P2 and P3. In addition, precipitation is corrected for wind-caused gauge undercatch in P2 (which is especially prevalent for snow) whereas no correction applied to P3. Comparison between P2 and P3 shows the hydrologic impact of neglecting the effect of wind on rain gauge accuracy. It should be noted that the correction for gauge undercatch in B0 is demonstrably excessive in some locations [*Dirmeyer et al.*, 2006; *Decharme and Douville*, 2006].

[9] Downwelling radiation is another driver for the land surface water cycle through its link to the energy budget. The surface radiation fields from the NCEP/DOE and ERA-40 reanalysis are used in R1 and R2, respectively whereas the satellite-based radiation products from the Surface Radiation Budget (SRB [*Stackhouse et al.*, 2000]) data set and the International Satellite Cloud Climatology Project (ISCCP [*Rosow and Zhang*, 1995]) are used for the baseline and R3 simulations respectively. Comparison among these sensitivity experiments shows the impact of the systematic errors in the reanalysis radiation and the choice of different radiation analyses on the simulation of the surface energy and water balances.

[10] In addition to precipitation and radiation, offline land surface simulation needs specification of land surface properties and other meteorological forcing data, such as near-surface air temperature, humidity, wind speed, and air pressure. The all-meteorological (M) series gives the broadest assessment as to the impact of overall differences in meteorological forcing between the NCEP/DOE and ERA-40 reanalysis. V1 and I1 are designed to test the response of land surface models to changes in vegetation boundary conditions. The global land cover classification of the International Geosphere-Biosphere Programme Data Information System (IGBP-DIS [*Loveland et al.*, 2000]) from the Earth Resources Observation and Science (EROS) Data Center (EDC) is used in the baseline (B0) experiment whereas a Moderate Resolution Imaging Spectroradiometer

(MODIS) derived land cover product from the University of Maryland [*Hansen et al.*, 2000] is used in the V1 experiment. Comparison between V1 and B0 shows the sensitivity of land surface hydrology to different vegetation data sets, and may help a modeling group decide which global vegetation data set to use with their model. The interannual vegetation case (I1) uses a fixed annual cycle of vegetation, instead of the interannually varying vegetation of all other cases. Comparison between I1 and B0 shows the sensitivity of the surface energy and water budgets to observed interannual variations in vegetation properties (leaf area index, greenness and vegetation cover fraction). In addition to assessing observed climate anomaly impacts, I1 is useful for a modeling group that is not currently using dynamic vegetation or vegetation phenology parameterizations, but is considering using or developing one for their land surface model.

[11] In the baseline experiment, the land surface model runs for a spin-up period from July 1982 to January 1986 so that the model can be initialized with a reasonably realistic climate at the beginning of 1986 that both reflects climate anomalies in the proceeding years, and is consistent with the model itself. The model is integrated over the global covering the 10-year period from 1986 to 1995 for the baseline run. The modeling sensitivity studies involve reintegrating the land surface model covering the same 10-year period. The restart file at 0000 UTC on 1 January 1986 from the baseline simulation (after completing spin-up) is used to initialize the land surface model for all sensitivity experiments except V1.

[12] The initialization strategy in GSWP-2 was chosen to allow an assessment of both the mean impact of different forcing data described in this paper and the transient response to shifts in the forcing data. This second point, not explored here, has consequences as observing networks continuously change and evolve as instruments, networks and satellite platforms come online or go offline. As

explained in the next section, the initial fields from the baseline simulations cannot be used in V1 because of differences in vegetation data for the experiments and the strategy used in the land model for discretization of soil layers. Even if we could start the V1 simulation from the baseline integration, a sudden global stepwise change in vegetation cover by replacing one data set with another has no comparable analog in the real world or observational analyses. The equilibrium land surface state will be very different because, in effect, one has a different global model. In all other sensitivity studies, the model itself remains identical.

## 2.2. SSiB Land Surface Model

[13] We used the SSiB land surface model, based on the work of *Xue et al.* [1991, 1996], *Dirmeyer and Zeng* [1997, 1999], and *Dirmeyer and Tan* [2001] for these sensitivity experiments. SSiB is the Simplified version of the Simple Biosphere Model (SiB [*Sellers et al.*, 1986]). SSiB follows the original “big leaf” philosophy behind SiB, but with some of the explicit representations of vegetation processes simplified to improve computational efficiency, and to reduce the number of parameters necessary to execute the model. A description of the offline version of SSiB, including a historical review of published research and applications, is given by *Dirmeyer and Zeng* [1997]. A number of modifications made to the previous versions are presented in this section.

[14] First, there are three soil layers in the original version of SSiB: surface, rooting, and recharge layer. In order to better simulate the vertical soil moisture profile and vertical movement of water between soil layers, the rooting layer is further broken into several sublayers. In the old version, the soil layers are hardwired with the equations of water and energy balances formulated for the specific three layers. In the new version, the soil can be divided into any number of layers, and the vertical discretization of soil layers has been generalized. In this study, we set six soil layers for both baseline and sensitivity experiments with the surface layer depth to be 3 cm thick; the additional layers are all incorporated into the rooting zone with equal root mass amount in each layer. The cumulative root profile for each vegetation type is calculated following *Gale and Grigal* [1987]. As a different vegetation data set is used in the V1 experiment and in SSiB the root zone thickness is a function of vegetation class, the vertical discretization of soil layers in V1 is different from the B0 run in many locations. As a result, this is the only occasion where we must reinitialize and spin-up the SSiB model. Second, in this new version, the soil water infiltration, percolation, diffusion, and drainage among soil layers are calculated with a semi-implicit scheme proposed by *Viterbo and Beljaars* [1995]. Third, a temporal disaggregation method is used for interpolating precipitation forcing to improve partitioning of infiltration and runoff when forcing data interval is large compared to the typical length of a convective rain event [*Dirmeyer and Zhao*, 2004]. Fourth, the river flow routing model developed by *Arora and Boer* [1999] is integrated into SSiB. Finally, new generalized drivers and data management software have been developed to facilitate controlling model input and output and maintaining as well as developing future code implementations.

## 2.3. In Situ Observations

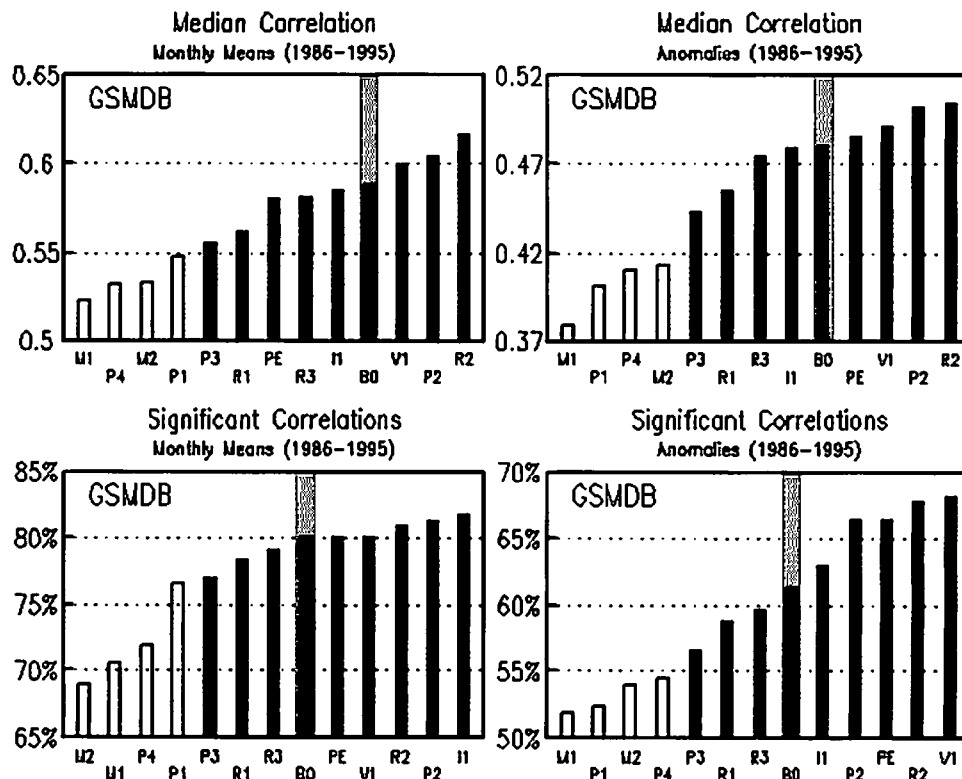
[15] The Global Soil Moisture Data Bank (GSMDB) provides soil moisture observations with long temporal records across the Northern Hemisphere midlatitudes, including North America, Europe and Asia. These data have been extensively used to evaluate spatial and temporal variability of soil moisture in nature [*Vinnikov et al.*, 1996; *Entin et al.*, 2000], and the ability of land surface models and AGCMs to simulate the observed characteristics of soil moisture [*Vinnikov and Yesserkepova*, 1991; *Robock et al.*, 1995; *Schlosser et al.*, 1997, 2000; *Yang et al.*, 1997; *Entin et al.*, 1999; *Dirmeyer et al.*, 2004; *Li et al.*, 2005; *Lu et al.*, 2005]. In this paper, we focus on station data in four regions (Illinois, USA; China; Mongolia and the former Soviet Union). The former Soviet Union (FSU) data are further divided into two categories, representing winter and spring cereal fields. The updated Chinese soil moisture observations [*Li et al.*, 2005] are used in this and the companion paper. Figure 1 in the companion paper shows a map of the station distribution for the Global Soil Moisture Data Bank [*Guo and Dirmeyer*, 2006]. These data sets span decades, although individual stations may have a much shorter record of observations.

[16] As in the companion paper, the columnar plant-available soil moisture for the top 1-m soil layer preprocessed by *Robock et al.* [2000] is used for the evaluation, and the instantaneous values are aggregated to get the monthly mean observations. No interpolation is applied to fill missing values, and data are used only when and where they are available.

## 3. Results

[17] Model outputs from the SSiB simulations for the GSWP-2 baseline and sensitivity experiments have been compared to GSMDB soil moisture observational data over China, USA (Illinois), Mongolia, and FSU. We have chosen in this study the plant-available soil moisture rather than total soil moisture to make these results comparable to those in the companion paper. The time series of the columnar plant-available soil moisture for the top 1-m soil layer is retrieved from the model simulations at the regular  $1^\circ \times 1^\circ$  latitude-longitude grid box that contains a particular station. The columnar plant-available soil moisture is calculated as a summation of soil moisture at each soil layer while subtracting the wilting point at that layer. The data are converted to columnar plant-available soil moisture within the top 1-m according to the soil layer depths at the specific model grid. If the model soil depth is shallower than 1 m, the deep layer plant-available soil moisture amount is extrapolated to fill in the additional depth. If the model soil depth is deeper than 1 m, only the columnar soil moisture summation for the top 1 m is calculated.

[18] Station counts over the domains are: China (40 stations), Illinois (19 stations), Mongolia (42 stations) and FSU (171 stations for each of the two categories). Those stations where more than 75% of the observational data are missing during the evaluation period are excluded. After applying the missing data threshold, the percentages of usable stations are as follows: China (90%), Illinois (100%), Mongolia (65%), FSU (46% spring cereal; 47% winter cereal). There are few measurements during winter



**Figure 1.** Median temporal correlation and significant counts between observed and simulated soil moisture for the time series of (left) monthly mean and (right) anomalies covering the period 1986–1995 for all stations regardless of regions. In addition, the products are shown in ranked order from lowest to highest score. All sensitivity and baseline integrations are included with the baseline run indicated by a shaded background. The sensitivity integrations where precipitation or all meteorological forcing are not hybridized with observational data are shown as open rectangles.

over FSU and Mongolia because of frozen ground, greatly reducing the number of useable stations for those regions.

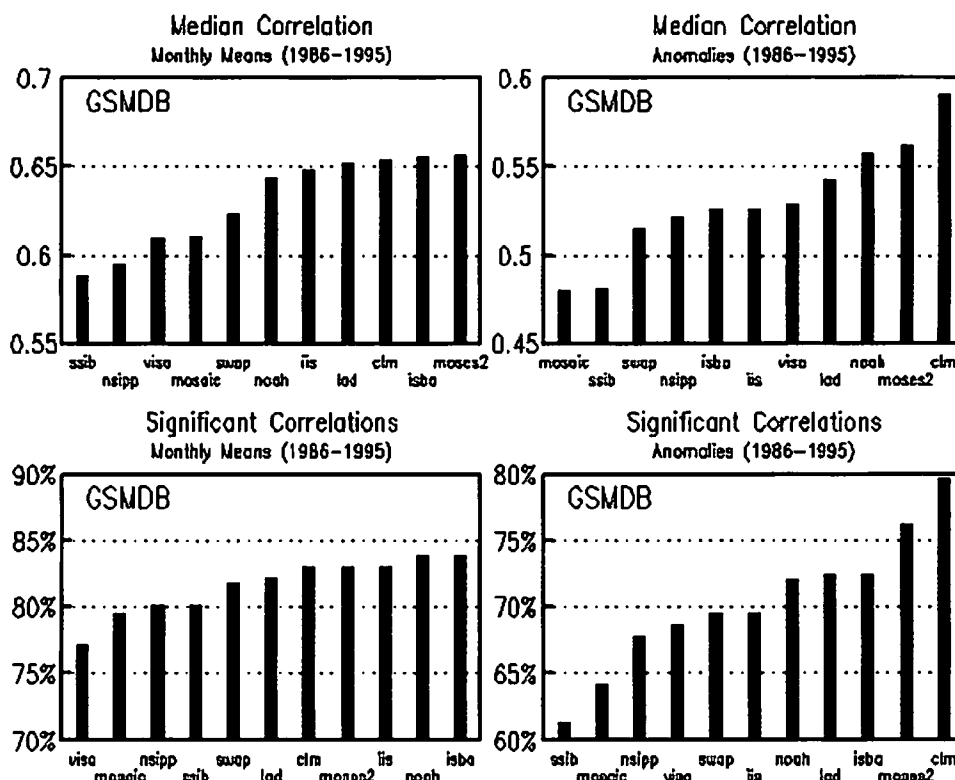
[19] Using available in situ soil moisture observations, we examine the model's capability to simulate the proper phasing of the annual cycle and interannual variability. The sensitivity experiments are a platform to examine the impact on these metrics of the changes in meteorological forcing and surface parameters.

[20] The median correlation between simulated soil moisture and observations used in the companion paper [Guo and Dirmeyer, 2006] and the work of Dirmeyer *et al.* [2004] provides an effective measure of the model's overall performance in simulating observed temporal variability of soil moisture over a number of stations scattered around the globe. The fraction of stations where the significance of temporal correlations is above a certain threshold gives another measure of model's skill. Figure 1 presents a comparison among baseline and sensitivity experiments of the median correlations between the model estimates and observations of soil moisture for both the total series and the anomalies, as well as the fraction of stations where the significance level is higher than 95%. The legend "GSMDB" in the figure indicates that all stations with less than 75% missing data during the 10-year period (1986–1995) are considered. Since an identical version of the model is used for all baseline and sensitivity experiments, the differences of model's performance in Figure 1 reflect

the model's response to the differences in the meteorological forcing and surface parameters among these experiments.

[21] Figure 1 allows for a few notable observations about overall performance. First, there is a clear indication that the quality of simulated soil moisture is significantly improved when the reanalysis data are hybridized with observational data sets. This is true for the median correlations and fraction of significant counts for all baseline and sensitivity experiments when all stations are considered regardless regions. In P1 and P4, the precipitation forcing uses the original precipitation from the reanalyses and in M1 and M2, all meteorological forcing data are pure reanalysis. P1, P4, M1, and M2 are the only experiments where precipitation has not been hybridized with observational data, and they rank the poorest in the median correlation and significance counts. This shows that the hybridization procedure generally removes systematic errors in the reanalysis fields, and improves the model estimates of soil moisture.

[22] Second, changes in all three fields of precipitation, radiation and vegetation type have strong impacts on the quality of simulated soil moisture. When the hybridization procedure is applied, changes in the meteorological forcing result in a comparable magnitude of changes in the median correlation and significance counts among the radiation series (R1–R3), precipitation series (P2, P3 and PE) of sensitivity experiments; the range of the vegetation type



**Figure 2.** Median temporal correlation and significant counts between observed and simulated soil moisture for the time series of (left) monthly mean and (right) anomalies covering the period 1986–1995 for all stations regardless of regions. In addition, the products are shown in ranked order from lowest to highest score. Baseline integrations by eleven participating land surface models in GSWP-2 are included (see Guo and Dirmeyer [2006] for details).

series (V1 and I1) tends to be smaller. These results indicate that the skill of soil moisture simulations is sensitive to the quality of precipitation, radiation, and vegetation classification. Third, the products resulting in a better simulation of phasing of seasonal cycle (with higher median correlation for the monthly means) may not necessarily produce a better simulation of interannual variability of soil moisture (with higher median correlation for the anomalies). Fourth, considering the skill in the median correlations and significant counts, V1, PE, P2, and R2 tend to produce a better overall (among all regions) quality in simulated soil moisture.

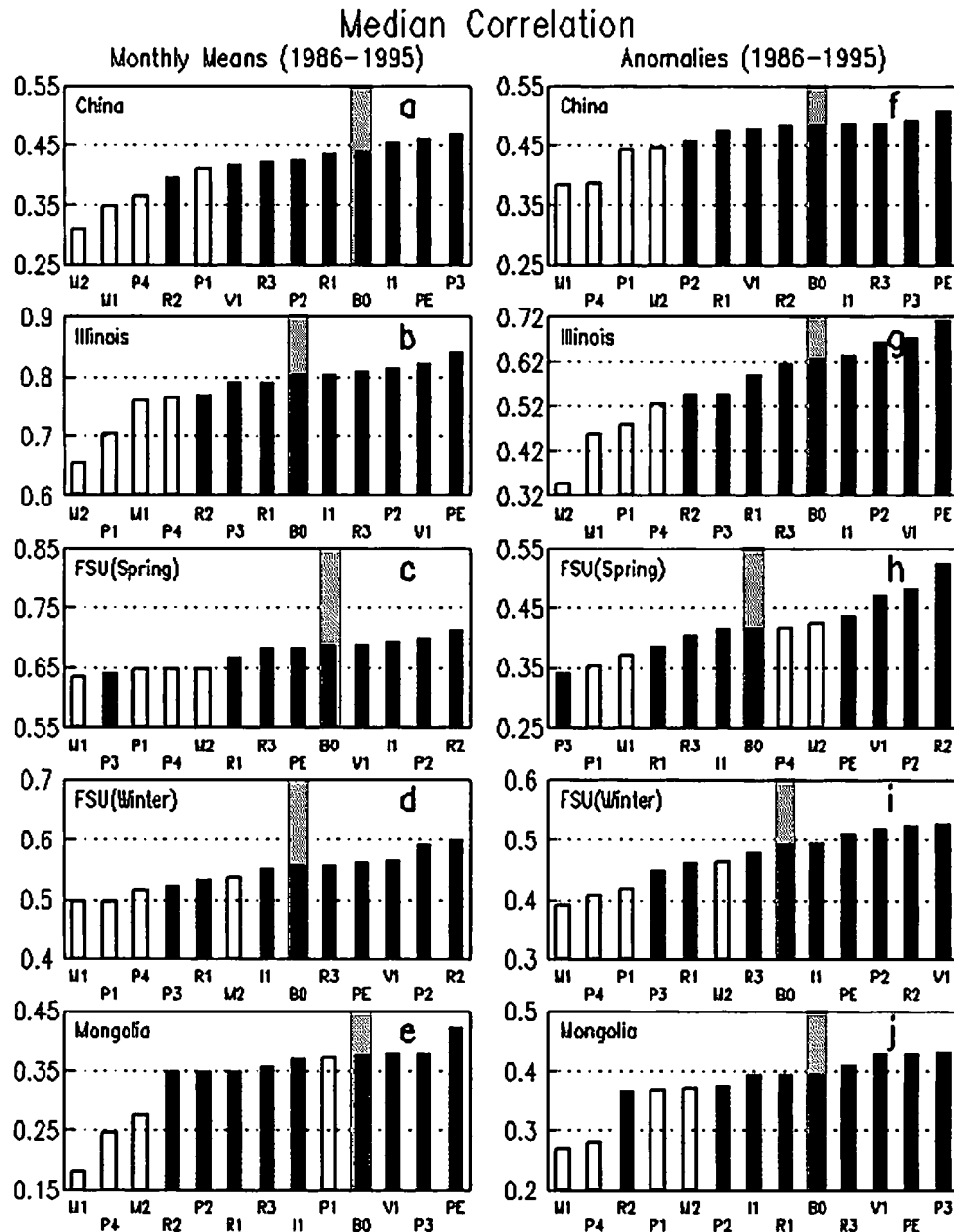
[23] Figure 2 shows the counterpart of Figure 1 for intermodel comparison (a reproduction of Guo and Dirmeyer [2006, Figures 2f, 3f, and 4f]). The range of skill changes observed in soil moisture simulations with the same land surface model driven with different external forcing is large considering the intermodel comparison regarding model's performance presented in the companion paper [Guo and Dirmeyer, 2006]. It is observed that the range of skill changes among sensitivity experiments with the same land surface model is as large as that resulting from different land surface models driven with the same meteorological forcing. The best model in Figure 2 simulates soil moisture with 7% and 19% more stations significantly correlated with observations than the poorest model for the total series and anomalies, respectively. In Figure 1, the case with the highest significance counts has 12% and 18% more significantly correlated stations than the worst case for the total series and anomalies, respectively. If differences among

versions of meteorological forcing fields are indicative of uncertainties in our knowledge of these drivers of the land surface climate, the impact of that uncertainty on land surface hydrology is as large as that resulting from the uncertainty among land surface models.

[24] Figure 3 compares median correlations between the model estimates and observations of soil moisture in terms of the total time series (Figure 3, left) and the anomalies (Figure 3, right) for stations over individual domains. Similar to what has been found in Figure 1, hybridization of reanalysis products with observational data substantially improves soil moisture simulations for most cases, and the median correlations for the sensitivity experiments P1, P4, M1, and M2 are smaller than for other experiments. The ERA-40 data set (M2) produce better soil moisture simulations than NCEP-DOE (M1) over all domains except for China for the monthly fields and Illinois, USA.

[25] Comparison of median correlation between R2 and B0 in Figure 3 shows that the improvements of soil moisture simulations in R2 found in Figure 1 only reflect the improvements over FSU (the FSU region has more stations than the other regions). In fact, the median correlations in R2 are smaller than that in baseline run over the other domains. Similarly, P2 produces smaller median correlations than B0 over both China and Mongolia whereas the overall scores produced by P2 are higher than B0 in Figure 1.

[26] Comparison of median correlations between B0 and the sensitivity experiments in Figure 3 shows that cases PE

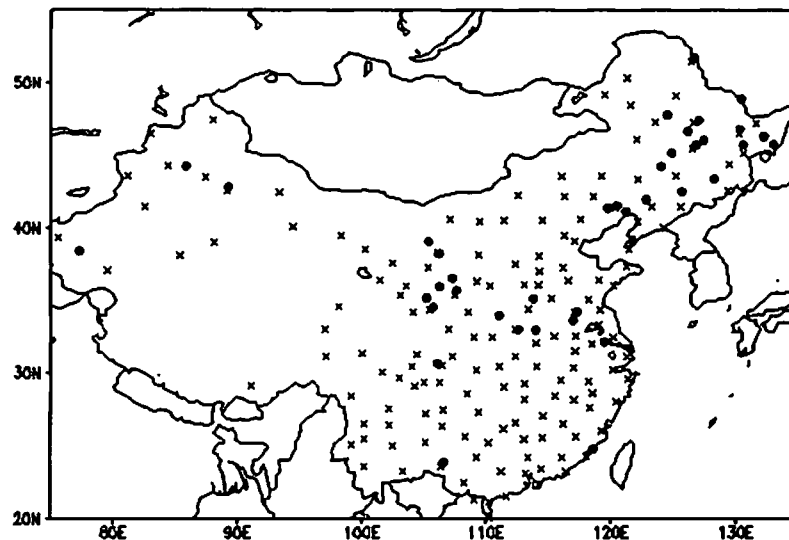


**Figure 3.** Median temporal correlations between simulated soil moisture and observations for (a–e) the monthly mean and (f–j) the anomalies over individual domains for each sensitivity and baseline run. In addition, the products are shown in ranked order, from lowest to highest correlations, based on the median values. The baseline run is indicated by a shaded background, and the sensitivity integrations where precipitation or all meteorological forcing are not hybridized with observational data are shown as open rectangles.

and V1 display improved quality of simulated soil moisture over all domains with the exception of Figures 3a, 3c and 3f, where a slight degradation occurs. The only difference in the meteorological forcing data between PE and B0 is that the ERA-40 precipitation is used in place of the NCEP-DOE precipitation. With the hybridization procedure applied to the reanalysis products, the monthly means of the two precipitation hybrid products would be the same, but the ERA-40 and NCEP-DOE products exhibit somewhat different characteristic in the diurnal cycle and synoptic variability. The differences in the skill between PE and B0

show the impact of different representations of the synoptic evolution of global weather on the surface simulation of the energy and water balances.

[27] The difference between I1 and B0 is that the 10-year mean seasonal cycle of time-varying vegetation parameters (leaf area index and greenness fraction) are used throughout the 10-year simulation in I1 whereas the observed interannual variations of vegetation are included in B0. Comparison between B0 and I1 indicates that the interannual variations of vegetation have little impact on the quality of soil moisture simulations. It is noted that in GSWP-2



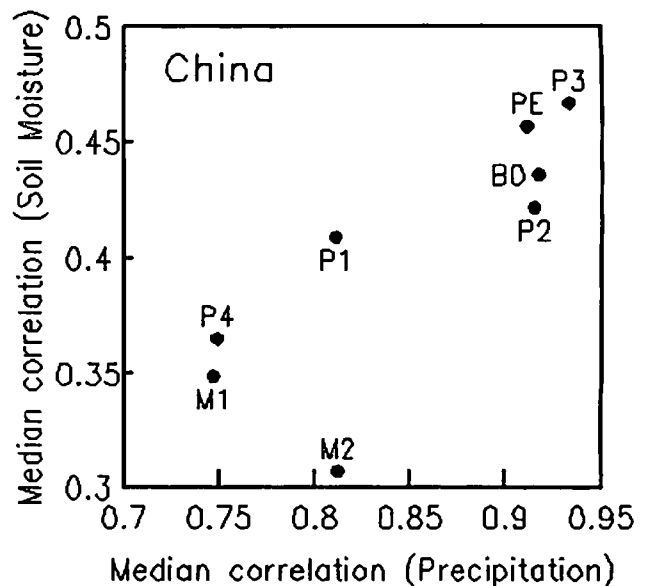
**Figure 4.** Soil moisture and precipitation observation stations for China. There are 40 stations for soil moisture (solid circle), and 160 stations for precipitation (cross).

experiments the simulated surface fluxes and state variables have no ability to feedback upon the atmospheric conditions. A different conclusion may be reached in a coupled land-atmosphere model system.

[28] As one of the most important drivers for the land surface water and energy balances, precipitation has a strong impact on soil moisture simulations. Thus we are motivated to use the observed precipitation and soil moisture data together to relate the quality of external meteorological forcing data to the resultant quality of simulated soil moisture. Figure 4 shows a map of soil moisture (40 stations indicated by closed circles) and precipitation observation stations (160 stations indicated by the exes) in China. We first calculate the median correlations between the precipitation observations over the 160 stations and the GSWP-2 precipitation data at the corresponding model grid boxes for each sensitivity, and then relate them to the median correlations between the observed soil moisture and model estimates over the 40 stations in China.

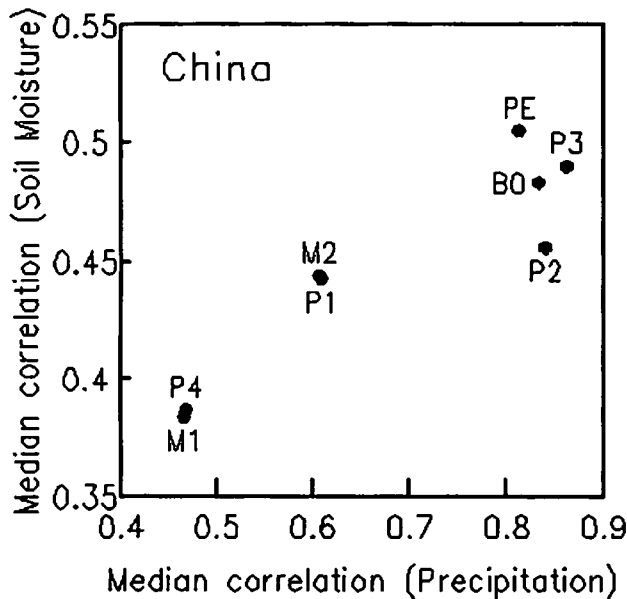
[29] Figure 5 shows the median correlations between the station observations and the GSWP-2 precipitation forcing data over the 160 stations versus median correlations between the observed soil moisture and model estimates over the 40 stations in China. The correlations are calculated from the monthly mean values. We only include experiments where precipitation or meteorological forcings are varied. If the median correlations represent the overall data quality over China for precipitation and soil moisture, Figure 5 demonstrates that the better the precipitation forcing, the better the soil moisture simulations. Figure 6 shows two clusters of experiments: M1, M2, P1, P4 and B0, P2, PE, P3. The later cluster has higher correlations than the former, and indicates that minimization of systematic errors in the reanalysis products via the hybridization procedure improves the soil moisture simulations substantially. In M1, M2, P1 and P4, precipitation data are not hybridized with observational data. Therefore P4 has the same median correlation for the precipitation as M1, and similarly P1 has the same median correlation for the precipitation as M2.

The differences in the median correlation for the soil moisture between the two pairs (P4 and M1, P1 and M2) reflect the improvements made by reducing systematic errors in the radiation, near-surface temperature and specific humidity. It seems that the ERA-40 has a better representation of precipitation over China than NCEP-DOE, but the soil moisture simulations with the original ERA-40 products are poorer than that with pure NCEP-DOE forcing due to other meteorological variables. When the hybrid NCEP-DOE products are used in the place of ERA-40 for all the meteorological forcing except the precipitation (P4 versus



**Figure 5.** Median correlations between the precipitation observations and the GSWP-2 precipitation data over the 160 stations versus median correlations between the observed soil moisture and model estimates over the 40 stations in China for the monthly fields.





**Figure 6.** Median correlations between the precipitation observations and the GSWP-2 precipitation data over the 160 stations versus median correlations between the observed soil moisture and model estimates over the 40 stations in China for the anomalies (with annual cycle removed).

P1), the soil moisture simulations suffer. When the precipitation is also hybridized with the observational data, the soil moisture simulations are further improved (e.g., PE).

[30] Figure 6 shows the counterpart of Figure 5 for the anomalies (mean annual cycle removed) of soil moisture and precipitation. Similarly, Figure 6 demonstrates that the products resulting in higher correlation with observations in the anomaly fields generally produce better simulations of soil moisture anomalies. Unlike in Figure 5, P1 has the same median correlation with M2, and P4 has the almost same median correlation with M1. A similar feature can be observed over Mongolia in Figure 3. It suggests that over both China and Mongolia, the quality of the simulation of soil moisture anomalies is largely controlled by the quality of precipitation anomalies alone whereas the quality of monthly soil moisture simulations is controlled by the quality in both precipitation and other meteorological forcing variables.

[31] In both Figures 5 and 6, precipitation after hybridization results in similar median correlations (PE, P2, P3 and B0), but they produce different qualities in the soil moisture simulations. In Figures 5 and 6, we calculate the correlations from monthly mean and anomalies of soil moisture and precipitation respectively. After applying hybridization, we would expect to have the similar median correlations among PE, P2, P3 and B0. The differences in median correlations for soil moisture can be explained by the combination of the impact of different characteristics in the diurnal cycle and synoptic variability on the simulation of surface energy and water balances, and differences between the observational data sets used for the hybridization.

#### 4. Discussion and Conclusions

[32] With the SSiB land surface model and different versions of GSWP-2 meteorological forcing data, we examined the response of soil moisture simulations to changes in external forcing data in an offline environment. The forcing data are taken from the ISLSCP Initiative II suite of gridded data sets, drawn from global reanalysis, in situ observations and satellite-based products. In comparison with gravimetric observations, it is found that the quality of simulated soil moisture is largely influenced by the external meteorological forcing data, and that hybridization of reanalysis products with observational data substantially improves soil moisture simulations. The soil moisture simulations are sensitive to the changes in precipitation, radiation, and vegetation class. The range of skill changes in soil moisture simulations for the same land surface model driven with different external forcing is as large as that resulted from different land surface models driven with the same meteorological forcing. If differences among versions of meteorological forcing fields are indicative of uncertainties in our knowledge of these drivers of the land surface climate, this indicates that the impact of that uncertainty on land surface hydrology is as large as that resulted from the uncertainty in land surface models itself. This also suggests that a fair evaluation of a land surface model's performance needs to consider the quality of external meteorological forcing data for the offline modeling, or the accuracy of the downward surface fluxes and near-surface meteorology when coupled to an AGCM. It is also found that driven with ERA-40 reanalysis, SSiB generally produces better soil moisture simulations than those driven with NCEP-DOE products. This suggests that with ERA-40 products hybridized with observational data may generate a multiyear global product of land surface state and fluxes variables with better overall quality. However, we do find regions where the quality of ERA-40 based simulations fall short of those driven by NCEP-DOE products.

[33] We also find that different factors control the quality of the simulation of the mean annual cycle of soil moisture (the mean content of water in the soil) versus the pattern of interannual variations. Over East Asia, for instance, precipitation and meteorological forcing both determine the quality of the mean soil wetness, whereas the simulation of soil moisture anomalies seems to be controlled predominantly by the precipitation.

[34] These results have implications for monitoring of the terrestrial hydrologic cycle and for land data assimilation. GSWP-2 has shown that land surface models can be used to provide quality estimates of soil moisture across space and time, and that a multimodel approach provides estimates superior to those from individual models [Dirmeyer et al., 2006; Z.-C. Guo et al., Improving the quality of simulated soil moisture with a multi-model ensemble approach, submitted to *Quarterly Journal of the Royal Meteorological Society*, 2006]. However, this study shows that the level of quality is strongly dependent on the character of the forcing data. All of the data sets used to force the land surface model are commonly used as validation or verification data for studies in the earth sciences. It may be true that a particular field from one data set can be safely declared to be inferior to another. For example, reanalysis precipitation

is likely to be less representative or reality than a gauge-based gridded analysis. However, even this assertion may not be true for every locality, as in situ observations do not exist within every grid box and interpolation in the gauge-based data set may introduce unrealistic details. This study suggests that as much effort should be put into improving the quality of global gridded data sets as is applied to model development. This conclusion is a strong validation of the philosophy behind the ISLSCP data initiatives. This study also shows a great deal of regional variation in terms of model sensitivity to data quality. More in-depth studies along this line may inform the development of targeted observational networks, thereby making the best use of limited resources. Such approaches can also optimize the effectiveness of land data assimilation to produce land-surface reanalyses and operational analyses that are more accurate and useful than what can be accomplished today.

[35] It is noted that this paper is based on the results from the SSiB land surface model. The responses and sensitivity of soil moisture simulations to external meteorological forcing may be different for different land surface models.

[36] **Acknowledgments.** We wish to thank all of the GSWP-2 participants who submitted model integrations, participated in meetings and workshops, and particularly the Inter-Comparison Center (ICC) at the University of Tokyo for collecting, checking, managing, and redistributing the large volume of data for the project. We are indebted to Randy Koster and two anonymous reviewers for their constructive comments on the earlier version of the manuscript. Much of the forcing data preparation and analysis work for GSWP-2 was supported by NASA grant NAG5-11579.

## References

- AMS Council (2001), Statement on seasonal to interannual climate prediction, *Bull. Am. Meteorol. Soc.*, **82**, 701–703.
- Arora, V. K., and G. J. Boer (1999), A variable velocity flow routing algorithm for GCMs, *J. Geophys. Res.*, **104**(D24), 30,965–30,979.
- Berg, A. A., J. S. Famiglietti, J. P. Walker, and P. R. Houser (2003), Impact of bias correction to reanalysis products on simulations of North American soil moisture and hydrological fluxes, *J. Geophys. Res.*, **108**(D16), 4490, doi:10.1029/2002JD003334.
- Betts, A. K., and A. C. M. Beljaars (2003), ECMWF ISLSCP-II near-surface dataset from ERA-40, *ERA-40 Proj. Rep. Ser.*, **8**, Eur. Cent. for Med.-Range Weather Forecasts, Reading, U. K. (Available at [http://www.ecmwf.int/publications/library/ecpublications/\\_pdf/ERA40\\_PRS\\_8.pdf](http://www.ecmwf.int/publications/library/ecpublications/_pdf/ERA40_PRS_8.pdf))
- Chen, F., R. A. Pielke, and K. Mitchell (2001), Development and application of land surface models for mesoscale atmospheric models: Problems and promises, in *Land Surface Hydrology, Meteorology and Climate: Observations and Modeling, Water Sci. and Appl. Ser.*, vol. 3, edited by V. Lakshmi, J. Albertson, and J. Schaake, pp. 107–135, AGU, Washington, D. C.
- Decharme, B., and H. Douville (2006), Uncertainties in the GSWP-2 precipitation forcing and their impacts on regional and global hydrological simulations, *Clim. Dyn.*, doi:10.1007/s00382-006-0160-6, in press.
- Dirmeyer, P. A. (2001), Climate drift in a coupled land-atmosphere model, *J. Hydrometeorol.*, **2**, 89–100.
- Dirmeyer, P. A., and L. Tan (2001), A multi-decadal global land-surface data set of state variable and fluxes, *COLA Tech. Rep. 102*, 43 pp., Cent. for Ocean-Land-Atmos. Stud., Calverton, Md. (Available at [http://grads.iges.org/pubs/ctr\\_102.pdf](http://grads.iges.org/pubs/ctr_102.pdf))
- Dirmeyer, P. A., and F. J. Zeng (1997), A two dimensional implementation of the Simple Biosphere (SiB) model, *COLA Tech. Rep. 48*, 30 pp., Cent. for Ocean-Land-Atmos. Stud., Calverton, Md.
- Dirmeyer, P. A., and F. J. Zeng (1999), An update to the distribution and treatment of vegetation and soil properties in SSiB, *COLA Tech. Rep. 78*, 25 pp., Cent. for Ocean-Land-Atmos. Stud., Calverton, Md.
- Dirmeyer, P. A., and M. Zhao (2004), Flux replacement as a method to diagnose coupled land-atmosphere model feedback, *J. Hydrometeorol.*, **5**, 1034–1048.
- Dirmeyer, P. A., A. J. Dolman, and N. Sato (1999), The Global Soil Wetness Project: A pilot project for global land surface modeling and validation, *Bull. Am. Meteorol. Soc.*, **80**, 851–878.
- Dirmeyer, P. A., Z. Guo, and X. Gao (2004), Comparison, validation and transferability of eight multi-year global soil wetness products, *J. Hydrometeorol.*, **5**, 1011–1033.
- Dirmeyer, P. A., X. Gao, M. Zhao, Z. Guo, T. Oki, and N. Hanasaki (2006), The Second Global Soil Wetness Project (GSWP-2): Multi-model analysis and implications for our perception of the land surface, *Bull. Am. Meteorol. Soc.*, in press.
- Douville, H. (2003), Assessing the influence of soil moisture on seasonal climate variability with AGCMs, *J. Hydrometeorol.*, **4**, 1044–1066.
- Entin, J. K., A. Robock, K. Y. Vinnikov, V. Zabelin, S. Liu, and A. Namkhai (1999), Evaluation of Global Soil Wetness Project soil moisture simulations, *J. Meteorol. Soc. Jpn.*, **77**, 183–198.
- Entin, J. K., A. Robock, K. Y. Vinnikov, S. E. Hollinger, S. Liu, and A. Namkhai (2000), Temporal and spatial scales of observed soil moisture variations in the extratropics, *J. Geophys. Res.*, **105**, 11,865–11,877.
- Fennessy, M. J., and J. Shukla (1999), Impact of initial soil wetness on seasonal atmospheric prediction, *J. Clim.*, **12**, 3167–3180.
- Gale, M. R., and D. F. Grigal (1987), Vertical root distribution of northern tree species in relation to successional status, *Can. J. For. Res.*, **17**, 829–834.
- Guo, Z.-C., and P. A. Dirmeyer (2006), Evaluation of Second Global Soil Wetness Project soil moisture simulations: I. Intermodel comparison, *J. Geophys. Res.*, **111**, D22S02, doi:10.1029/2006JD007233.
- Hall, F. G., E. Brown de Colstoun, G. J. Collatz, D. Landis, P. Dirmeyer, A. Betts, G. 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.*, doi:10.1029/2006JD007366, in press.
- 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.
- Hirabayashi, Y., S. Kanae, I. Struthers, and T. Oki (2005), A 100-year (1901–2000) global retrospective estimation of the terrestrial water cycle, *J. Geophys. Res.*, **110**, D19101, doi:10.1029/2004JD005492.
- Huffman, G. J., R. F. Adler, P. Arkin, A. Chang, R. Ferraro, A. Gruber, J. Janowiak, A. McNab, B. Rudolf, and U. Schneider (1997), The Global Precipitation Climatology Project (GPCP) combined precipitation dataset, *Bull. Am. Meteorol. Soc.*, **78**, 5–20.
- International GEWEX Project Office (2002), The Second Global Soil Wetness Project science and implementation plan, *IGPO Publ. Ser. 37*, 69 pp., Silver Spring, Md.
- 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**, 1631–1648.
- Koster, R. D., et al. (2004), Regions of strong coupling between soil moisture and precipitation, *Science*, **305**, 1138–1140.
- Li, H., A. Robock, S. Liu, X. Mo, and P. Viterbo (2005), Evaluation of reanalysis soil moisture simulations using updated Chinese soil moisture observations, *J. Hydrometeorol.*, **6**, 180–193.
- Liston, G. E., Y. C. Sud, and G. K. Walker (1993), Design of a global soil moisture initialization procedure for the Simple Biosphere model, *NASA Tech. Memo. 10,4590*, 138 pp., NASA Goddard Space Flight Cent., Greenbelt, Md.
- Loveland, E. R., B. C. Reed, J. F. Brown, D. O. Ohlen, Z. Zhu, L. Yang, and J. W. Merchant (2000), Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data, *Int. J. Remote Sens.*, **21**, 1303–1330.
- Lu, C.-H., M. Kanamitsu, J. O. Roads, W. Ebisuzaki, K. E. Mitchell, and D. Lohmann (2005), Evaluation of soil moisture in the NCEP-NCAR and NCEP-DOE global reanalyses, *J. Hydrometeorol.*, **6**, 391–408.
- Meeson, B. W., F. E. Corprew, J. M. P. McManus, D. M. Myers, J. W. Closs, K. J. Sun, D. J. Sunday, and P. J. Sellers (1995), ISLSCP Initiative I—Global data sets for land-atmosphere models, 1987–1988, vol. 1–5 [CD-ROM], NASA, Washington, D. C.
- Pal, J. S., and E. A. B. Eltahir (2001), Pathways relating soil moisture conditions to future summer rainfall within a land-atmosphere system, *J. Clim.*, **12**, 1227–1242.
- Qian, T., A. Dai, K. E. Trenberth, and K. W. Oleson (2006), Simulation of global land surface conditions from 1948 to 2004: Part I: Forcing data and evaluations, *J. Hydrometeorol.*, **7**, 953–975.
- Robock, A., K. Y. Vinnikov, C. A. Schlosser, N. A. Speranskaya, and Y. Xue (1995), Use of midlatitude soil moisture and meteorological observations to validate soil moisture simulations with biosphere and bucket models, *J. Clim.*, **8**, 15–35.
- Robock, A., C. A. Schlosser, K. Y. Vinnikov, N. A. Speranskaya, J. K. Entin, and S. Qiu (1998), Evaluation of the AMIP soil moisture simulations, *Global Planet. Change*, **19**, 181–208.
- Robock, A., K. Y. Vinnikov, G. Srinivasan, J. K. Entin, S. E. Hollinger, N. A. Speranskaya, S. Liu, and A. Namkhai (2000), The global soil moisture data bank, *Bull. Am. Meteorol. Soc.*, **81**, 1281–1299.

- Rossow, W. B., and Y.-C. Zhang (1995), Calculation of surface and top of atmosphere radiative fluxes from physical quantities based on ISCCP data sets: 2. Validation and first results, *J. Geophys. Res.*, **100**, 1167–1197.
- Rudolf, B., H. Hauschild, W. Reuth, and U. Schneider (1994), Terrestrial precipitation analysis: Operational method and required density of point measurements, in *Global Precipitation and Climate Change, NATO ASI Ser. I*, vol. 26, edited by M. Desbois and F. Desalmond, pp. 173–186, Springer, New York.
- Schlosser, C. A., A. Robock, K. Y. Vinnikov, N. A. Speranskaya, and Y. Xue (1997), 18-year land-surface hydrology model simulations for a mid-latitude grassland catchment in Valdai, Russia, *Mon. Weather Rev.*, **125**, 3279–3296.
- Schlosser, C. A., et al. (2000), Simulations of a boreal grassland hydrology at Valdai, Russia: PILPS phase 2(d), *Mon. Weather Rev.*, **128**, 301–321.
- Schnur, R., and D. P. Lettenmaier (1997), A global gridded data set of soil moisture for use in general circulation models, paper presented at 13th Conference on Hydrology, Am. Meteorol. Soc., Long Beach, Calif.
- Sellers, P. J., Y. Mintz, Y. C. Sud, and A. Dalcher (1986), A simple biosphere model (SiB) for use within general circulation models, *J. Atmos. Sci.*, **43**, 505–531.
- Sheffield, J., G. Goteti, and E. F. Wood (2006), Development of a 50-yr high-resolution global dataset of meteorological forcings for land surface modeling, *J. Clim.*, **19**, 3088–3111.
- Shukla, J., and Y. Mintz (1982), Influence of land-surface evapotranspiration on the Earth's climate, *Science*, **215**, 1498–1501.
- Simmons, A. J., and J. K. Gibson (2000), The ERA-40 project plan, *ERA-40 Proj. Rep. Ser. 1*, 63 pp., Eur. Cent. for Med.-Range Weather Forecasts, Reading, U. K. (Available at [http://www.ecmwf.int/publications/library/ecpublications/pdf/ERA40\\_PRS\\_1.pdf](http://www.ecmwf.int/publications/library/ecpublications/pdf/ERA40_PRS_1.pdf))
- Stackhouse, P. W., Jr., S. K. Gupta, S. J. Cox, M. Chiacchio, and J. C. Mikovitz (2000), The SRB project release 2 data set: An update, *WCRP GEWEX News*, **10**(3), 4.
- Taylor, C. M., and R. J. Ellis (2006), Satellite detection of soil moisture impacts on convection at the mesoscale, *Geophys. Res. Lett.*, **33**, L03404, doi:10.1029/2005GL025252.
- Vinnikov, K. Y., and I. B. Yesserkepova (1991), Soil moisture: Empirical data and model results, *J. Clim.*, **4**, 66–79.
- Vinnikov, K. Y., A. Robock, N. A. Speranskaya, and C. A. Schlosser (1996), Scales of temporal and spatial variability of midlatitude soil moisture, *J. Geophys. Res.*, **101**, 7163–7174.
- Viterbo, P., and A. C. M. Beljaars (1995), An improved land surface parameterization scheme in the ECMWF model and its validation, *J. Clim.*, **8**, 2716–2748.
- Xue, Y., P. J. Sellers, J. L. Kinter, and J. Shukla (1991), A simplified biosphere model for global climate studies, *J. Clim.*, **4**, 345–364.
- Xue, Y., F. J. Zeng, and C. A. Schlosser (1996), SSiB and its sensitivity to soil properties: A case study using HAPEX-Mobilhy data, *Global Planet. Change*, **13**, 183–194.
- Yang, Z.-L., R. E. Dickinson, A. Robock, and K. Y. Vinnikov (1997), Validation of the snow submodel of the Biosphere-Atmosphere Transfer Scheme with Russian snow cover and meteorological observational data, *J. Clim.*, **10**, 353–373.
- Zhao, M., and P. Dirmeyer (2003), Production and analysis of GSWP-2 near-surface meteorology data sets, *COLA Tech. Rep. 159*, 22 pp., Cent. for Ocean-Land-Atmos. Stud., Calverton, Md. (Available online at [ftp://grads.iges.org/pub/ctr/ctr\\_159.pdf](ftp://grads.iges.org/pub/ctr/ctr_159.pdf))

---

P. A. Dirmeyer, X. Gao, Z. Guo, Z.-Z. Hu, and M. Zhao, Center for Ocean-Land-Atmosphere Studies, Calverton, MD 20705, USA. (guo@cola.iges.org)

## Comparison of ERA40 and NCEP/DOE near-surface data sets with other ISLSCP-II data sets

Alan K. Betts,<sup>1</sup> Mei Zhao,<sup>2</sup> P. A. Dirmeyer,<sup>2</sup> and A. C. M. Beljaars<sup>3</sup>

Received 7 February 2006; revised 10 May 2006; accepted 1 June 2006; published 19 August 2006.

[1] The fields of 2-m temperature, relative humidity, precipitation, downward short-wave and long-wave radiation, net radiation, and sensible and latent heat flux from the ERA40 and NCEP/DOE reanalyses are compared with each other and with other independent ISLSCP-II data sets, where available. There are differences in the climatologies of the different data sets, but generally, they show consistent patterns of the major seasonal anomaly fields. The anomaly patterns are coherent, showing warm, dry seasonal biases associated with reduced precipitation and cloudiness, and the converse. This confirms that major changes in the atmospheric circulation patterns, with the associated differences in surface temperature, humidity, precipitation, cloud fields and incoming surface SW and LW radiation fluxes are captured by both reanalyses and the comparison ISLSCP-II data sets.

**Citation:** 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.

### 1. Introduction

[2] Near-surface meteorology data sets were extracted for the years 1986–1995 for the second International Land-Surface Climatology Project (ISLSCP-II) from the European Centre for Medium-range Weather Forecasts (ECMWF) reanalysis (ERA40 [Simmons and Gibson, 2000; Källberg et al., 2004; Uppala et al., 2005; <http://www.ecmwf.int/research/era/>]), and by the Center for Ocean-Land Atmosphere Studies (COLA) from the National Centers for Environmental Predictions (NCEP)/Department of Energy (DOE) Atmospheric Model Intercomparison Project (AMIP)-II Reanalysis [Kanamitsu et al., 2002]. For brevity we will refer to the two reanalyses as ERA40 and NCEP2. In this paper we compare five basic climate parameters; near-surface temperature, relative humidity, precipitation, incoming short-wave and long-wave radiation fluxes from the two reanalyses with other ISLSCP-II data sets (F. G. Hall et al., The ISLSCP Initiative II global data sets: Surface boundary conditions and atmospheric forcings for land-atmosphere studies, submitted to *Journal of Geophysical Research*, 2006, hereinafter referred to as Hall et al., submitted manuscript, 2006). We also intercompare the surface sensible and latent heat fluxes from the two reanalyses, for which there is no corresponding global observational time series over land (although we have an ocean climatology). The purpose of this analysis, which is purely qualitative, is to give users of the ISLSCP-II data

sets a broad visual overview of the differences and similarities between some key data sets.

#### 1.1. Concept of Reanalysis and Data Assimilation

[3] Routine analyses are produced by operational meteorological centers in real time several times each day using the current version of the centers' global forecast and analysis system. These models are constantly being updated and improved, so that over time fundamental climatological properties of the model are also changed, sometimes drastically. This makes a long time series of operational analyses useless for examining long-term trends or variations in climate. A reanalysis is a way to produce a dynamically consistent global analysis of the state of the atmosphere over an extended period of time (many years or decades) with no gaps in space or time. This is done by using a "frozen" version of the analysis model, and performing a retrospective analysis using the historic archive of observations. This allows for the use of more observational data, as many high-quality observations are not available to the operational centers in real time.

[4] Observational networks do not cover the entire globe uniformly, but have gaps in space and vary in their coverage over time. However, observational measurements provide the best estimate of the state of the atmosphere where they are taken. On the other hand, a geophysical fluid-dynamical model of the atmosphere containing parameterizations of important physical processes like radiative transfer, convection, turbulent transfer and diffusion of heat, moisture and momentum, can provide a complete global simulation of the atmosphere and is also used for forecasting purposes as it can be integrated forward in time. However, the models are imperfect representations of the atmosphere (as many small-scale processes are parameterized), and are prone to systematic errors, drift, and limitations owing to their finite spatial and temporal resolutions. A reanalysis is a combi-

<sup>1</sup>Atmospheric Research, Pittsford, Vermont, USA.

<sup>2</sup>Center for Ocean-Land-Atmosphere Studies, Calverton, Maryland, USA.

<sup>3</sup>European Center for Medium-Range Weather Forecasts, Reading, UK.

nation of model and measurement, using observations to constrain the dynamical model to optimize between the properties of complete coverage and accuracy. The insertion of observational information into the model integration is called data assimilation. Operational meteorological centers developed data assimilation as a means to generate initial conditions for dynamical numerical models that are consistent both with the model and the observed state of the atmosphere, thereby improving forecasts. However, the data assimilation practiced by operational centers is an adjustment of state variables, essentially an additional term in the predictive equations called an analysis increment. Because of the addition of these increments, reanalyses frequently do not locally conserve mass, energy, water or momentum. This can make operational reanalyses challenging to use in certain science applications.

[5] Although the reanalysis model is frozen, there are significant changes with time in the observational data sets, especially during the last few decades in satellite data. The ISLSCP-II decade of 1986–1995 does include significant changes in the satellite observations; and especially the introduction of microwave sensor data in 1987. Note that the changes with time in the observational systems means that these reanalyses still cannot determine trends on decadal timescales, although, as we shall show, the seasonal anomalies are coherent and useful.

## 1.2. Description of the Two Reanalyses

### 1.2.1. ERA40 Reanalysis System

[6] The European Centre for Medium-range Weather Forecasts (ECMWF) reanalysis (ERA40 [Källberg *et al.*, 2004; Uppala *et al.*, 2005]) covers the period from September 1957 to August 2002. Links can be found at <http://www.ecmwf.int/research/era/> for many aspects of ERA40, including documentation of the cycle 23r4 Integrated Forecast System that was used for this reanalysis; and a summary and discussion of the observations available at different times during the 40-year reanalysis. The ERA40 model has 60 atmospheric levels in the vertical, from the top of the model at 0.1 hPa to the lowest model level at about 10 m above the surface. The spectral resolution is  $T_L$ -159 (triangular truncation at wave number 159) with a corresponding resolution of about 125 km in grid point space. A so-called reduced Gaussian grid (N80, described at <http://www.ecmwf.int/products/data/technical/gaussian/>) is used for many physical processes and the land surface parameters. The ISLSCP products have been interpolated from this grid to the uniform 1° × 1° ISLSCP Earth grid, as much as possible consistent with the land-sea mask definitions. The ERA40 land sea mask (LSM) and the ISLSCP LSM are used to ensure that when possible only land points are transformed into land points, and only sea points are used for sea points. The ERA40 analysis uses the so-called three-dimensional variational method, where a cost function in relation to observations and background model field is minimized. The weighting of the different parts of the cost function is controlled by estimates of observation errors and model background errors. The spreading of observations in the horizontal and the vertical is controlled by horizontal and vertical correlation of the background errors. Most satellite observations (e.g., from infrared sensors) are used by computing radiances from the model fields (using a

forward model) and by comparing them with the satellite radiances. The analysis system uses a wide range of other observations, from conventional radiosonde and synoptic (SYNOP) observations, to ocean winds from satellite scatterometry. The ISLSCP period of 1986–1995 spans changes in the satellite observations used in the analysis. The satellite microwave data from the Special Sensor Microwave/Imager (SSM/I) was introduced in 1987 and the European Remote-Sensing Satellite (ERS) data in 1991.

[7] The analysis of T and q at the 2-m level and the snow depth analysis are part of a separate surface analysis, which uses a successive correction method. Because large areas over land do not have snow depth observations, a weak relaxation is applied to a snow depth climatology that is specified. ERA40 uses an optimal interpolation of soil water [Douville *et al.*, 2000], which adds soil water increments based on analysis increments in 2-m T and q. The land-surface scheme for ERA40 and its parameters are discussed by Van den Hurk *et al.* [2000] and its performance over the boreal forest is discussed by Betts *et al.* [2001]. ERA40 has a four layer soil model with a representation of frozen and unfrozen soil. In each grid box, there are tiles for bare soil and two classes of vegetation, short and tall, each with a specified vegetation type with a fixed set of parameters: fractional area, rooting distribution, leaf area index, roughness and canopy resistance. None of the vegetation parameters vary with time. Snow is represented by a single layer, which lies on top of bare ground and short vegetation, so that it is directly coupled to the atmosphere; but under the canopy for tall vegetation, with a separate energy balance that is less strongly coupled to the atmosphere. A report series evaluating ERA40 is available at <http://www.ecmwf.int/research/era/Products/>.

[8] This paper extends the work of Betts and Beljaars [2003], which compared intercompared ERA40 with other ISLSCP-II data sets. Evaluations of ERA40 have been made using river basin budgets for the Mississippi, Mackenzie, and Amazon basins [Betts *et al.*, 2003a, 2003b, 2005; Betts and Viterbo, 2005]. A useful time series analysis of the trends and variability in the CRU, ERA40 and the earlier NCEP/NCAR analyses of surface air temperature is given by Simmons *et al.* [2004]. A global evaluation of the hydrological cycle in ERA40 is given by Hagemann *et al.* [2005]. The ERA40 system has a spin-up of the precipitation field at high latitudes for the first 24–36 hours, associated with a problem in the moisture analysis [Betts *et al.*, 2003b]. The ERA40 reanalysis has a high bias in precipitation over the tropical oceans, stemming from a problem in the use of satellite radiances in the analysis of humidity [Troccoli and Källberg, 2004]. ERA40 has a known error in the diurnal cycle of precipitation over land (a bias toward precipitation too early in the day) which is larger in the tropics [Betts and Jakob, 2002] than the mid-latitudes. Betts [2006] analyzes the coupling of the diurnal cycle of temperature in ERA40 to the net long-wave radiation field. Here we average over the diurnal cycle. There is also a cold bias over ice-covered oceans in both the Arctic and the Antarctic [Betts and Beljaars, 2003], relating to the assimilation of infrared data satellite data, which affects part of the ISLSCP-II period, 1989–1996 (it was identified and largely corrected as the reanalysis progressed). The eruption of Pinatubo in 1991, which put

volcanic aerosol into the stratosphere, impacts the infrared radiances, which in turn impact the analyzed tropical circulation and rainfall, mainly over the tropical oceans.

[9] The output from ERA40 is an analysis every 6 hours at the standard synoptic times: 0000, 0600, 1200 and 1800 UTC. From each analysis 6-hour forecasts are run; and twice a day from the 0000 and 1200 UTC analyses, these forecasts are extended to 36 hours, with fields archived every 3 hours. The ISLSCP-II data includes both analyses every 6 hours, and model forecast data every 3 hours. In this paper, the ERA40 data comes primarily from monthly averages of the four daily 0–6 hour short-range forecasts, except for precipitation, where we shall show the monthly averages from 24–36 hour forecasts. All flux variables are averages accumulated during a model forecast, while some variables, such as temperature, humidity and wind are instantaneous fields. The ERA40 ISLSCP-II data set comprises some fixed fields, soil and snow variables (monthly time step), surface fluxes (monthly and 3-hourly), near surface variables (monthly, 3-hourly) at the lowest model level (roughly 10 m above the surface); and a level about 100 m above the surface (intended for driving off-line land-surface models in a loosely coupled mode). There are also the conventional 2-m temperature and dew point, and the 10-m wind that are computed using the boundary layer model from the predicted variables at the surface and the lowest model level. In addition, there is a separate 2-m analysis of temperature and dew point (6-hourly).

### 1.2.2. NCEP2 Reanalysis System

[10] The Center for Ocean-Land Atmosphere Studies (COLA) near-surface meteorology data set for ISLSCP-II has been derived from the NCEP/DOE AMIP-II reanalysis (NCEP2) [Kanamitsu *et al.*, 2002] that covers the years from 1979 to 2003. The purpose of this NCEP/DOE Reanalysis is to provide an improved version of the original NCEP/National Center for Atmospheric Research (NCAR) reanalysis [Kalnay *et al.*, 1996; Kistler *et al.*, 2001] for use by the AMIP-II project [Gleckler, 1996; <http://www-pcmdi.llnl.gov/projects/amip/NEWS/amipn18.pdf>] for general circulation model validation. The NCEP2 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 NCEP2 reanalysis to the ISLSCP 1° Earth grid, the reanalysis data set was regridded using bilinear interpolation from its native T62 Gaussian grid resolution (192 × 94 grid boxes globally) to the 1° spatial resolution required by ISLSCP-II. When possible, NCEP2 land grid points are mapped to ISLSCP land grid points, and NCEP2 water grid points are mapped to ISLSCP water grid points. On occasions where there is no overlap of like surfaces between the two grids, a straight interpolation is performed. The NCEP2 reanalysis scheme is also a three-dimensional variational scheme called spectral statistical interpolation [Parrish and Derber, 1992]. The NCEP reanalysis system is thoroughly reviewed by Kalnay *et al.* [1996]: here we summarize the error fixes, model changes, new components and boundary conditions implemented for NCEP2 [Kanamitsu *et al.*, 2002]. Errors found and corrected include misregistration of PAOBS (subjective sea level pressure analyses) over the Southern Hemisphere for 1979–1992,

mishandling of observed snow cover data between 1974–1994, a problem in humidity diffusion that led to extremely noisy snow analysis maps, spatial discontinuities in the relationships between relative humidity and cloudiness, and problems in the representation of ocean albedo and snowmelt. Major changes to the model physics for NCEP2 were the implementation of a new planetary boundary layer scheme [Hong and Pan, 1996] and short-wave radiation parameterization [Chou and Lee, 1996]. Minor changes include an increase in the frequency of full radiation calculations from 8 to 24 times daily, retuning of the stratus cloud and convective parameterizations, improved cloud top cooling and calculation of radiation on the full Gaussian grid. Two new system components may have a particularly strong impact on the surface fields in ISLSCP-II. An assimilation of precipitation was incorporated to improve the simulation of soil wetness. The difference of pentad-averaged reanalysis rainfall from the Xie and Arkin [1997] precipitation estimate is subtracted from infiltration into the top soil layer in an effort to improve the land surface hydrology. Similarly, adjustments to observed snow cover analyses based on model predicted snow cover replaced a simple empirical estimate for more realistic winter hydrology. Finally, several fixed fields were changed, including new data for desert albedos, ozone, Northern Hemisphere snow cover; and the sea surface temperatures and sea-ice data used are consistent with AMIP-II. The land surface scheme used in NCEP2 is the same as for the original NCEP/NCAR reanalysis. It is a relatively simple two-layer soil with simple parameterizations of surface flux exchanges, including restrictions on evapotranspiration caused by plant vascular systems [Pan and Mahrt, 1987]. This scheme includes no interannual or seasonal variations in vegetation cover, and minimal spatial variability of surface properties.

[11] Some data comparisons of the COLA NCEP2 data set are contained in a report by Zhao and Dirmeyer [2003], who discuss the production of a hybrid data set for the second Global Soil Wetness Project [Dirmeyer *et al.*, 2006]. There are several studies of the hydrologic cycle of the NCEP2 reanalysis. Roads *et al.* [2002] evaluated the NCEP2 water and energy budgets on a global scale. Roads *et al.* [2003] is a comprehensive analysis of the water and energy budgets for the Mississippi river basin comparing the two NCEP reanalyses with operational model products and an off-line hydrologic model [Maurer *et al.*, 2001]. Roads [2003] evaluates the tropical precipitation of both NCEP reanalyses against TRMM data. Lu *et al.* [2005] compared the simulations of soil wetness to observations and the earlier NCEP/NCAR reanalysis and found some improvements, although Dirmeyer *et al.* [2004] describe shortcomings in the ability of NCEP2 to simulate interannual variations of soil wetness where long-term observations are available.

[12] The analysis increment for the NCEP2 reanalysis is 6 hours, and output data are routinely reported every hour or 6 hours. In order to satisfy the ISLSCP-II requirement for 3-hourly data, twice daily 36 hour forecasts were made and hourly output were obtained from the 24–36 hour forecasts. The 24–36 hour forecasts were chosen instead of the 0–12 hour forecasts to minimize initial “spin-up/spin-down” problems. The 1-hour data were later combined to produce 3-hourly data. Time averaging was performed on the native reanalysis grid before regridding. The fields from the



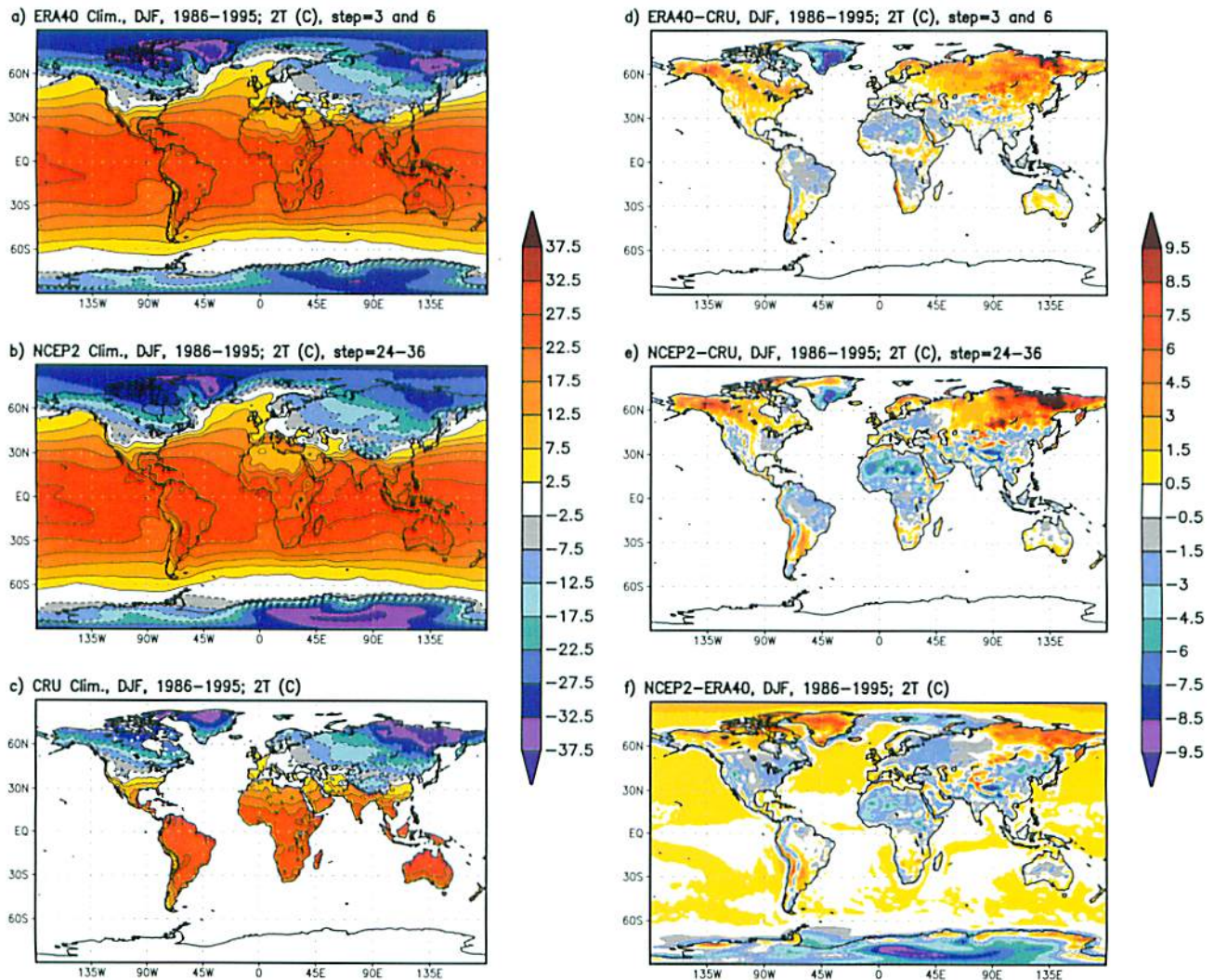
**Table 1.** Some Differences Between the NCEP2 and ERA40 Reanalysis Products

Property	NCEP/DOE	ERA40
Original horizontal resolution	T62 (approximately 1.9° on Gaussian grid)	T159 (approximately 125 km on a reduced Gaussian grid)
Full period of coverage	1979-current	Sep 1957 to Aug 2002
Analysis increment	6 hours	6 hours
Soil moisture nudging	based on pentad precipitation errors	based on near-surface temperature and humidity errors
ISLSCP source data	24–36 hour forecasts	analyses and 0–6 hour forecasts (plus 0–12, 12–24 and 24–36 hour, for precipitation)

NCEP2 reanalysis that are provided for ISLSCP-II are near surface meteorological fields, fluxes of heat, moisture and momentum at the surface, and land surface state variables, all with a spatial resolution of 1° in both latitude and longitude. 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).

### 1.3. Differences Between NCEP2 and ERA40 Reanalyses

[13] Note that there are fundamental differences between the NCEP2 and ERA40 reanalysis products (Table 1), which are not evident to the user of the ISLSCP-II versions of the products, since they have been coregistered on the same 1° grid, and presented with the same time structure. These differences may contribute beyond what might be expected simply from the disparities in model physics and assimilated data streams. For example, the differences in the forecast intervals used are a response to very different spin-up



**Figure 1.** (a–c) Two-meter temperature distribution from ERA40, NCEP2 and the CRU data (over land) for the 10-year (1986–1995) DJF climatologies and (d–f) difference fields for ERA40-CRU, NCEP2-CRU, and NCEP2-ERA40.



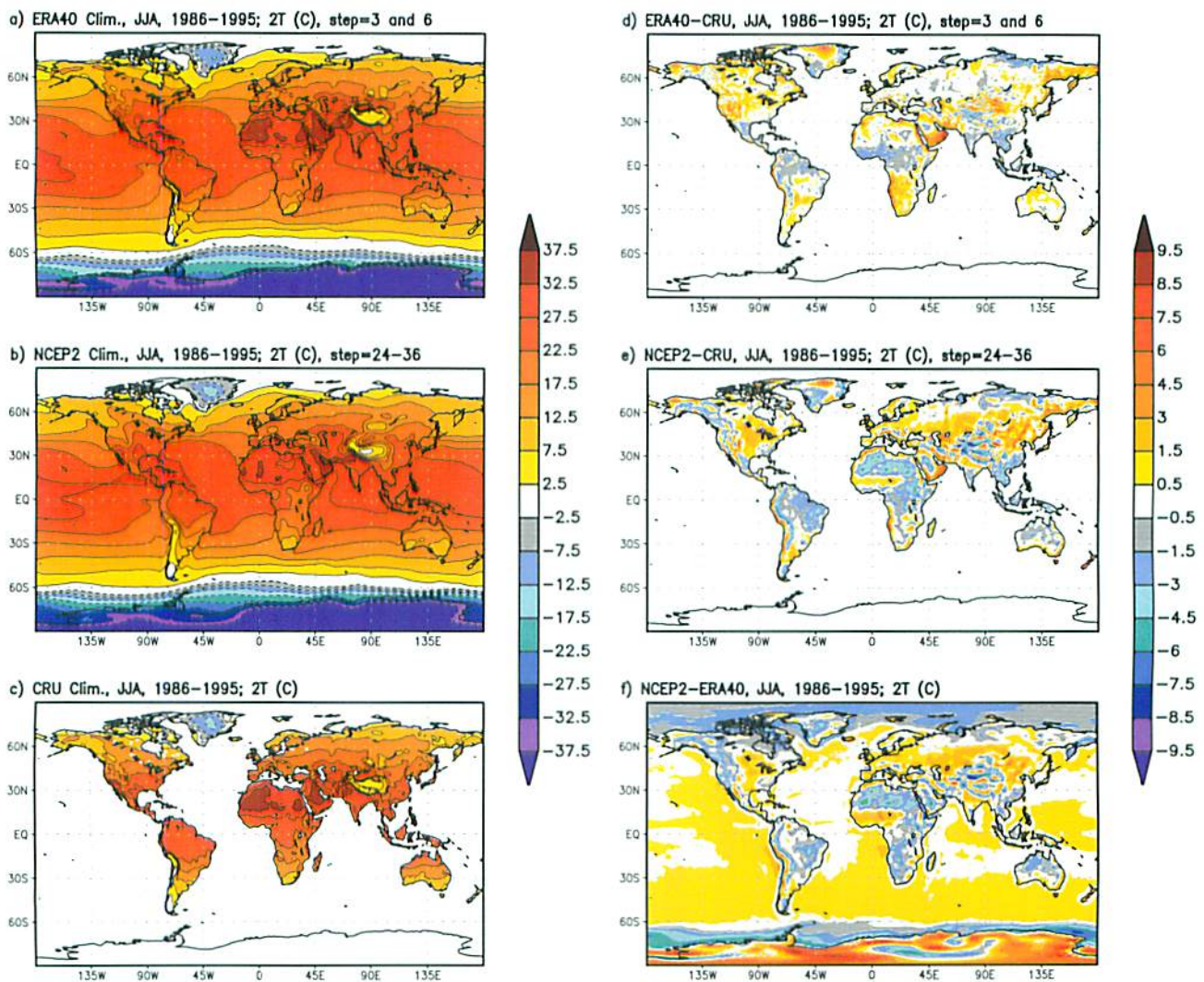


Figure 2. (a–e) As Figure 1 but for JJA climatologies and difference fields.

behaviors in the two products, particularly for precipitation and surface fluxes.

#### 1.4. Comparison With ISLSCP-II Data

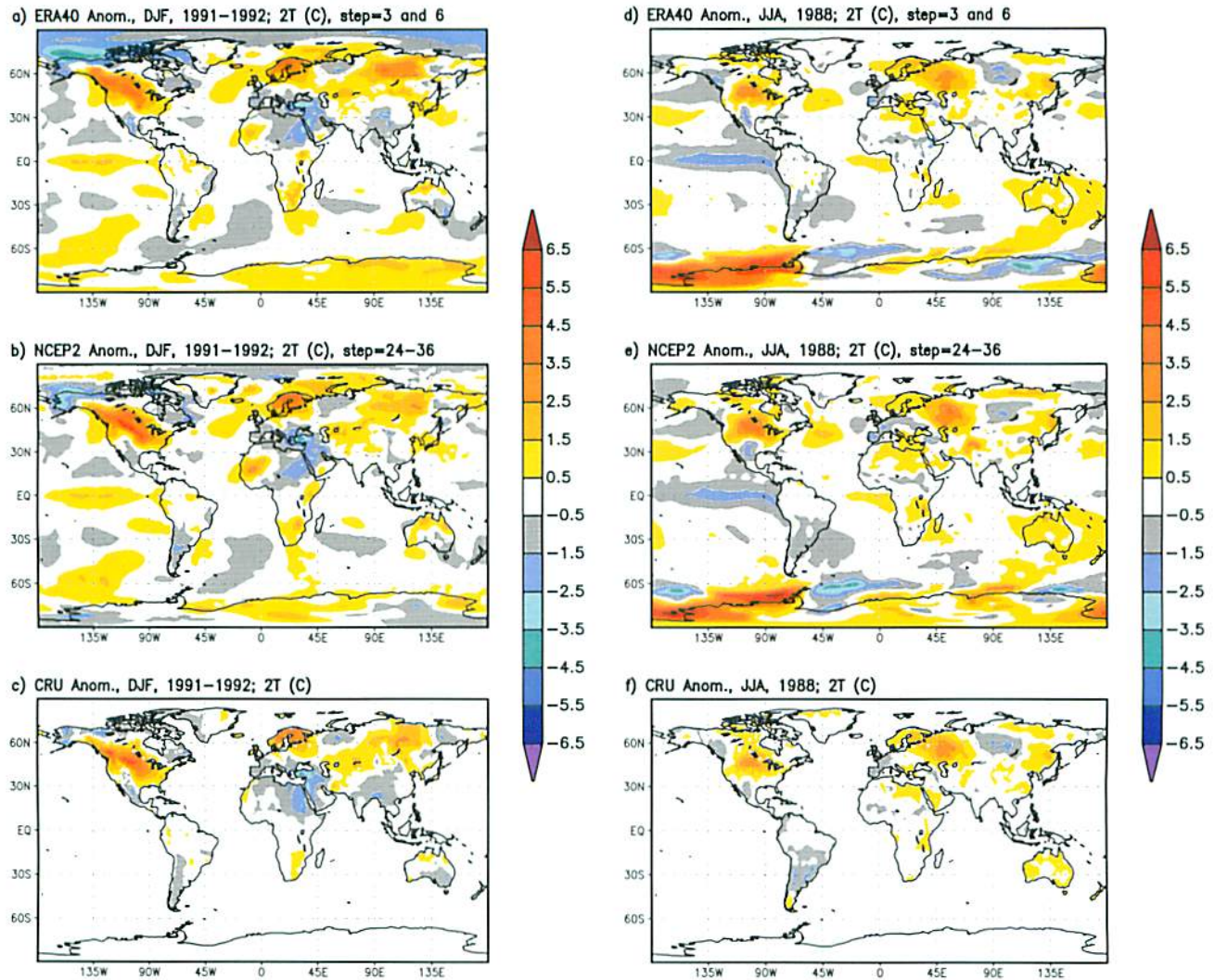
[14] In this paper, we show comparisons of the two reanalyses with three other ISLSCP-II data sets (Hall et al., submitted manuscript, 2006). For temperature and relative humidity, we compare with the ISLSCP-II surface meteorology data set regridded from the Climatic Research Unit, University of East Anglia data set [New et al., 1999, 2000]. This we shall refer to as the CRU data set. Temperature is a primary variable with extensive coverage, which is directly analyzed. Where observations are sparse and over mountainous terrain, this gridded temperature analysis may not be representative. However, New et al. [2000] treat vapor pressure (from which we compute relative humidity, RH) as a secondary variable, and the gridded analysis uses not only station observations where available (converting monthly mean RH to vapor pressure for some stations) but also synthetic data (found by estimating monthly dew point from monthly mean minimum temperature), which is added to the analysis in regions of sparse data.

[15] For precipitation we compare with the Global Precipitation Climatology Project (GPCP) Version 2 Satellite-

Gauge (SG) combination global gridded monthly precipitation data set [Adler et al., 2003], which is considered the best all-round monthly precipitation in the ISLSCP-II set, that is derived from observations as opposed to reanalysis (Hall et al., submitted manuscript, 2006). Over land, the GPCP SG 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 and other ungauged regions. Thus the GPCP SG is globally complete, albeit with reduced confidence over the oceans and at high latitudes. For brevity, this will be referred to in this paper as the GPCP precipitation. Further discussion of this GPCP data set, and the other precipitation data sets in the ISLSCP-II collection is given by Hall et al. (submitted manuscript, 2006).

[16] For the surface radiation, we compare the downwelling short-wave (SWdown) and long-wave (LWdown) radiation fluxes with the corresponding radiation fluxes from the surface radiation budget data set (for which we use the acronym SRB) developed by Stackhouse et al. [2004] and Cox et al. [2004] from the ISCCP cloud data [Rossow and





**Figure 3.** Two-meter temperature anomaly fields from ERA40, NCEP2 and the CRU data (over land) for (a–c) DJF 1991–1992 and (d–f) JJA 1988.

Schiffer, 1999]. The algorithms used for the SW fluxes are documented by *Pinker and Laszlo* [1992] and *Gupta et al.* [1992] for the LW radiation fluxes. Meteorological profile information is developed at the processing grid resolution from the NASA Data Assimilation Office Goddard Earth Observing System version 1 (GEOS-1) reanalysis [*Schubert et al.*, 1995]. Note that this is a different and earlier reanalysis, which has a different set of temperature and humidity biases from NCEP2 and ERA40.

[17] For the surface sensible and latent heat fluxes, we have no comparison data set over land, but we will compare the reanalyses over the oceans with a monthly climatology, derived from marine observations [*Da Silva et al.*, 1994]. This climatology does end however in 1993, so we use the 1986–1993 climatology for comparison.

## 2. Comparison of Reanalyses With Selected ISLSCP-II Products

[18] The advantage of surface fields from model reanalyses is that they have complete coverage at 3-hourly time resolution (in this ISLSCP-II data). In contrast, surface

observations are not uniformly distributed globally, and are sparse over many regions in the tropics, where only monthly mean data may be available, or even just climatology. Surface observations can be interpolated to a common grid, as is the case for our CRU data set, but where observations are sparse; the gridded analysis may not be representative. Models not only assimilate data, but in regions of missing data, the global model will compute a complete set of fields from continuity and the dynamic and thermodynamic equations. This means that model products have biases, related to the specific set of model equations and the parameterizations for unresolved physical processes. The NCEP and ECMWF analysis-forecast systems differ in their model structure, physical parameterizations and horizontal and vertical resolution, and in their methods of processing the input observations. Consequently there are differences between the model surface fields. In this short paper, we cannot analyze in detail all the differences between the reanalyses. Our intent is to help the users of the ISLSCP-II data assess which products might be useful for different purposes. In many cases the differences between reanalyses and other data sets will give an estimate in our uncertainty in a



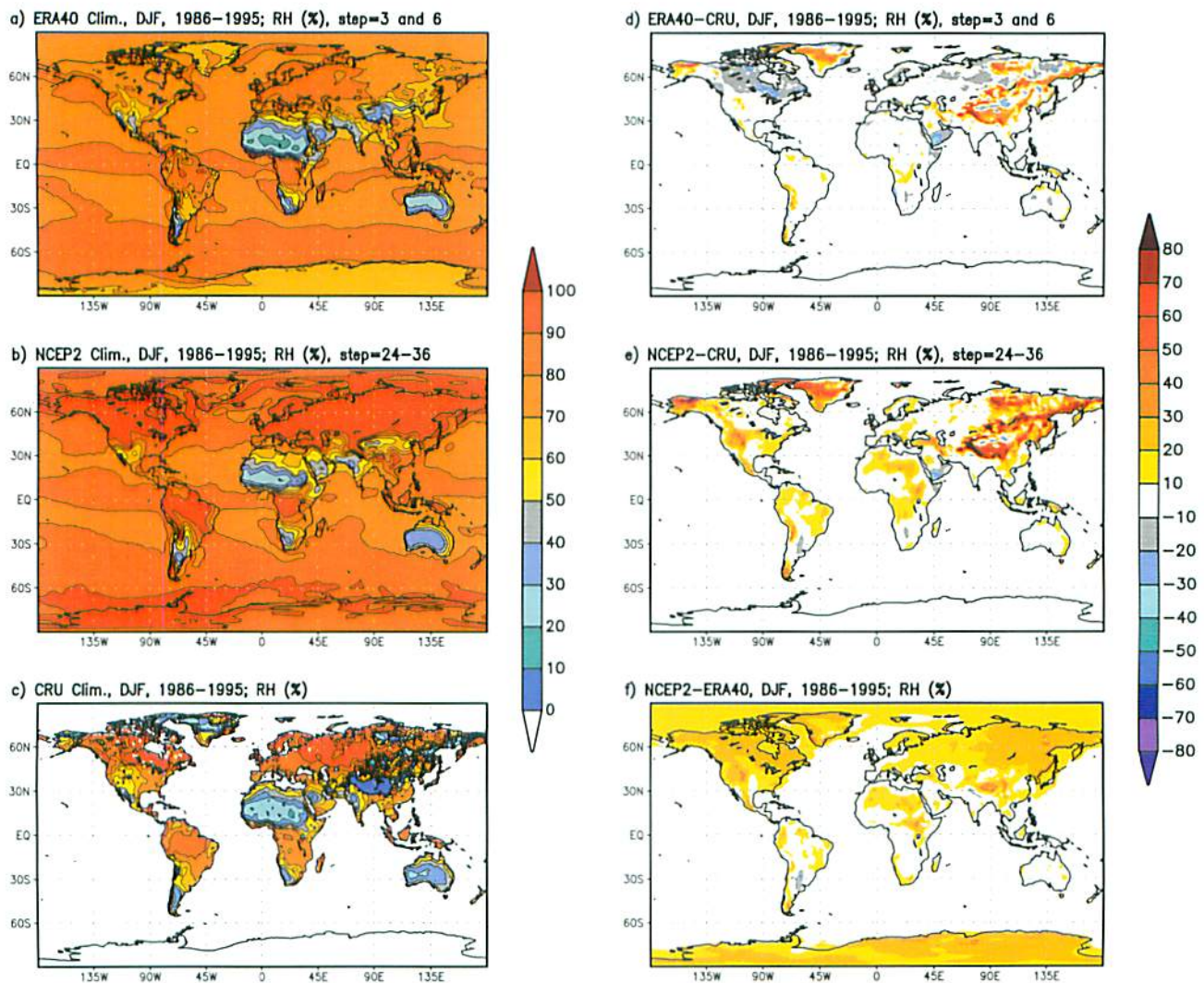


Figure 4. (a–e) As Figure 1 but for 2-m RH.

given variable. We will give a broad overview on seasonal timescales of differences at the surface by comparing the reanalysis fields with independent data from the ISLSCP-II data collection. The ISLSCP data set contains several other near-surface products, some derived directly from surface observations, and some from satellite observations. We will present the 10-year mean climatology for DJF (December, January, February) and JJA (June, July, August) for the two reanalyses and a corresponding data set; and the difference fields of the reanalyses from this data set, as well as the difference of the two reanalyses. We will also compare anomaly fields for two seasons (DJF, 1991–1992 and JJA, 1988), where the anomalies are computed from the separate climatologies for each reanalysis or other data set. The first, DJF, 1991–1992, has a warm anomaly in the eastern Pacific, the onset of an El Niño event; and the second, JJA 1988 has a cold anomaly in the eastern Pacific, and also corresponded to a drought over the continental United States. A full set of these anomaly fields for all four seasons is available for ERA40 at <http://www.ecmwf.int/research/demeter/d/inspect/catalog/research/era/diagnostics/ISLSCP-II/>.

## 2.1. Surface Temperature and Relative Humidity Fields

### 2.1.1. The 2-m Temperature

[19] Figures 1 and 2 show the 2-m temperature distribution from ERA40, NCEP2 and the CRU data (over land) for the 10-year (1986–1995) DJF and JJA climatologies and selected difference fields. The three climatologies on the left (Figures 1a–1c and 2a–2c) show that the two reanalyses are similar to each other and to the CRU analysis over land. The very low JJA temperatures (below  $-55^{\circ}\text{C}$ ) over the Antarctic icecap do not appear with the contouring shown. Three difference fields are shown on the right in Figures 1d–1f and 2d–2f; for ERA40–CRU, NCEP2–CRU, and NCEP2–ERA40 respectively. ERA40 tends to be a little cooler in the tropics than the CRU analysis, and a little warmer in mid and high latitudes, especially over the boreal forest in the northern winter. These ERA40 temperature biases are consistent with those shown in *Betts et al.* [2003b] for the Mackenzie River basin and *Betts et al.* [2005] for the Amazon basin. The pattern of the NCEP2 temperature bias differs from ERA40. The cool bias over the Amazon and Sahara is a little larger for NCEP2 than ERA40, but over the Sahel in JJA, the two reanalyses have opposite biases. The



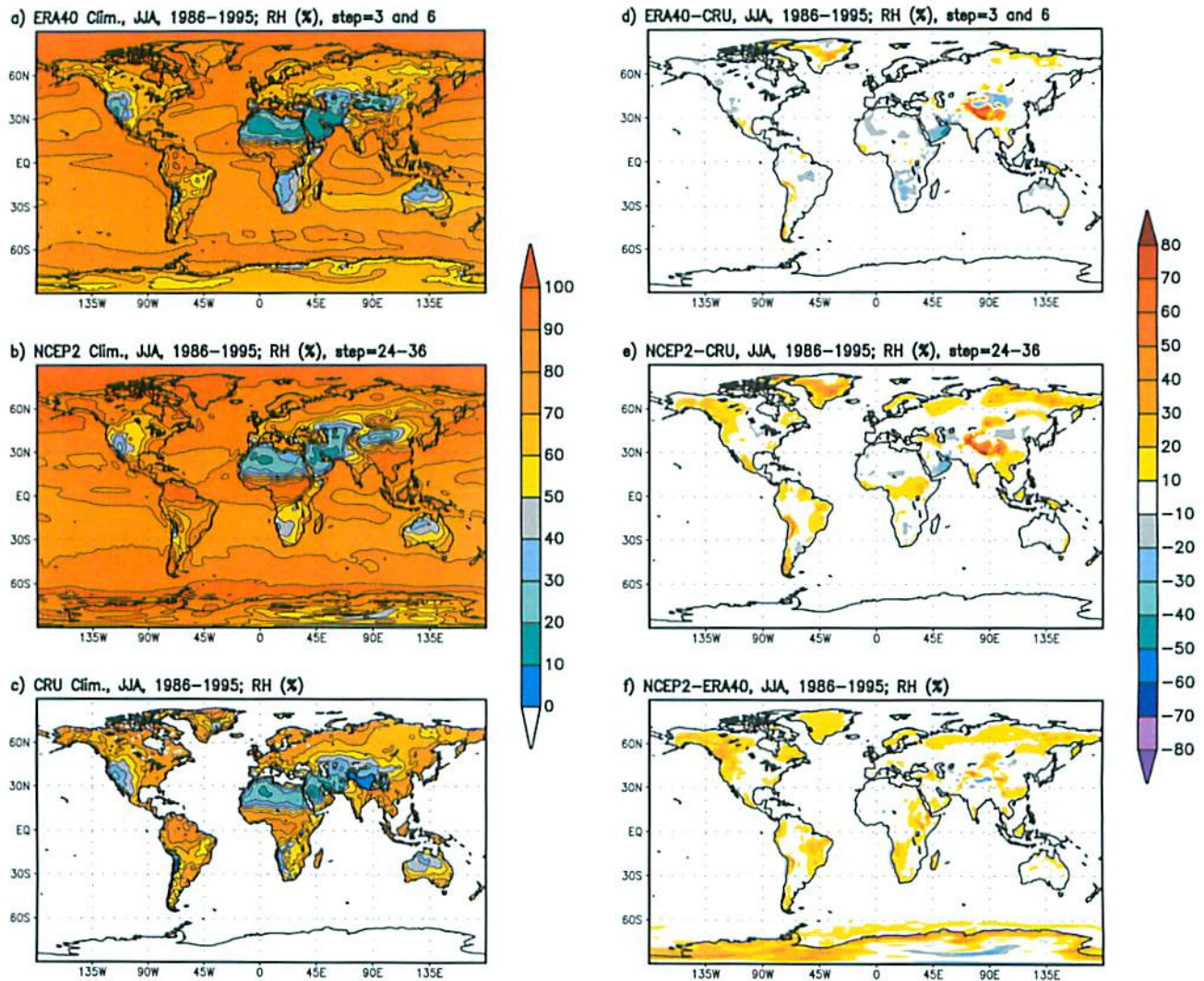


Figure 5. (a–e) As Figure 2 but for 2-m RH.

Eurasian warm bias during DJF is larger in NCEP2 further to the east than in ERA40. Although sea surface temperature is specified, NCEP2 has a slightly warmer 2-m temperature over much of the oceans. The comparisons over regions of sparse data and over high terrain should be viewed with some caution. There are differences in the orography used in the two analyses, and in mountainous regions data coverage is generally sparse and stations are often limited to the valleys. In polar regions where we have no CRU data, NCEP2 is warmer than ERA40 in the winter hemisphere and cooler in the summer hemisphere. ERA40 has a known cold bias over ice-covered oceans in both the Arctic and the Antarctic, relating to the assimilation of infrared satellite data (which was identified as the reanalysis progressed, so that only the years, 1989–1996, are affected). The difference pattern in Figures 1f and 2f resemble Figures 1e and 2e more closely than they do Figures 1d and 2d; which means that, away from polar regions, ERA40 has generally smaller biases over land than NCEP2.

[20] Figure 3 shows the anomaly fields (each from their own climatologies) for two seasons, DJF, 1991–1992 (Figures 3a–3c) and JJA, 1988 (Figures 3d–3f) for the two reanalyses and the CRU analysis over land. Despite the

differences in the climatologies, the anomaly patterns are quite similar, showing that both reanalyses capture the broad character of differences in seasonal weather regimes.

#### 2.1.2. The 2-m Relative Humidity

[21] Figures 4 and 5 show the relative humidity (RH) distribution (with respect to water saturation) from ERA40, NCEP2 and the CRU data (over land) for the 10-year DJF and JJA climatologies. The differences between the reanalyses are small over the oceans. Over land, NCEP2 generally has a higher RH than ERA40. With respect to the CRU analysis, ERA40 is drier at most high northern latitudes in DJF, while elsewhere the ERA40 bias is small. The NCEP2 RH bias is generally positive over land. As a result, RH in NCEP2 is generally higher than in ERA40, which would correspond to a lower mean lifting condensation level. Figure 6 shows the anomaly fields for DJF 1991–1992, and JJA, 1988 (the same seasons as Figure 3). There is some correspondence in the anomaly patterns of RH: generally the ERA40 anomaly extremes are larger than those of the CRU data set, and for NCEP2 they are larger still. Comparing Figures 3 and 6, we see the correspondence in the summer hemisphere of warm-dry and cool-wet anomalies, as



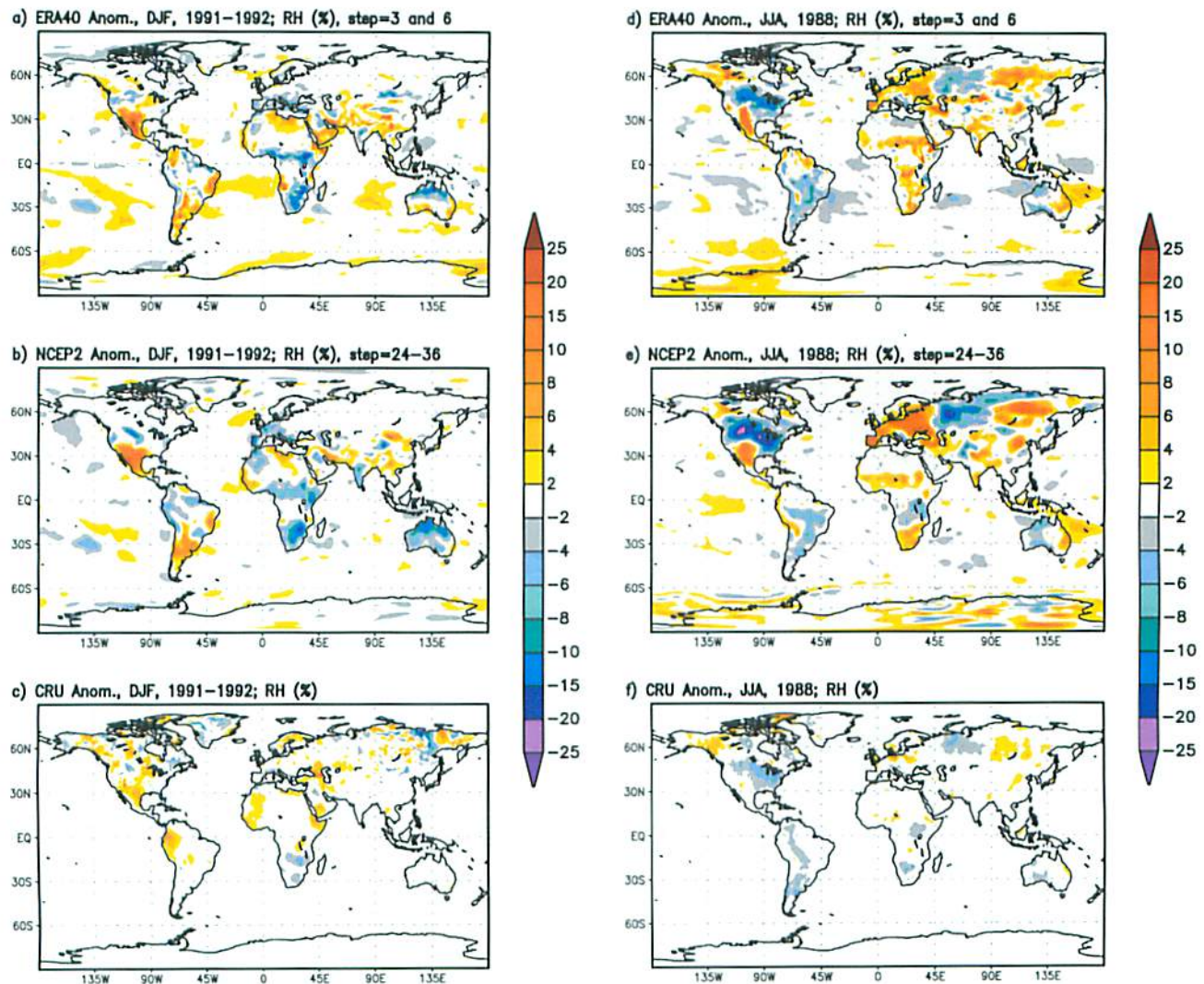


Figure 6. (a–e) As Figure 3 but for 2-m RH.

expected: for example, the warm, dry conditions associated with the summer drought over the United States in JJA 1988.

[22] However *New et al.* [2000] treat vapor pressure (from which we compute relative humidity, RH) as a secondary variable, and the gridded analysis uses not only station observations where available (converting monthly mean RH to vapor pressure for some stations) but also synthetic data (found by estimating monthly dew point from monthly mean minimum temperature) in regions of limited observations. As a result there is considerably more uncertainty in their humidity analysis than in their temperature analysis, especially over regions of sparse data.

## 2.2. Precipitation Fields

[23] Precipitation in the reanalyses is most affected by the spin-up of the dynamic fields in midlatitudes [*Betts et al.*, 2003b]. For ERA-40, the ISLSCP-II data set includes monthly precipitation computed from the 0–6, 0–12, 12–24 and 24–36 hour forecasts so user can assess this spin-up. *Betts and Beljaars* [2003] includes some examples. For NCEP2, the ISLSCP-II data set is derived from the 24–36 hour forecasts as discussed in section 1.2.2. In this section we compare the 24–36 hour forecast precipitation

for both ERA40 and NCEP2 with the observationally based GPCP data set, which combines a standard gauge analysis over land (with climatological bias correction) with a community-based satellite-only product to improve estimates where gauges are sparse. Thus this GPCP analysis is globally complete, albeit with reduced confidence over regions such as open oceans and at high latitudes, where it is a satellite-only product. Figures 7 and 8 show the precipitation distribution from ERA40, NCEP2 and the GPCP data for the 10-year DJF and JJA climatologies. They have the same structure as earlier figures with the three climatologies on the left and the difference fields on the right. Although the climatologies are generally similar, the reanalyses have significant biases. 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 in the use of satellite radiances in the analysis of humidity [*Troccoli and Källberg*, 2004]. ERA40 also has a negative DJF bias over the Amazon. For NCEP2, the biases over the tropics are smaller than in



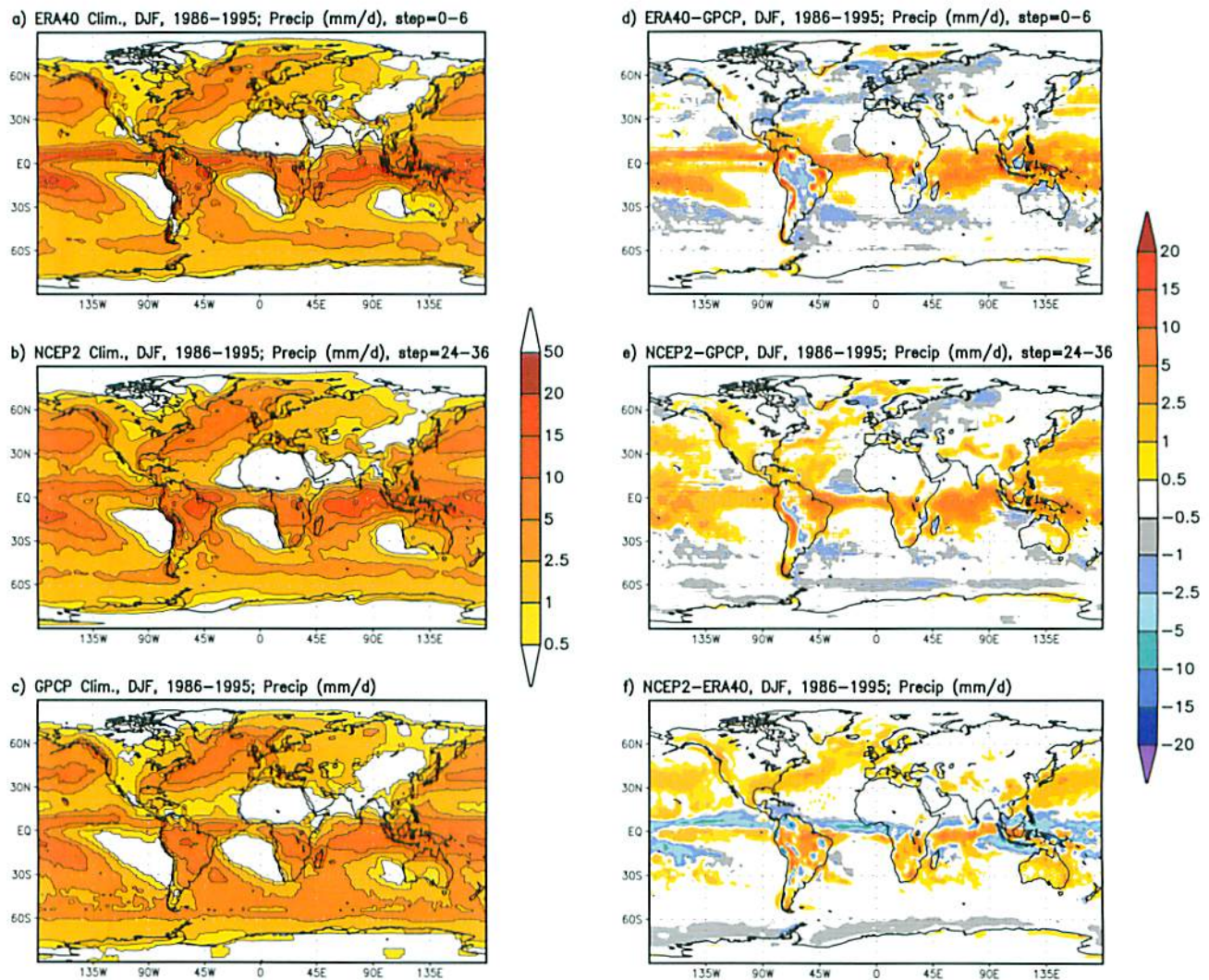


Figure 7. (a–e) As Figure 1 but for 24–36 hour forecast precipitation and GPCP precipitation.

ERA40. In midlatitudes, the NCEP2 biases 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 JJA, the ITCZ precipitation in both reanalyses does not extend as far north as in the GPCP analysis. The biases for ERA40 are a little smaller with the 24–36 hour forecast precipitation than for the 0–6 hour forecast precipitation (not shown, although the ISLSCP-II data set includes both for ERA40), so for most users 24–36 hour precipitation is the preferred choice for the reanalyses. ERA40 has a known error in the diurnal cycle of precipitation over land (a bias toward precipitation too early in the day) which is larger in the tropics [Betts and Jakob, 2002] than the midlatitudes. Similar biases in the diurnal cycle exist for NCEP2. However, the seasonal means that we show average over the diurnal cycle in the reanalyses.

[24] Figure 9 compares the ERA40, NCEP2 and GPCP precipitation anomalies for DJF, 1991–1992 and JJA, 1988. Despite the differences in their means, the anomaly patterns are remarkably similar, especially considering that the three analyses differ in their native horizontal resolution. 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 blend of satellite data for the cloud field and rain-gage data, primarily over land. Again, if Figures 3, 6, and 9 are compared, we see 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 differences for the two seasons shown.

### 2.3. Surface Radiation Budget

[25] The SRB data set [Stackhouse *et al.*, 2004; Cox *et al.*, 2004] computes the SW and LW radiation fluxes from the ISCCP cloud data set [Rossow and Schiffer, 1999], using the algorithms of Pinker and Laszlo [1992] and Gupta *et al.* [1992], and meteorological profile information from the



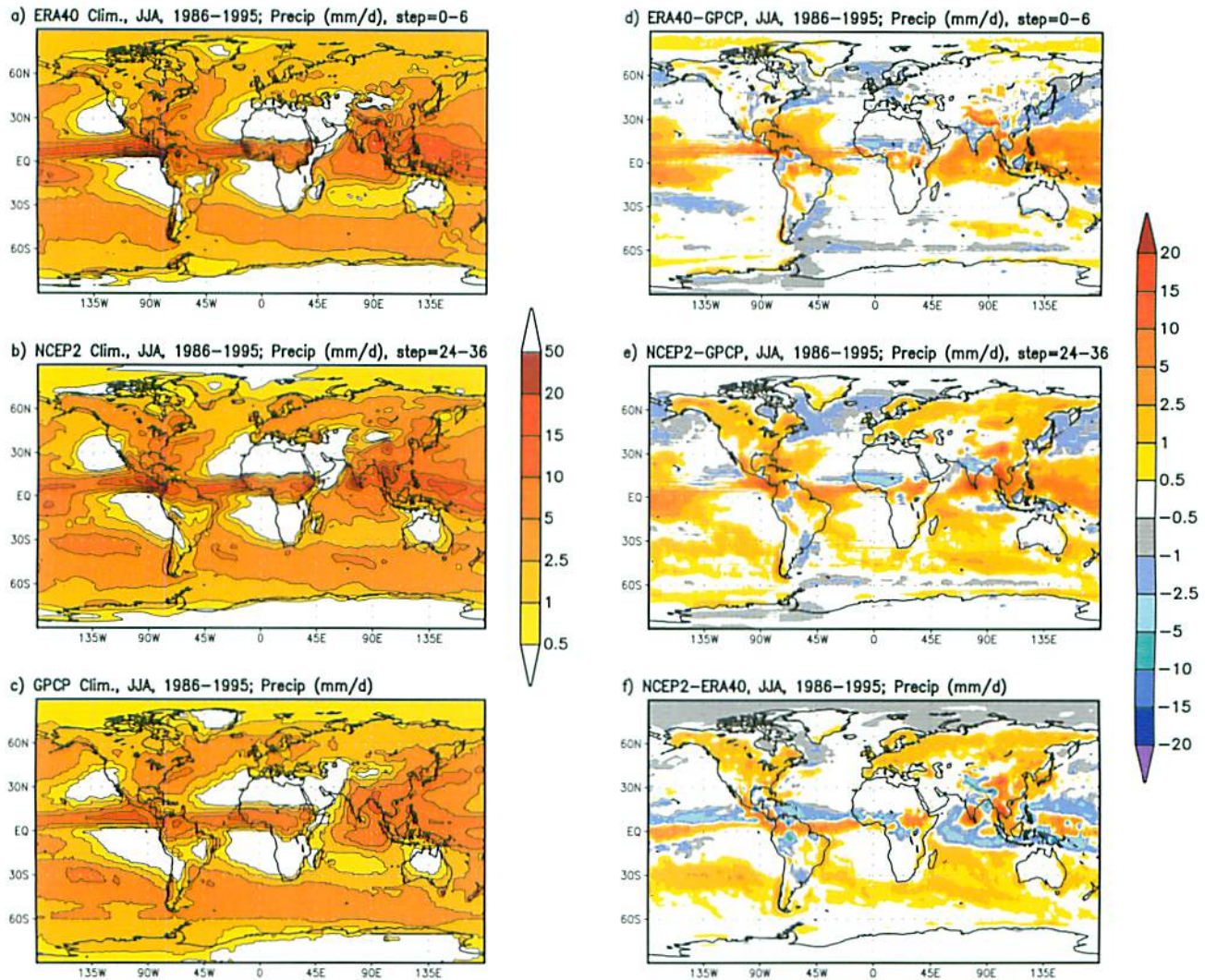


Figure 8. (a–e) As Figure 7 but for JJA climatology.

GEOS-1 reanalysis [Schubert *et al.*, 1995]. This is an earlier reanalysis than NCEP2 and ERA40, with a different set of biases in temperature and humidity. The ERA40 and NCEP2 reanalyses use their own radiation codes, meteorological profiles and model cloud fields to compute their radiation fluxes. The satellite-based data sets for short-wave are generally similar to each other, because they typically use cloud information from the International Satellite Cloud Climatology Project (ISCCP). One known problem with the SRB SW data set is that it underestimates the downward short-wave flux over the central part of the Tibetan Plateau [Masuda, 2004], and the bias extends to a wide area of elevated terrain in western China, and overestimates the downward surface shortwave flux over regions of China with high absorbing aerosol optical thickness [Hayasaka *et al.*, 2006].

[26] The downward surface SW flux, SWdown, can be written as the difference between the clear-sky flux and the surface SW cloud forcing (SSWCF):

$$\text{SWdown} = \text{SWdown}(\text{clear}) + \text{SSWCF}$$

For both reanalyses and SRB data set, SWdown(clear) is a model calculation from the top-of-the-atmosphere (TOA)

incoming flux, modified by the atmospheric absorption. Since all three calculations use different radiation models, together with slightly different atmospheric structure and composition, some small differences (not shown) can be expected in their estimates of SWdown(clear). However, the primary differences between the three estimates of SWdown come from the computation of the impact of the cloud field, SSWCF. Whereas, the SRB estimate is based on the TOA flux measurements of the observed cloud field, ERA40 and NCEP2 calculate their cloud fields using their model parameterizations. We presume therefore that the SRB SWdown estimate is closer to the truth (since it is based on an observed cloud field); so that the model differences from SRB represent primarily model bias, associated with errors in the model cloud fields.

### 2.3.1. Incoming Short-Wave Radiation Flux, SWdown

[27] Figure 10 shows the 10-year climatology of surface SWdown for southern summer DJF from ERA40, NCEP2 and SRB (Figures 10a–10c) and the difference of ERA40 and NCEP2 from SRB climatology (Figures 10d and 10e), and the difference, NCEP2-ERA40 (Figure 10f). The three difference patterns have some similarities. For DJF, SWdown for ERA40 is systematically low in the tropics (–10 to



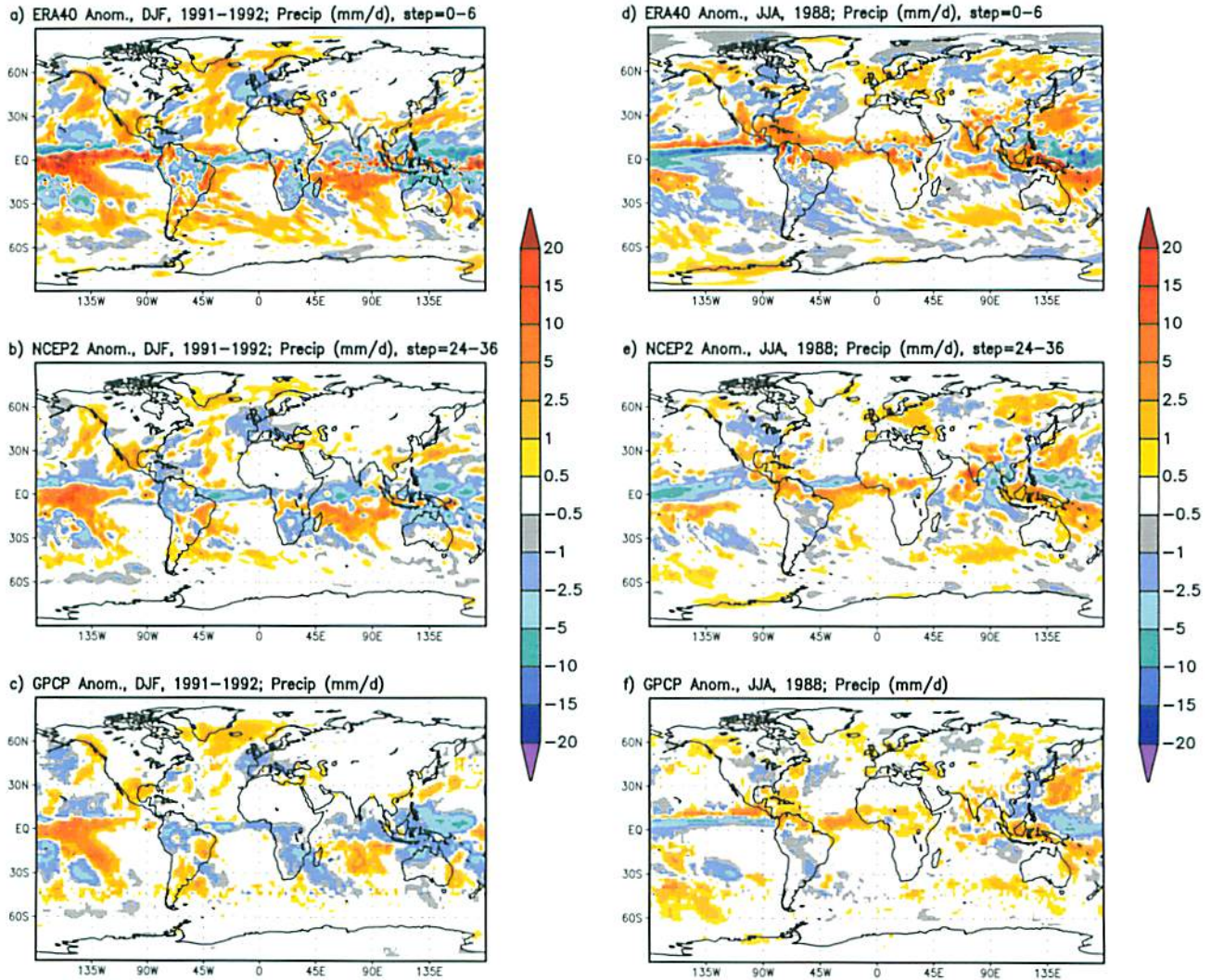


Figure 9. (a–e) As Figure 3 but for precipitation anomalies.

$-50 \text{ W m}^{-2}$ ), suggesting too much reflective cloud cover. SWdown is too large (suggesting too little cloud), for the southern ocean, for the stratocumulus areas off the western edge of continents, and for some continental regions. For NCEP2, the bias pattern is similar and the biases are generally larger (e.g.,  $-20$  to  $-90 \text{ W m}^{-2}$  in the tropics), except over Antarctica (where the uncertainty of the SRB estimate becomes large because of the difficulty of distinguishing clouds over background ice). Figure 11 is the corresponding plot of SWdown for northern summer, JJA. Again we see that the reanalyses have low biases over the tropical oceans, and high biases over the stratocumulus regimes and parts of the northern continents. Again the NCEP2 biases are larger, especially over the northern continents, where clearly NCEP2 has too little cloud cover. The ERA40 biases over the northern continents are mixed and generally smaller. The low bias of the SRB SWdown data over the Tibetan plateau [Masuda, 2004] is consistent with the positive difference we see for ERA40-SRB. In the tropics, differences in the location and width of the ITCZ is responsible for the banded structure near the equator in the

difference NCEP2-ERA40. Comparing Figures 10 and 11 with the corresponding Figures 7 and 8 for the precipitation bias shows that some (but not all) of the model errors in cloud cover over the tropical oceans are not surprisingly associated with errors in the model precipitation field. Figure 12 compares anomaly fields for ERA40, NCEP2 and SRB (from their own climatologies) for DJF, 1991–1992 (Figures 12a–12c) and JJA 1988 (Figures 12d–12f). Encouragingly, despite the biases in their climatologies, the major anomaly signals are similar in the two reanalyses and the SRB data set except at very high latitudes. Comparing Figures 3, 6, 9, and 12, the coherence of the anomaly patterns can be again seen in the summer hemispheres, with the association of high SWdown with higher temperature but lower precipitation and RH. For this pair of seasons, the SRB anomalies are generally closer to the ERA40 anomalies over the tropical oceans.

[28] Differences in SWup are not shown here: they reflect differences in background albedo, as well as the albedo with snow cover. The two reanalyses have different specified background albedos, and the SRB data set albedo differs substantially from ERA40 [Betts and Beljaars, 2003]. The ISLSCP-II data collection contains six other albedo prod-



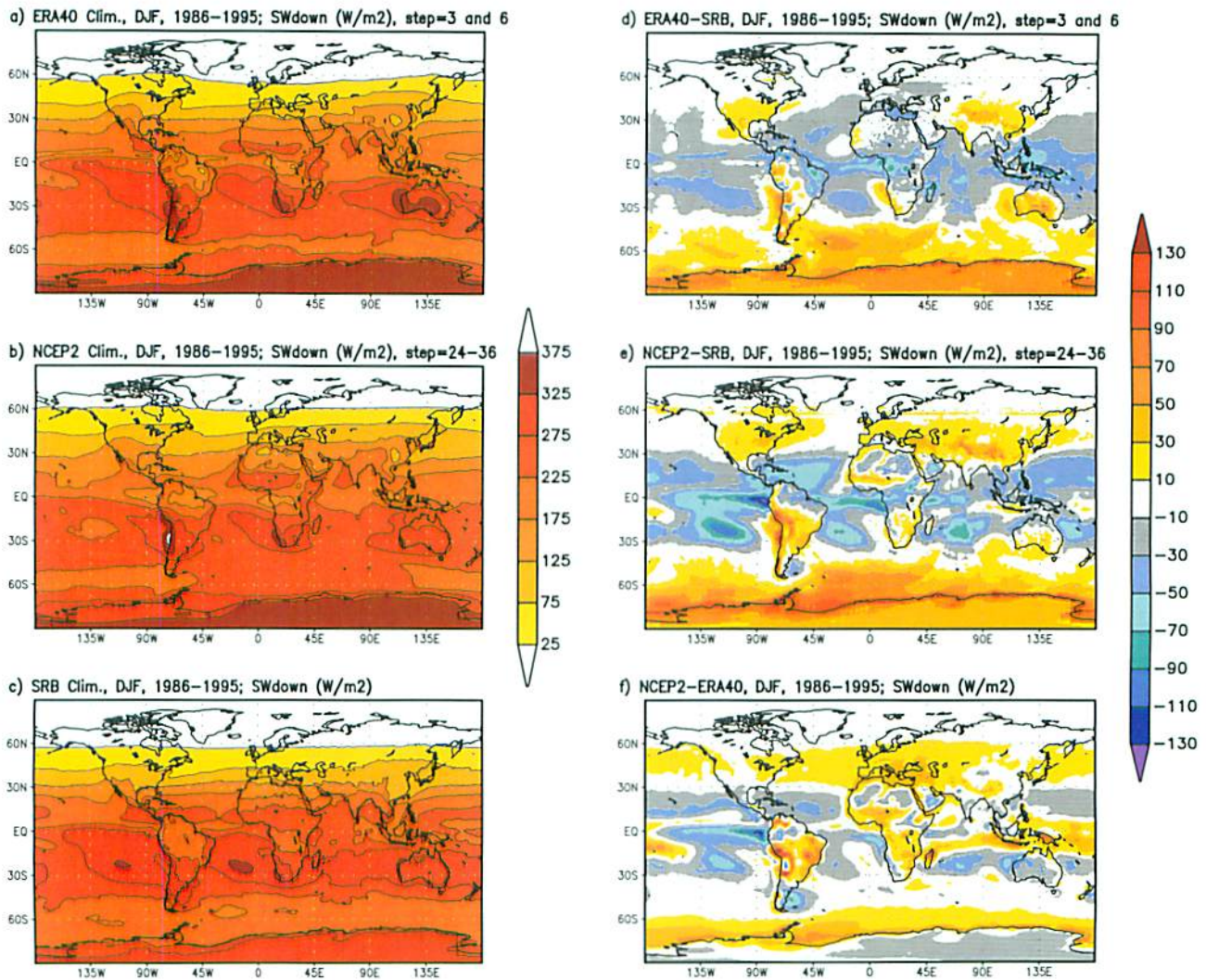


Figure 10. (a–e) As Figure 1 but for surface SWdown from reanalyses and SRB data.

ucts (Hall et al., submitted manuscript, 2006); and this important surface parameter is still not known to sufficient accuracy for climate studies.

### 2.3.2. Incoming Long-Wave Radiation Flux, LWdown

[29] LWdown depends on the atmospheric structure and composition, as well as on the cloud base temperature, and retrievals from satellite data often differ significantly. Uncertainty in cloud base temperature is a source of uncertainty in the SRB data set, and it uses atmospheric temperature and humidity profiles from GEOS-1, a different reanalysis. For the ERA40 and NCEP2 reanalyses, differences in the LW flux are primarily related to differences in model cloud cover and atmospheric temperature and moisture structure. Generally LWdown will have a low bias if low- and medium-level cloud cover is underestimated, because these clouds are nearly black in the infrared.

[30] Figure 13 shows the DJF climatology and difference fields, and Figure 14 is the corresponding JJA comparison. For ERA40 the differences from SRB are generally small over the oceans. The stratocumulus regions, which have too little low cloud, have correspondingly reduced LWdown in ERA40 and NCEP2. There are some regions in the sub-

tropics where LWdown is higher in ERA40 and NCEP2, which correspond to regions with a high cloud bias. At higher latitudes, especially in the summer hemisphere, NCEP2 has a lower LWdown over the oceans than ERA40 and SRB (by  $-10$  to  $-20 \text{ W m}^{-2}$ ). For DJF over the northern continents, ERA40 has a higher LWdown than SRB, especially over western Eurasia; for NCEP2, this region of higher LWdown is further east. When we compare the difference field (NCEP2-ERA40 in Figure 13f with Figures 1f and 10f, we see the influence of the temperature difference in eastern Russia and the added impact of reduced cloud cover over Europe in NCEP2. For JJA over the northern continents, ERA40 and SRB agree quite closely, but NCEP2 has a lower LWdown ( $-10 \text{ W m}^{-2}$ ) over land as well as over the ocean. The high values of SWdown in Figure 11e suggest that the bias of LWdown in Figure 14e is also caused by too little cloud cover. There are surprisingly large differences between the reanalyses over the Arctic in JJA: with NCEP2 values a little less than SRB and ERA40 much greater. At high latitudes in the winter hemisphere, the SRB estimates of the cloud field have larger uncertainties. In addition, large cold biases in winter in the



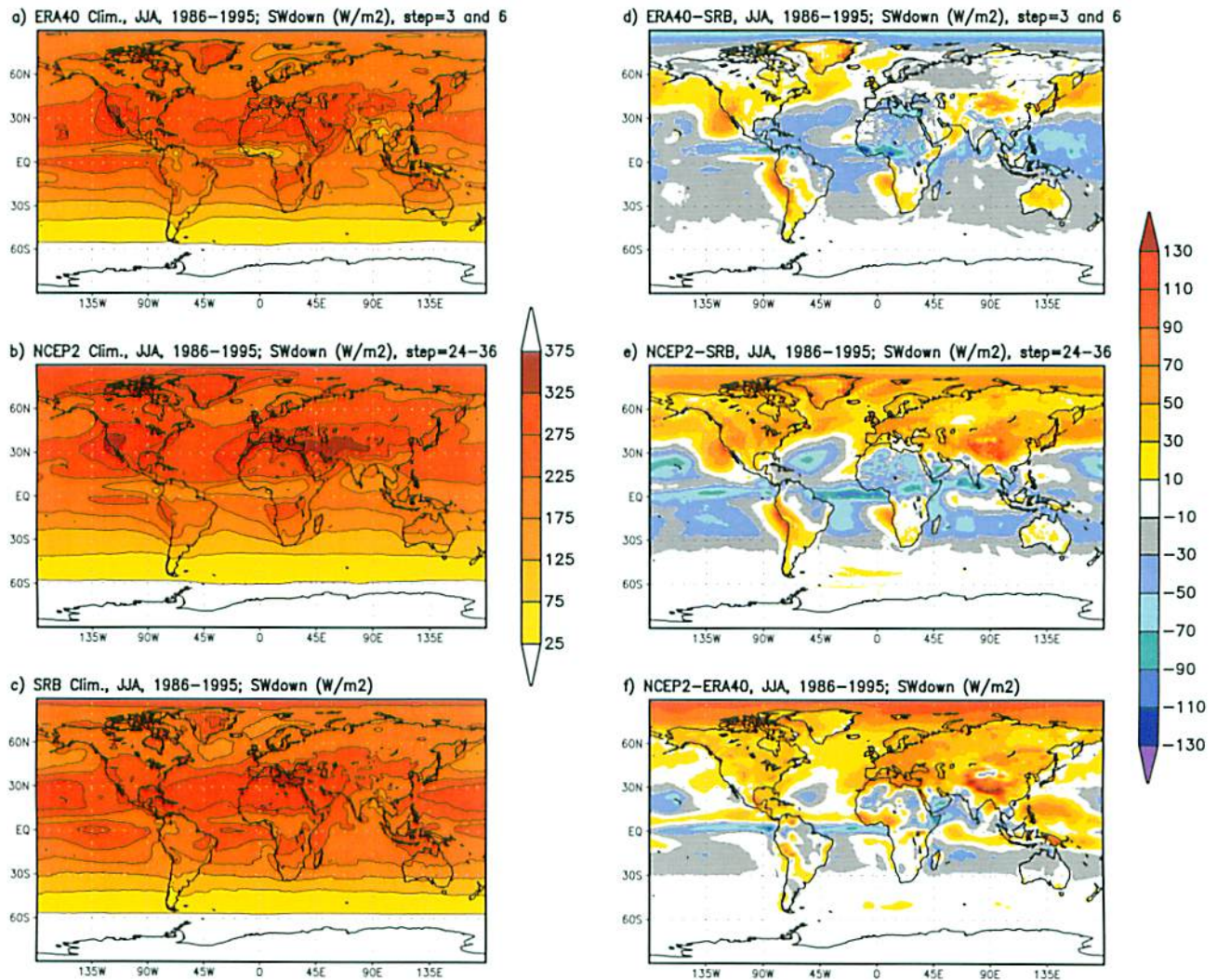


Figure 11. (a–e) As Figure 2 but for surface SWdown from reanalyses and SRB data.

skin temperatures in the GEOS-1 reanalysis adversely affect the SRB LWdown; so we cannot say with any confidence that the SRB LWdown is more accurate over the continents. Differences in the upward long-wave fluxes reflect primarily differences in radiometric skin temperature, and we do not show this comparison. The GEOS-1 reanalysis has cold biases in wintertime skin temperature, which give a large negative bias in upward LW flux [see *Betts and Beljaars, 2003*]. In addition, there is still considerable uncertainty in model skin temperatures, because these depend strongly on poorly known (as well as conceptually questionable) roughness lengths for heat [e.g., *Betts and Beljaars, 1993*].

[31] Figure 15 shows that the anomaly fields for the two reanalyses and SRB are rather similar and rather small ( $< \pm 20 \text{ W m}^{-2}$ ), suggesting again that both reanalyses represent the major changes in the atmospheric circulation. Comparing Figures 3 and 15 for DJF, shows the association of the LWdown anomalies with temperature anomalies in winter.

### 2.3.3. Net Radiation Flux, Rnet

[32] Figure 16 compares the JJA climatology and difference fields for JJA for Rnet. Comparing with the patterns in

Figure 11 for SWdown and Figure 14 for LWdown, we see that Rnet differences are dominated by SWdown differences in the tropics, while LW flux differences are dominant at high latitudes. There are of course differences in the surface albedo and the upward LW fluxes, which we have not shown. Note the large differences in Rnet over the Sahara, with NCEP2 having the lowest values, which we shall see in the next section lead to low values of sensible heat flux.

### 3. Surface Sensible and Latent Heat Fluxes (SH and LH)

[33] The partition of the surface Rnet into sensible and latent heat fluxes is of fundamental importance over land in driving the diurnal cycle of the atmospheric boundary layer. However, the two reanalyses differ substantially in this energy partition, and we have no comparison ISLSCP-II data product for evaluation. We therefore compared the sensible and latent heat fluxes from ERA40 and NCEP2 with each other, and over the oceans with the Da Silva climatology [*Da Silva et al., 1994*]. This covers the years 1945–1993, and has both the long-term mean and the



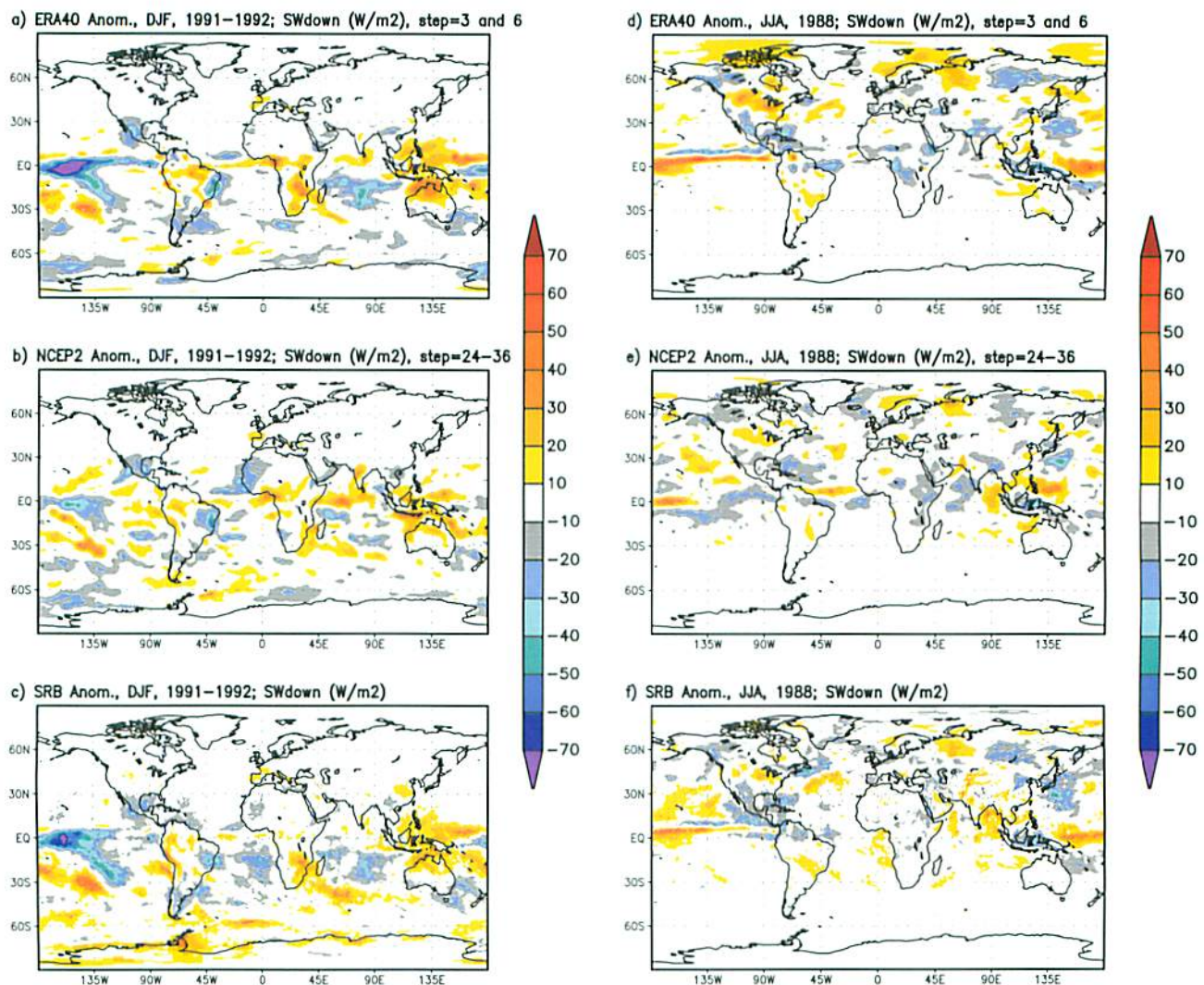


Figure 12. (a–e) As Figure 3 but for surface SWdown anomalies.

monthly anomaly fields. For our comparison we use the 8-year Da Silva climatology for 1986–1993. Six additional Figures for the climatologies, difference and anomaly fields for SH and LH are available as auxiliary material<sup>1</sup>.

### 3.1. Sensible Heat Flux

[34] ERA40 has a higher SH over most of the ocean in both seasons than the Da Silva climatology, except for a band in the southern ocean and over parts of the Gulf Stream and Kuroshio currents. In DJF in the Arctic Circle, ERA40 has a negative heat flux over the ice, whereas the Da Silva climatology has an unrealistically large upward heat flux. The NCEP2 SH fluxes are less than the Da Silva climatology over much of the oceans especially at midlatitudes in winter; so that over most of the oceans NCEP2 has a smaller SH flux than ERA40. Over land as well the NCEP2 SH flux is generally less than ERA40, except for a few regions in summer. Over the Sahara NCEP2 has rather low values of SH heat flux in JJA because of the low values

of Rnet (see Figure 16). The reanalyses have a similar structure in their anomalies however, although the NCEP2 anomalies are generally larger, consistent with Figure 3. The Da Silva anomaly fields have considerable uncertainty as observations are sparse in many regions, and we have averaged only three months. (See Figures S1, S2, and S3 in auxiliary material.)

### 3.2. Latent Heat Flux

[35] Over the oceans, evaporation is generally higher in NCEP2 than ERA40 except in the summer hemisphere, where ERA40 is higher. There is a tendency for evaporation in ERA40 over the oceans to be higher in southern hemisphere midlatitudes and lower in some regions of the tropics than the Da Silva climatology (which is again unrealistic over the Arctic in winter). Over land, NCEP2 has generally a higher latent heat flux than ERA40, which is consistent with its lower SH flux. The anomaly fields show some similarities between the reanalyses, but again the NCEP2 anomalies are larger than those of ERA40. There are considerable differences over the tropical oceans between the reanalyses for JJA 1988. Again, the Da Silva anomalies may be less

<sup>1</sup>Auxiliary materials are available in the HTML. doi:10.1029/2006JD007174.



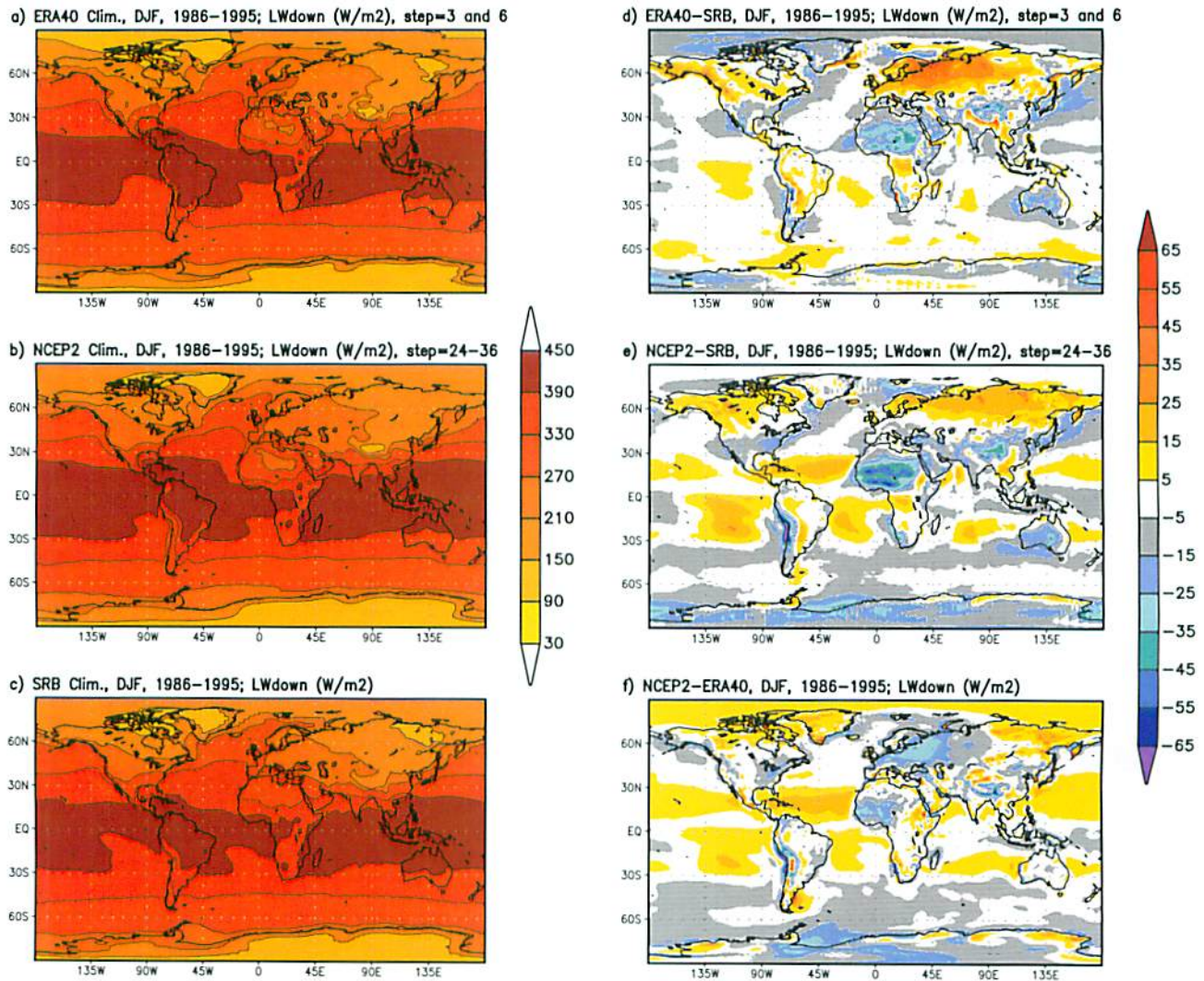


Figure 13. (a–e) As Figure 10 but for surface LWdown.

reliable on a seasonal basis. The summer drought over the USA in 1988 is picked up in both reanalyses; with a stronger signal in the NCEP2 reanalysis (consistent with Figure 6). (See Figures S4–S6 in auxiliary material.)

#### 4. Discussion

[36] We have given a broad overview on seasonal time-scales of differences at the surface by comparing the reanalysis fields with independent data from the ISLSCP-II data collection. Our purpose is to help the users of these data to assess which products might be useful for different purposes. In many cases the differences between reanalyses and other data sets give an estimate in our uncertainty in a given variable. Models not only assimilate data, but in regions of missing data a global model analysis will compute a complete set of fields using short-term forecasts of the model. Consequently, the ERA40 and NCEP2 reanalyses have biases that relate to their specific analysis-forecast systems and their choice of physical parameterizations. The comparison ISLSCP-II data sets each have their own issues of bias and representivity. In situ surface observations can be

interpolated to a common grid, as is the case for our CRU data set, but where observations are sparse; the gridded analysis may not be representative. The GPCP precipitation analysis is a blend of in situ precipitation measurements and satellite data, and biases over the open oceans and the arctic regions are unknown. The SRB data set is derived from satellite visible and infrared observations using parameterized radiation codes and atmospheric profiles coming from another earlier reanalysis with its own biases.

[37] It is clear from the figures that there are systematic differences in the climatologies of the two reanalyses and the other ISLSCP-II data sets that we have shown for comparison. For temperature and RH, we compared the reanalyses with the CRU data set [New *et al.*, 1999]. ERA40 has generally smaller biases of temperature over land than NCEP2. The CRU humidity analysis has more uncertainty in some regions because synthetic observations are added where there are few observations [New *et al.*, 2000]. ERA40 has small biases in RH, often slightly negative, and NCEP2 has in many regions a small positive RH bias. The NCEP2 RH anomalies are larger than those for ERA40, which in turn are larger than those in the CRU analysis.



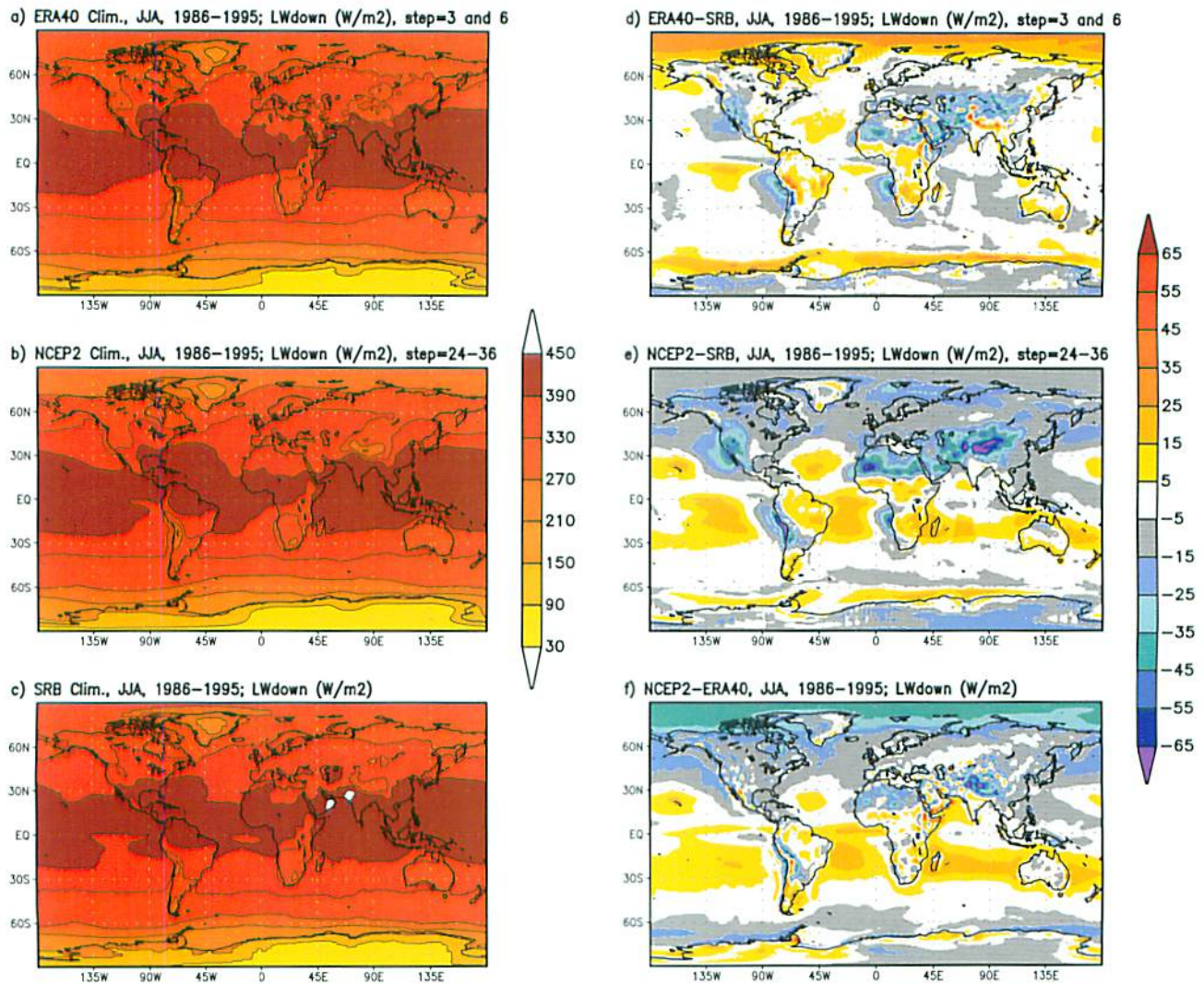


Figure 14. (a–e) As Figure 11 but for surface LWdown.

[38] For precipitation we compared with the GPCP Satellite-Gauge combination monthly precipitation data set [Adler *et al.*, 2003]. Over the tropical oceans, both reanalyses have more rainfall than GPCP, with ERA40 greater than NCEP2. This is a known error in the ERA40 reanalysis [Troccoli and Källberg, 2004]. ERA40 also has a negative bias over the Amazon in DJF. 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 from GPCP are generally smaller for ERA40.

[39] The incoming radiation fluxes at the surface are a critical component of the surface radiation budget, which are heavily modified by clouds. So we compared the downward SW and LW fluxes from the reanalyses with those in the SRB data set [Stackhouse *et al.*, 2004], which computes them from the ISCCP cloud data set [Rossow and Schiffer, 1999] using atmospheric profiles from the GEOS-1 reanalysis [Schubert *et al.*, 1995]. The SRB downward SW flux, which is derived from the observed cloud field, probably has smaller biases than the reanalysis fluxes,

which calculate their cloud fields using their model parameterizations. The biases in the climatology of the reanalyses are significant: they suggest too much reflective cloud over the tropical oceans (more for NCEP2 than ERA40), except for the stratocumulus regimes, where reanalysis cloud cover is too low. Over the northern continents in summer, NCEP2 has too little cloud cover, so that downward SW is too large. The corresponding ERA40 biases are mixed and generally smaller.

[40] The downward LW depends on the atmospheric structure and composition, as well as cloud base temperature. Uncertainty in cloud base temperature is a source of uncertainty in the SRB data set, as are atmospheric boundary layer temperatures, which come from the GEOS-1 reanalysis. For the reanalyses, differences in the LW flux are primarily related to differences in model cloud cover and atmospheric temperature and moisture structure. Generally, over the tropical oceans, the differences in downward LW for the reanalyses come from differences in cloud cover; so that there is a positive bias where model cloud cover is too high (trade-wind regimes) and a negative bias where model cloud cover is too low (stratocumulus regimes, and northern



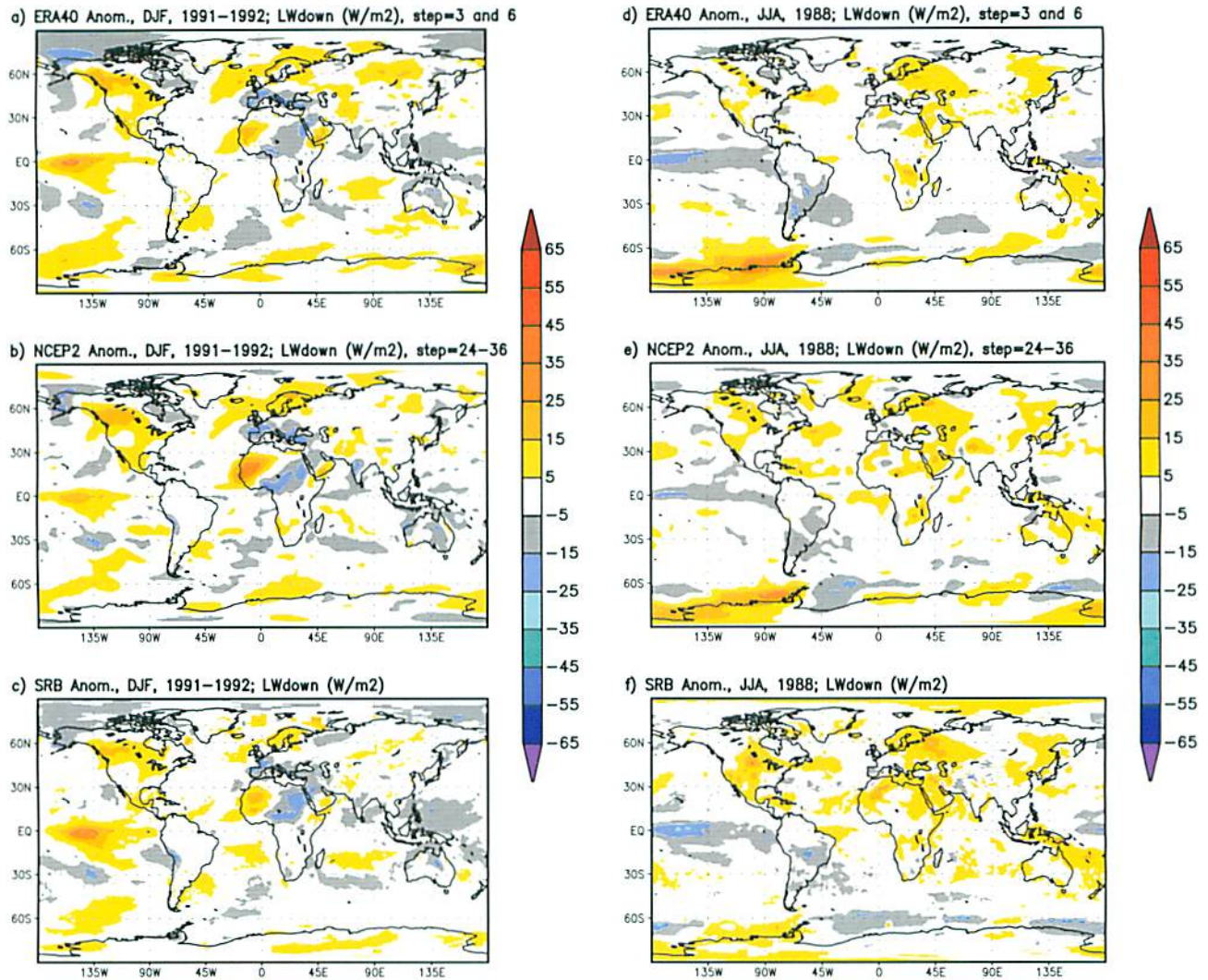


Figure 15. (a–e) As Figure 12 but for surface LWdown anomalies.

summer midlatitudes for NCEP2). Over the winter continents, the SRB LW has significant biases coming from a cold bias in the near-surface temperatures in the GEOS-1 reanalysis. We showed a single comparison of the surface net radiation flux for the northern summer. In the tropics, differences in cloud cover lead to differences in SWdown and Rnet, while at high latitudes the substantial LW flux differences dominate. Rnet of course drives the important surface energy budget over land.

[41] The two reanalyses differ considerably in their SH and LH fluxes over land, which are of importance to ISLSCP; yet we have no comparison data set over land for evaluation. For many regions over land NCEP2 has more evaporation and reduced SH in comparison to ERA40, but larger seasonal anomalies. The larger seasonal anomalies for the NCEP2 reanalysis are consistent with *Dirmeyer et al.* [2004], who noted that the annual range of soil wetness was smaller for ERA40 than NCEP2. *Ferranti and Viterbo* [2006] concluded from a study of the European summer of 2003 that variability in the European Centre model climate is dampened by the soil moisture increments. It is likely that the different methods used in the reanalyses'

land surface models for controlling drifts of soil wetness (see sections 1.2.1 and 1.2.2) are responsible, so uncertainty remains in the energy flux partition over land.

[42] These biases in the climatologies of the reanalyses, as well as the uncertainties in the data sets which we have compared them with, must be considered by users of the data sets in the ISLSCP-II collection. In contrast, a striking and encouraging feature of all our comparisons is that the anomaly fields derived from each climatology show similar features in both data sets and reanalyses. The anomalies are also coherent in the sense that we see warm, dry seasonal biases associated with reduced precipitation and cloudiness, and the converse. This confirms that major changes in the atmospheric circulation patterns, with the associated differences in surface temperature, humidity, precipitation, cloud fields and incoming surface SW and LW radiation fluxes are captured by both reanalyses and our other comparison ISLSCP-II data sets. Thus the anomaly fields are useful in identifying seasonal changes in the climatological patterns. For precipitation and SWdown, the large-scale anomalies for the higher resolution ERA40 appear to be a little closer to the comparison ISLSCP-II data set than are the NCEP2



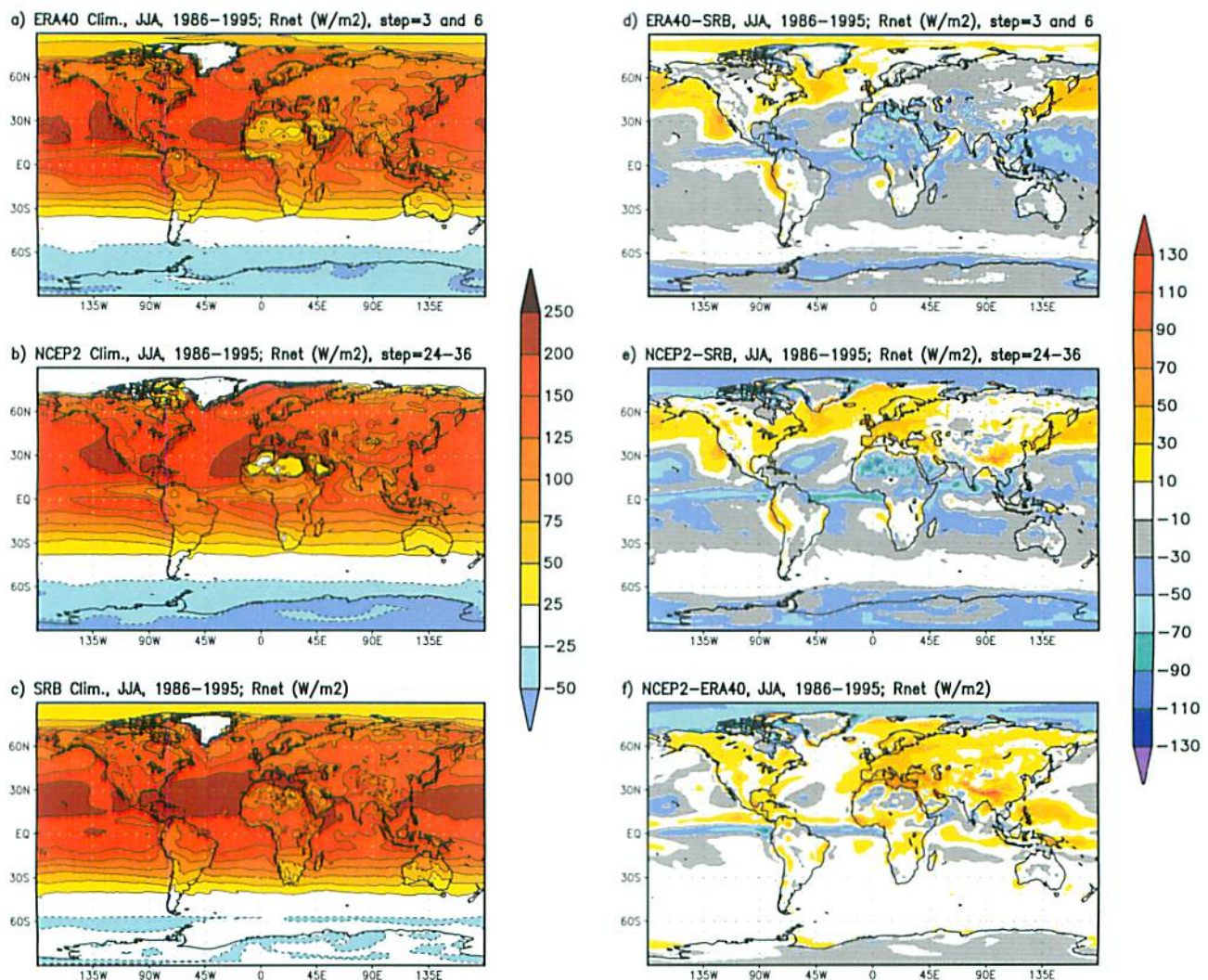


Figure 16. (a–e) As Figure 11 but for surface Rnet.

anomalies, but there are regions over land where this is not so. Our advice to the user of these ISLSCP-II data can only be very general, since these data will be used for such a wide variety of analyses.

[43] **Acknowledgments.** Alan Betts acknowledges support from NSF under grant ATM-0529797 and from the NASA under NEWS grant NNG05GQ88A and grant NAS5-11578 (which also supported the production of the ISLSCP-II data set at ECMWF). The COLA portion of this work was supported by NASA grant NAG5-11579 as well as by omnibus support at the Center for Ocean-Land-Atmosphere Studies, from NSF grant ATM 9814265, NOAA grant NA96GP0056 and NASA grant NAG5-8202. The authors would like to thank both reanalysis teams and all the ISLSCP-II data providers, especially George Huffman for providing the GPCP data set and Paul Stackhouse and Shashi Gupta for the SRB data set.

## References

- Adler, R. F., et al. (2003), The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–Present), *J. Hydrometeorol.*, **4**, 1147–1167.
- Betts, A. K. (2006), Radiative scaling of the nocturnal boundary layer and the diurnal temperature range, *J. Geophys. Res.*, **111**, D07105, doi:10.1029/2005JD006560.
- Betts, A. K., and A. Beljaars (1993), Estimation of effective roughness length for heat and momentum from FIFE data, *Atmos. Res.*, **30**, 251–261.
- Betts, A. K., and A. C. M. Beljaars (2003), ECMWF ISLSCP-II near-surface dataset from ERA-40, *ERA-40 Proj. Rep. Ser.* **8**, Eur. Cent. for Med.-Range Weather Forecasts, Reading, U. K. (Available at [http://www.ecmwf.int/publications/library/ecpublications/\\_pdf/era40/ERA40\\_PRS\\_8.pdf](http://www.ecmwf.int/publications/library/ecpublications/_pdf/era40/ERA40_PRS_8.pdf)).
- 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., and P. Viterbo (2005), Land-surface, boundary layer and cloud-field coupling over the south-western Amazon in ERA-40, *J. Geophys. Res.*, **110**, D14108, doi:10.1029/2004JD005702.
- Betts, A. K., P. Viterbo, A. C. M. Beljaars, and B. J. J. M. van den Hurk (2001), Impact of BOREAS on the ECMWF Forecast Model, *J. Geophys. Res.*, **106**, 33,593–33,604.
- Betts, A. K., J. H. Ball, M. Bosilovich, P. Viterbo, Y.-C. Zhang, and W. B. Rossow (2003a), Intercomparison of water and energy budgets for five Mississippi sub-basins between ECMWF reanalysis (ERA-40) and NASA-DAO fvGCM for 1990–1999, *J. Geophys. Res.*, **108**(D16), 8618, doi:10.1029/2002JD003127.
- Betts, A. K., J. H. Ball, and P. Viterbo (2003b), Evaluation of the ERA-40 surface water budget and surface temperature for the Mackenzie River basin, *J. Hydrometeorol.*, **4**, 1194–1211.
- Betts, A. K., J. H. Ball, P. Viterbo, A. Dai, and J. A. Marengo (2005), Hydro-meteorology of the Amazon in ERA-40, *J. Hydrometeorol.*, **6**, 764–774.
- Chou, M.-D., and K.-T. Lee (1996), Parameterizations for the absorption of solar radiation by water vapor and ozone, *J. Atmos. Sci.*, **53**, 1203–1208.
- Cox, S. J., P. W. Stackhouse Jr., S. K. Gupta, J. C. Mikovitz, M. Chiacchio, and T. Zhang (2004), The NASA/GEWEX Surface Radiation Budget Project: Results and analysis, paper presented at International Radiation Symposium, Int. Radiat. Comm., Busan, South Korea, 23–27 Aug.

- Da Silva, A., C. C. Young, and S. Levitus (1994), *Atlas of Surface Marine Data*, vol. 1, *Algorithms and Procedures*, NOAA Atlas NESDIS 6, 83 pp., U.S. Dep. of Commer., Washington, D. C.
- Dirmeyer, P. A., Z. Guo, and X. Gao (2004), Comparison, validation and transferability of eight multi-year global soil wetness products, *J. Hydrometeorol.*, **5**, 1011–1033.
- Dirmeyer, P. A., X. Gao, M. Zhao, Z. Guo, T. Oki, and N. Hanasaki (2006), The Second Global Soil Wetness Project (GSWP-2): Multi-model analysis and implications for our perception of the land surface, *Bull. Am. Meteorol. Soc.*, in press.
- Douville, H., P. Viterbo, J.-F. Mahfouf, and A. C. M. Beljaars (2000), Evaluation of optimal interpolation and nudging techniques for soil moisture analysis using FIFE data, *Mon. Weather Rev.*, **128**, 1733–1756.
- Ferranti, L., and P. Viterbo (2006), The European summer of 2003: Sensitivity to soil water initial conditions, *J. Clim.*, in press.
- Gleckler, P. (Ed.) (1996), AMIP II guidelines, *AMIP Newsl.*, **8**, 20 pp. (Available at <http://www-pcmdi.llnl.gov/projects/amip/NEWS/amipn8.pdf>)
- 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.
- Hagemann, S., K. Arpe, and L. Bengtsson (2005), Validation of the hydrological cycle of ERA-40, *ERA-40 Proj. Rep.*, **24**, 31 pp., Eur. Cent. for Med.-Range Weather Forecasts, Reading, U. K. (Available at [http://www.ecmwf.int/publications/library/ecpublications/\\_pdf/era40/ERA40\\_PRS24.pdf](http://www.ecmwf.int/publications/library/ecpublications/_pdf/era40/ERA40_PRS24.pdf))
- Hayasaka, T., K. Kawamoto, G. Shi, and A. Ohmura (2006), Importance of aerosols in satellite-derived estimates of surface shortwave irradiance of China, *Geophys. Res. Lett.*, **33**, L06802, doi:10.1029/2005GL025093.
- Hong, S.-Y., and H.-L. Pan (1996), Nonlocal boundary layer vertical diffusion in a medium-range forecast model, *Mon. Weather Rev.*, **124**, 2322–2339.
- Källberg, P., A. Simmons, S. Uppala, and M. Fuentes (2004), The ERA-40 archive, *ERA-40 Proj. Rep.*, **17**, 31 pp., Eur. Cent. for Med.-Range Weather Forecasts, Reading, U. K. (Available at [http://www.ecmwf.int/publications/library/ecpublications/\\_pdf/era40/ERA40\\_PRS17.pdf](http://www.ecmwf.int/publications/library/ecpublications/_pdf/era40/ERA40_PRS17.pdf))
- 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.
- Lu, C.-H., M. Kanamitsu, J. O. Roads, W. Ebisuzaki, and K. E. Mitchell (2005), Evaluation of soil moisture in the NCEP-NCAR and NCEP-DOE global reanalyses, *J. Hydrometeorol.*, **6**, 391–408.
- Masuda, K. (2004), Surface radiation budget: Comparison between global satellite-derived products and land-based observations in Asia and Oceania, paper presented at International Radiation Symposium 2004, Int. Radiat. Comm., Busan, South Korea, Aug. (Available at <http://www.jamstec.go.jp/frcgc/research/p2/masuda/radcmp/>)
- Maurer, E. P., G. M. O'Donnell, D. P. Lettenmaier, and J. O. Roads (2001), Evaluation of the land surface water budget in NCEP/NCAR and NCEP/DOE reanalyses using an off-line hydrologic model, *J. Geophys. Res.*, **106**(D16), 17,841–17,862.
- 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. Jones (2000), Representing twentieth-century space-time climate variability. Part II: Development of 1901–1996 monthly grids of terrestrial surface climate, *J. Clim.*, **13**, 2217–2238.
- Pan, H.-L., and L. Mahrt (1987), Interaction between soil hydrology and boundary-layer development, *Boundary Layer Meteorol.*, **38**, 185–202.
- Parrish, D. F., and J. C. Derber (1992), The National Meteorological Center's spectral statistical-interpolation analysis system, *Mon. Weather Rev.*, **120**, 1747–1763.
- Pinker, R. T., and I. Laszlo (1992), Modeling surface solar irradiance for satellite applications on a global scale, *J. Appl. Meteorol.*, **31**, 194–211.
- Roads, J. (2003), The NCEP/NCAR, NCEP/DOE and TRMM tropical atmosphere hydrologic cycles, *J. Hydrometeorol.*, **4**, 826–840.
- Roads, J., M. Kanamitsu, and R. Stewart (2002), CSE water and energy budgets in the NCEP-DOE reanalysis II, *J. Hydrometeorol.*, **3**(3), 227–248.
- Roads, J., et al. (2003), GCIP water and energy budget synthesis (WEBS), *J. Geophys. Res.*, **108**(D16), 8609, doi:10.1029/2002JD002583.
- Rossow, W. B., and R. A. Schiffer (1999), Advances in understanding clouds from ISCCP, *Bull. Am. Meteorol. Soc.*, **80**, 2261–2287.
- Schubert, S., C.-K. Park, C.-Y. Wu, W. Higgins, Y. Kondratyeva, A. Molod, L. Takacs, M. Seablom, and R. Rood (1995), A multiyear assimilation with the GEOS-1 system: Overview and results, *NASA Tech. Memo. 104606*, 6. (Available at <http://gmao.gsfc.nasa.gov/pubs/docs/Schubert124.pdf>)
- Simmons, A. J., and J. K. Gibson (2000), The ERA-40 project plan, *ERA-40 Proj. Rep. Ser.*, **1**, 63 pp., Eur. Cent. for Med.-Range Weather Forecasts, Reading, U. K.
- Simmons, A. J., P. D. Jones, V. da Costa Bechtold, A. C. M. Beljaars, P. W. Källberg, S. Saarinen, S. M. Uppala, P. Viterbo, and N. Wedi (2004), Comparison of trends and variability in CRU, ERA-40 and NCEP/NCAR analyses of monthly-mean surface air temperature, *ERA-40 Proj. Rep. Ser.*, **18**, Eur. Cent. for Med.-Range Weather Forecasts, Reading, U. K. (Available at [http://www.ecmwf.int/publications/library/ecpublications/\\_pdf/era40/ERA40\\_PRS18.pdf](http://www.ecmwf.int/publications/library/ecpublications/_pdf/era40/ERA40_PRS18.pdf))
- Stackhouse, P. W., Jr., S. K. Gupta, S. J. Cox, J. C. Mikovitz, T. Zhang, and M. Chiacchio (2004), 12-year surface radiation budget data set, *GEWEX News*, **14**(4), 10–12.
- 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\\_PRS13.pdf](http://www.ecmwf.int/publications/library/ecpublications/_pdf/era40/ERA40_PRS13.pdf))
- Uppala, S. M., et al. (2005), The ERA-40 reanalysis, *Q. J. R. Meteorol. Soc.*, **131**, 2961–3012.
- Van den Hurk, B. J. J. M., P. Viterbo, A. C. M. Beljaars, and A. K. Betts (2000), Offline validation of the ERA-40 surface scheme, *ECMWF Tech. Memo*, **295**, 43 pp., Eur. Cent. for Med.-Range Weather Forecasts, Reading, U. K. (Available at [http://www.ecmwf.int/publications/library/ecpublications/\\_pdf/tm/001-300/tm295.pdf](http://www.ecmwf.int/publications/library/ecpublications/_pdf/tm/001-300/tm295.pdf))
- Xie, P., and P. A. Arkin (1997), Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs, *Bull. Am. Meteorol. Soc.*, **78**, 2539–2558.
- Zhao, M., and P. A. Dirmeyer (2003), Production and analysis of GSWP-2 near-surface meteorology data sets, *Publ. 159*, Cent. for Ocean-Land-Atmos. Stud., Calverton, Md. (Available at [ftp://grads.iges.org/pub/ctr/ctr\\_159.pdf](ftp://grads.iges.org/pub/ctr/ctr_159.pdf))

A. C. M. Beljaars, European Center for Medium-Range Weather Forecasts, Reading RG2 9AX, UK. ([anton.beljaars@ecmwf.int](mailto:anton.beljaars@ecmwf.int))

A. K. Betts, Atmospheric Research, Pittsford, VT 05763, USA. ([akbetts@aol.com](mailto:akbetts@aol.com))

P. A. Dirmeyer and M. Zhao, Center for Ocean-Land-Atmosphere Studies, Calverton, MD 20705, USA. ([dirmeyer@cola.iges.org](mailto:dirmeyer@cola.iges.org); [mzhao@cola.iges.org](mailto:mzhao@cola.iges.org))



# Revisiting a hydrological analysis framework with International Satellite Land Surface Climatology Project Initiative 2 rainfall, net radiation, and runoff fields

Randal D. Koster,<sup>1</sup> Balázs M. Fekete,<sup>2</sup> George J. Huffman,<sup>3</sup> and Paul W. Stackhouse Jr.<sup>4</sup>

Received 9 February 2006; revised 17 July 2006; accepted 4 August 2006; published 23 November 2006.

[1] The International Satellite Land Surface Climatology Project Initiative 2 (ISLSCP-2) data set provides the data needed to characterize the surface water budget across much of the globe in terms of energy availability (net radiation) and water availability (precipitation) controls. The data, on average, are shown to be consistent with Budyko's decades-old framework, thereby demonstrating the continuing relevance of Budyko's semiempirical relationships. This consistency, however, appears only when a small subset of the data with hydrologically suspicious behavior is removed from the analysis. In general, the precipitation, net radiation, and runoff data also appear consistent in their interannual variability and in the phasing of their seasonal cycles.

**Citation:** Koster, R. D., B. M. Fekete, G. J. Huffman, and P. W. Stackhouse Jr. (2006), Revisiting a hydrological analysis framework with International Satellite Land Surface Climatology Project Initiative 2 rainfall, net radiation, and runoff fields, *J. Geophys. Res.*, **111**, D22S05, doi:10.1029/2006JD007182.

## 1. Introduction

[2] The International Satellite Land Surface Climatology Project Initiative 2 (ISLSCP-2) data set [Hall *et al.*, 2006] provides global, coregistered, high-resolution ( $1^\circ \times 1^\circ$ ) fields of land surface meteorological and hydrological data for scientific analysis. Naturally, many of the variables in the data set are intrinsically related to each other. This raises the question of data consistency: given that different measurement systems are used to generate, for example, the rainfall and runoff components of the data set, are the values provided for these two variables mutually consistent? To address this question, the variables must be examined jointly in a sensible way.

[3] Budyko [1958, 1974] pioneered the joint analysis of land surface hydroclimatological variables. He recognized that evaporation from the land surface requires both water at the surface and incident energy at the surface, the latter for the change in phase of water. Mean annual evaporation ( $E$ ), he argued, is thus effectively controlled by two things: the mean annual precipitation ( $P$ ) and the mean annual net radiation ( $R_{net}$ ). When  $P \ll R_{net}/\lambda$ , where  $\lambda$  is the latent heat of vaporization, the evaporation is water limited, and  $E \approx P$ . On the other hand, when  $P \gg R_{net}/\lambda$ , the evaporation is energy limited, and  $E \approx R_{net}/\lambda$ . When  $P$  and  $R_{net}/\lambda$  are of

the same order,  $E$  will be lower than both, since additional issues then come into play. In essence, Budyko realized that a region's water balance cannot be estimated from precipitation or net radiation data alone; the two forcings must be considered together in a single framework.

[4] The joint controls are summarized in Budyko's plot of  $E/P$  versus dryness index  $D$ , defined as  $D = R_{net}/P\lambda$  (Figure 1). The solid curve in Figure 1 is defined by the equation

$$E/P = \mathcal{F}(D) = [D (\tanh 1/D)(1 - \cosh D + \sinh D)]^{1/2} \quad (1)$$

The curve is a semiempirical fit to the limited observations available to Budyko when he performed his analysis. When  $D$  is large, water-limited conditions prevail, and the curve approaches the horizontal line (equivalent to the condition  $E = P$ ). When  $D$  is small, energy-limited conditions prevail, and the curve approaches the 1:1 line, equivalent to the condition  $E = R_{net}/\lambda$ . For all values of  $D$ , the curve allows a first-order estimate of the annual evaporation rate.

[5] The curve works well on average for the regions investigated by Budyko [1974], and it describes well the evaporation rates generated by various atmospheric general circulation models [Koster *et al.*, 2001]. As a research framework, it can be used to understand interannual variations in the water cycle; Koster and Suarez [1999] use (1) to derive the following equation for the standard deviation of annual evaporation,  $\sigma_E$ :

$$\sigma_E/\sigma_P = \mathcal{F}(D) - D\mathcal{F}'(D) \quad (2)$$

where  $\sigma_P$  is the standard deviation of annual precipitation,  $\mathcal{F}(D)$  is the function described in (1), and  $\mathcal{F}'(D)$  is the first derivative of  $\mathcal{F}(D)$  with respect to  $D$ . Equation (2),

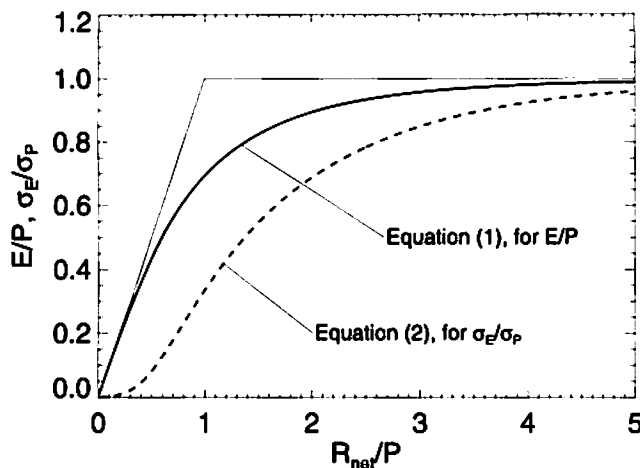
<sup>1</sup>Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.

<sup>2</sup>Water Systems Analysis Group, Complex Systems Research Center, Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, Durham, New Hampshire, USA.

<sup>3</sup>Science Systems and Applications, Inc., NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.

<sup>4</sup>NASA Langley Research Center, Hampton, Virginia, USA.





**Figure 1.** Curves derived by *Budyko [1958]* for  $E/P$  and *Koster and Suarez [1999]* for  $\sigma_E/\sigma_P$ .

illustrated by the dotted line in Figure 1, successfully characterizes the interannual variability of evaporation in a number of AGCMs [*Koster et al., 2001*].

[6] Although Budyko's semiempirical relationships are useful for characterizing first-order controls on the mean and variability of annual evaporation and runoff and for guiding further inquiry into the global surface water balance, they have rarely been tested on a global scale using purely observational data. *Milly and Dunne [2002]*, as part of a broad analysis addressing controls on runoff variability, examined observed precipitation, radiation, and runoff data over a number of basins across the globe and found some discrepancies between the observations and Budyko-based estimates of evaporation. Now, with the advent of the global  $1^\circ \times 1^\circ$  ISLSCP-2 observational data set, we can test the Budyko framework even more comprehensively. Several questions suggest themselves: Do Budyko's relationships still hold up? Is the framework still valid for exploring controls on global hydrology? Can the framework be used to identify deficiencies or inconsistencies in the data sets? In this paper, we use the ISLSCP-2 precipitation, net radiation, and runoff fields to revisit and reevaluate Budyko's framework. We describe the data sets used in section 2, and we identify some clear inconsistencies and some suspicious behavior among the data in section 3. In section 4, we test the Budyko framework with the global data.

## 2. Data Used

### 2.1. Precipitation

[7] The Global Precipitation Climatology Project (GPCP) Version 2 Satellite-Gauge (SG) precipitation data set, consisting of globally complete  $2.5^\circ \times 2.5^\circ$  monthly estimates, is based on a variety of satellite data sets plus gauge analyses [*Adler et al., 2003*] (see also [ftp://precip.gsfc.nasa.gov/pub/gpcp-v2/doc/V2\\_doc](http://precip.gsfc.nasa.gov/pub/gpcp-v2/doc/V2_doc)). The SG data set extends from 1979 to present. Beginning in July 1987, but not including December 1987, the data set incorporates Special Sensor Microwave/Imager (SSM/I) passive microwave estimates at low latitudes and midlatitudes and Television-Infrared Observation Satellite (TIROS) Operational Vertical Sounder (TOVS) estimates at higher latitudes.

Over the latitude band  $40^\circ\text{N-S}$ , infrared brightness temperatures (IR  $T_b$ s) are converted into precipitation estimates by applying the Adjusted Geostationary Operational Environmental Satellite (GOES) Precipitation Index (AGPI [*Adler et al., 1994*]), which depends on calendar month calibrations computed with approximately time/space matched IR  $T_b$ 's and SSM/I rain estimates. A MultiSatellite (MS) estimate is composited from these inputs, and then the SG product is produced in two steps: (1) The MS estimate is adjusted to the large-scale gauge average for each grid box over land. (2) The gauge-adjusted MS estimate and the gauge analysis are combined in a weighted average, where the weights are the inverse (estimated) error variance of the respective estimates.

[8] The period before the start of SSM/I observations requires a more approximate scheme. During the period 1986 to June 1987, plus December 1987, outgoing long-wave radiation (OLR) precipitation index (OPI) data are climatologically calibrated by the 1988–1996 GPCP SG estimates and used in place of the SSM/I–TOVS component. The MS field is built from geo-AGPI estimates where available ( $40^\circ\text{N-S}$ ) and calibrated OPI estimates elsewhere, and the results are combined with the gauge data as in the recent era to produce the SG product.

[9] *Hall et al. [2006]* provide a discussion of SG data quality specifically directed at ISLSCP-2 users. The ancillary data documentation and supplemental fields (e.g., rain gauge density) provided by ISLSCP-2 contain important additional information on the precipitation data's uncertainty. This uncertainty has many sources: a lack of gauges in many parts of the world; the tendency to locate gauges in developed areas (e.g., in the valleys rather than at high elevations, where more of the rain falls); inhomogeneity in instrumentation and reporting methods; and uneven skill by remote sensing algorithms, which have particular difficulties, for example, over frozen surfaces. Overall, the data we examine here tend to be more uncertain in mountains, deserts, high latitudes, and areas lacking development and/or suffering societal upheavals. Biases, however, are considered relatively small, even in gauge-sparse regions. The exception is in mountainous regions, where the gauge (and therefore the SG product) will generally underestimate the true precipitation because of the aforementioned biased siting of gauges. The data account explicitly for gauge undercatch.

### 2.2. Net Radiation

[10] The Surface Radiation Budget (SRB) project at NASA is a component of the Global Energy and Water Cycle Experiment (GEWEX), under the auspices of the World Climate Research Programme (WCRP). SRB produces estimates of surface radiative flux quantities by processing satellite observations, reanalysis meteorology, and ozone measurements through parameterized radiation models. For the ISLSCP-2 data set, monthly averaged SRB (Release 2.0) solar (SW, or shortwave) and thermal infrared (LW, or longwave) data [*Cox et al., 2004; Stackhouse et al., 2004*] were supplied at a resolution of  $1^\circ$  latitude  $\times$   $1^\circ$  longitude with the algorithms described below.

[11] The radiation fluxes for ISLSCP-2 were computed using the International Satellite Cloud Climatology Project (ISCCP) "DX" data [*Rossow and Schiffer, 1999*], using an

averaging approach analogous to that of Rossow *et al.* [1996] to produce cloud and surface properties at a  $1^\circ$  spatial and 3-hourly temporal resolution. The needed meteorological profile information (e.g., temperature and humidity) was provided by the Goddard Earth Observing System (GEOS) v.1 reanalysis [Schubert *et al.*, 1995], which was generated by the Data Assimilation Office (DAO, now the Global Modeling and Assimilation Office) of NASA's Goddard Space Flight Center. Additional data used include the  $1.25^\circ$  longitude  $\times$   $1^\circ$  latitude column ozone fields derived from Total Ozone Mapping Spectrometer (TOMS) data. High-resolution classification maps of land surface type helped in the assignment of surface spectral albedo and emissivities [Charlock and Alberta, 1996; Wilber *et al.*, 1999].

[12] The shortwave fluxes submitted to ISLSCP-2 were computed from an upgraded version of the algorithm of Pinker and Ewing [1985]. This algorithm computes a broadband solar flux for each time stamp using a spectral two-stream delta-Eddington model to map broadband reflected fluxes at the top of the atmosphere (TOA) to transmitted fluxes at the surface. The reflected fluxes at TOA are computed using narrowband-to-broadband relationships on the visible radiances and angular distribution models (ADMs) from the Earth Radiation Budget Experiment (ERBE). The broadband surface albedo is retrieved at every step using the ISCCP background surface radiance and assumed column aerosol, precipitable water and column ozone. The spectral shape of the albedo is fixed according to the surface type prevalent in the grid box. This surface albedo is then used to infer the transmitted flux, weighing the contributions of cloudy and clear-sky fluxes. The surface albedo and the transmitted flux are used to infer the reflected surface flux of the grid box. The model has been updated with new water vapor parameterizations and averaging schemes.

[13] The SRB SW Release 2.0 monthly averaged fluxes were compared to surface measurements across the globe from the Baseline Surface Radiation Network (BSRN [Ohmura *et al.*, 1998]) and the Global Energy Balance Archive (GEBA [Gilgen and Ohmura, 1999]). Comparisons to the GEBA data set suggest that the SRB downwelling fluxes have a bias and RMSE of 1.8% and 15.2%, respectively, whereas comparisons with the BSRN data suggest a bias and RMSE of  $-2.2\%$  and  $14.3\%$ , respectively. Estimating errors in upward fluxes from surface measurements is problematic because of spatial heterogeneity, so we estimate the uncertainty in the net fluxes using the above comparisons between the downwelling fluxes and flux albedo differences between SRB and ISCCP-FD [Zhang *et al.*, 2004], as documented by Zhang *et al.* [2006]. Zhang *et al.* [2006] show that SRB albedos in snow-free regions are 2% higher (4.6% standard deviation) than ISCCP-FD values. For snow covered landmass areas, the differences are much larger, and SRB albedos are  $11.2\%$  (8.3% standard deviation) lower than ISCCP-FD values. However, subsequent analysis shows that ISCCP-FD and SRB differences in the downwelling SW tend to compensate for albedo differences. Using these differences as a proxy for the uncertainty in the SRB surface SW net fluxes, the uncertainty range is expected to be within  $\pm 10 \text{ W m}^{-2}$  overall (about 8% of the globally averaged surface net flux), with

the largest uncertainties over snow covered areas. However, because the polar regions are excluded from this study, much of these snow covered areas are not considered. Last, we note that SW errors vary with altitude to a maximum negative bias of  $50 \text{ W m}^{-2}$  in summer in the Tibetan plateau region. In this case, the surface albedo was also too low so that the net flux is more likely to be biased high.

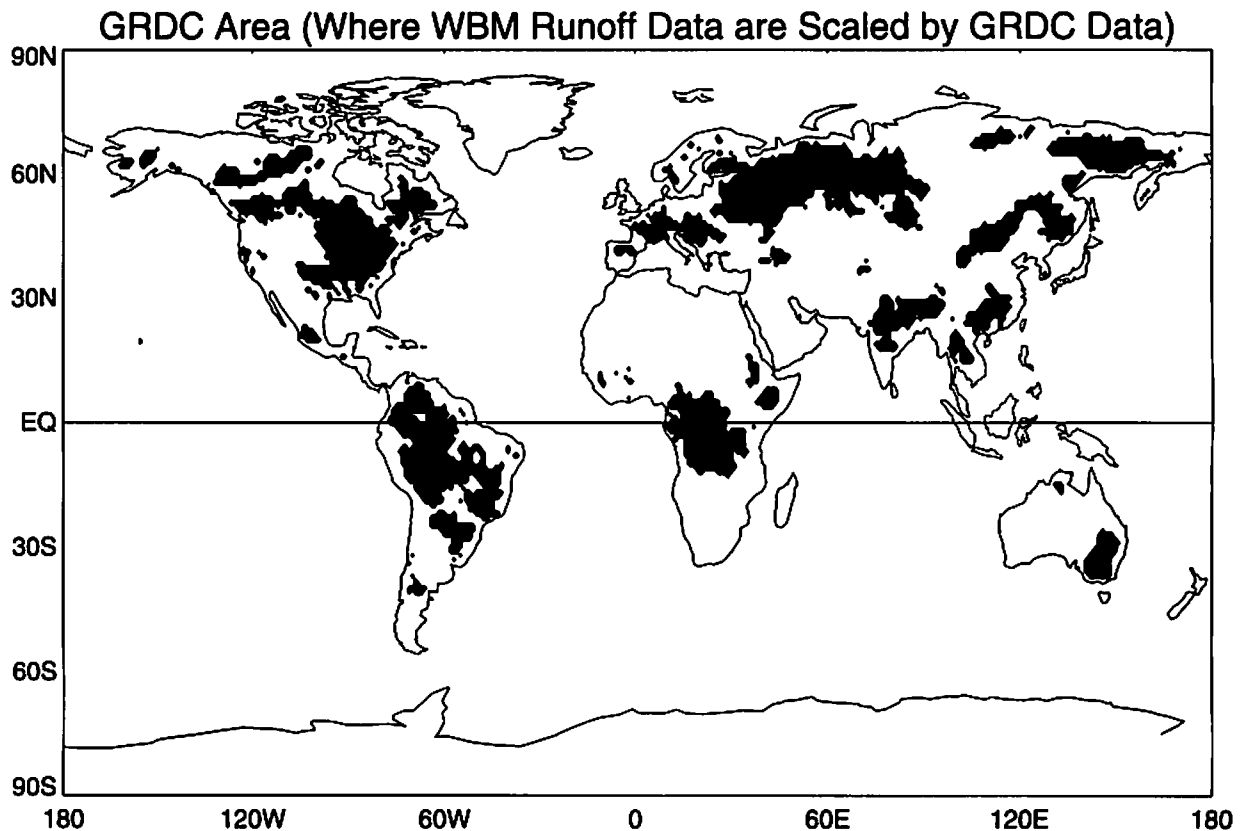
[14] For the ISLSCP-2 longwave fluxes, the GEWEX SRB Longwave Quality Check algorithm was used (SRB Rel. 2.0). This algorithm, a slightly upgraded version of the algorithm of Gupta *et al.* [1999], uses broadband parameterizations of narrow band ( $10 \text{ cm}^{-1}$ ) radiative transfer calculations to compute a clear-sky longwave flux given the meteorological profile (water vapor and temperature) of the grid box. The model uses cloud fraction and the cloud top temperatures to prescribe the effects of clouds on the clear-sky downwelling flux. TOA fluxes are currently not computed with this algorithm, but the model does allow for nonblack surface emittances. These nonblack broadband surface emittances are used in conjunction with the diurnally varying GEOS-1 surface temperatures to compute the upwelling longwave fluxes.

[15] When the ISLSCP downwelling longwave fluxes are compared to BSRN measurements (from 1992 to 1995), the mean bias is about  $5 \text{ W m}^{-2}$  (with model fluxes higher), and the random error is about  $\pm 15 \text{ W m}^{-2}$ . This mean bias is considered to be within the uncertainty for BSRN measurements. On the basis of estimates of surface temperature reliability for the GEOS-1 modeling system, we infer a mean bias of order  $\pm 2\text{--}4 \text{ W m}^{-2}$  and an RMSE of about  $12\text{--}18 \text{ W m}^{-2}$  for the upwelling longwave fluxes in the ISLSCP data set. The LW net fluxes are estimated to be within  $10 \text{ W m}^{-2}$  of truth, with the largest potential errors in drier areas over snow covered surfaces.

[16] For the present paper, we utilize an SRB product that represents a combination of the solar and longwave fluxes at the surface: the net radiation at the surface,  $R_{\text{net}}$ . This flux is simply the sum of the net shortwave and net longwave fluxes. Throughout the text, we normalize the  $R_{\text{net}}$  values with the latent heat of vaporization,  $\lambda$ , taken here to be  $2.45 \times 10^6 \text{ J/kg}$  (a value typical of liquid-vapor phase changes at the surface of the Earth). As discussed in section 1, the normalization allows the net radiation to be expressed in terms of the amount of water it can evaporate. The latent heat of vaporization is about 15% larger for evaporation from snow (sublimation). For simplicity (and because the ISLSCP-2 precipitation data are not partitioned into rainfall and snowfall data), this difference is ignored in the present study. A simple test in which winter midlatitude  $R_{\text{net}}$  values were normalized by the higher latent heat of sublimation showed very little impact of this change on this study's results.

### 2.3. Runoff

[17] The gridded runoff data set included in the ISLSCP-2 archive is an update to the UNH-GRDC composite runoff fields of Fekete *et al.* [2002], a data set that combines observed river discharge from the Global River Data Center (GRDC) with simulated water balance model (WBM) estimates. (A runoff value for a given cell represents the runoff generated within the cell only and does not include runoff flowing into the cell from upstream.) In essence,



**Figure 2.** Locations (shaded) where the raw WBM runoff data have been “corrected” for the ISLSCP-2 data set through a scaling with observed stream gauge measurements, as archived by GRDC.

WBM estimates were “corrected” with the GRDC data through the application of scaling factors, factors that forced the gridded WBM annual estimates across gauged areas to sum up to the corresponding observed discharge values. The corrections were computed from annual totals and thus did not affect the seasonality of the WBM estimates.

[18] The original UNH-GRDC data set (<http://www.grdc.sr.unh.edu>) consists of monthly climatologies at  $0.5^\circ \times 0.5^\circ$  spatial resolution. To generate the fields for ISLSCP-2, the raw WBM monthly means were revised through the application of climate forcing (air temperature, precipitation, vapor pressure, solar radiation, wind speed) from the Climate Research Units (CRU) data set [New *et al.*, 1999, 2000]. In addition, a 10-year monthly time series of runoff for the 1986–1995 period was developed for ISLSCP-2.

[19] Of course, GRDC streamflow observations do not span the Earth’s land area. In nonmonitored regions, the runoff estimates are derived from WBM estimates alone. Furthermore, in monitored regions, GRDC corrections were not applied when GRDC and WBM data values appeared inconsistent. Potential reasons for inconsistencies between GRDC measurements and the raw WBM estimates are many. For example, although discharge is thought to be the most accurately measured component of the hydrological cycle, the uncertainty of discharge increments (the difference in discharge between adjacent stream gauges) can be large. (For reference, the discharge measurements at the gauges have an accuracy of 5–20% [Hageman and Dümenil, 1998; Rantz *et al.*, 1982].) Also, WBM runoff

estimates are subject to uncertainty in measured precipitation fields, and observed discharge may be influenced by human activities such as water withdrawal, interbasin transfer, and reservoir operation (though reservoir operation probably has a small impact on the long-term mean annual discharge). For both the climatology and the 10-year time series, the WBM estimates were scaled with GRDC data only if the scaling factor was in the 0.5–2.0 range. Regions with scaling factors outside this range were left uncorrected, reflecting (in part) our lack of knowledge regarding which data set is better.

[20] We restrict our analyses below to regions for which the GRDC scaling factors are applied, that is, to regions for which the WBM runoff estimates are indeed corrected with stream gauge measurements. These regions are illustrated in Figure 2. The corrected long-term averages in the shaded areas of Figure 2 were determined from at least five GRDC-corrected years of  $1^\circ \times 1^\circ$  data. (A  $1^\circ \times 1^\circ$  data value was considered “corrected” if all four contributing  $0.5^\circ \times 0.5^\circ$  quadrants were themselves corrected.) If a grid cell included years of data that were not scaled with GRDC measurements, those years were not used to compute the averages.

### 3. Regions and Time Period Considered

#### 3.1. GRDC-Corrected Basins

[21] The aim of the present paper is to examine independently measured precipitation, net radiation, and runoff values to evaluate their mutual consistency within the Budyko framework. Note that the runoff and precipitation

data in the unshaded land areas of Figure 2 are not truly independent of each other, since the runoff values are derived solely from the WBM analysis, an analysis that uses observed precipitation as an input. Only the runoff in the shaded areas of Figure 2 can be considered largely independent of the precipitation measurements, since the long-term runoff means there are scaled to GRDC stream gauge measurements (see section 2.3). Thus, for the present study, we utilize only data from the shaded areas in Figure 2.

[22] The hydrological regimes of the stream gauge monitored and nonmonitored areas (the shaded and unshaded land areas in Figure 2) are substantially different. For the original UNH-GRDC composite runoff data set, *Fekete et al.* [2002] show that roughly 60% of the nonmonitored landmass is dry (without any organized river network), implying that the remaining 40% of the nonmonitored land area must be substantially wetter than the monitored land area, since the monitored and nonmonitored areas for that data set are roughly equal in area and produce about the same discharge to the oceans. While the monitored area is smaller for the ISLSCP version of the data (due, for example, to differences in data period), the nonmonitored area in Figure 2 still shows a distinct relative prevalence of deserts.

### 3.2. Points With Hydrologically Inconsistent Data

[23] Data within a handful of ISLSCP-2 land grid cells, roughly 2% of all land grid cells (and roughly 3% of the grid cells within the GRDC-corrected area) show a clear hydrological inconsistency: The long-term average of the locally generated runoff,  $Q$ , in these cells exceeds the long-term average precipitation,  $P$ . These cells are marked with a dot in Figure 3a. The runoff  $Q$  does not exceed  $P$  by more than 10% in about a third of the marked cells, and it does not exceed  $P$  by more than 50% in about 5/6 of these cells.

[24] Areas failing the  $Q \leq P$  test include parts of the Amazon, the Pacific Northwest of the United States, and the Himalayas. In one sense, this is not surprising; rain gauge densities in the Amazon are low, and rainfall estimates are relatively uncertain in the Pacific Northwest and the Himalayas, where the topography is complex. (Indeed, *Adam et al.* [2006] offer an approach that uses the Budyko equation to account for the underestimation of precipitation over topographically complex regions.) We do not claim here, however, that poor rainfall estimates are necessarily responsible for the inconsistencies, because runoff estimates have their own sources of error (section 2.3). Unequivocally identifying which data set is most “at fault” in these regions is beyond the scope of this study. The relevant result is that researchers should be cautious about using the ISLSCP-2 precipitation and runoff data together in these regions. The points identified in the plot are excluded from the analyses in section 4 below.

### 3.3. Points With Hydrologically Suspicious Behavior

[25] Another 3% of all land grid cells (4% of the cells in the GRDC-corrected area) require special consideration because the water and energy variables at these cells violate a basic assumption of Budyko’s analysis: that evaporation not exceed the net radiation. We compute the long-term average evaporation,  $E$ , from the ISLSCP-2 data as  $E = P - Q$ . Figure 3b shows, with dots, the locations where the ISLSCP-2 data fail the  $E \leq R_{\text{net}}/\lambda$  test. This occurs, for

example, on the southern coast of Alaska and across parts of northern Europe. Note that if we had scaled the net radiation (see section 2.2) by the latent heat of sublimation rather than by the latent heat of vaporization in latitudes poleward of  $40^\circ$ , to account (rather overconservatively) for the presence of snow, the points failing the  $E \leq R_{\text{net}}/\lambda$  test would increase to 5% of all land grid cells (10% of the cells in the GRDC-corrected area). Again, though, the main results of the analysis below would remain unchanged.

[26] Note that having  $E$  exceed  $R_{\text{net}}/\lambda$  is not necessarily incorrect, since the energy needed to evaporate water could come from a negative sensible heat flux (effectively, from external warm air advected over an area) as well as from the net radiation. (Indeed, for northern Europe, *Milly and Dunne* [2002] faced the same problem and questioned the correctness of the  $E \leq R_{\text{net}}/\lambda$  assumption there.) Still, the violation of the  $E \leq R_{\text{net}}/\lambda$  condition does seem to point to ISLSCP-2 data inconsistencies in western tropical South America, for which sensible heat flux is presumably positive throughout the year. Given the inconsistency with the Budyko framework, these grid cells are not included in the analyses of section 4.

[27] A larger subset of the land grid cells, roughly 13% of all land points (20% in the GRDC Area), are identified as having another form of suspicious behavior: the standard deviation of yearly runoff,  $\sigma_Q$ , at these points exceeds that of the yearly precipitation,  $\sigma_P$ . These points are located with dots in Figure 3c. We must emphasize that a  $\sigma_Q \leq \sigma_P$  criterion is somewhat subjective, since it is not based on water balance considerations. Conceivably, for example, a region could receive the same total precipitation every year ( $\sigma_P = 0$ ) but have a nonzero runoff variability ( $\sigma_Q > 0$ ) because of interannual variations in the subyearly temporal distribution of the precipitation. In addition, with  $\sigma_Q$  and  $\sigma_P$  computed from a small number of annual totals, the  $\sigma_Q \leq \sigma_P$  criterion could be violated through sampling error alone. Still, given that annual precipitation is the main driver of annual runoff, so that interannual precipitation variability is a first-order driver of interannual runoff variability, and given that soil can act as a low-pass filter, so that higher frequencies of precipitation are partially filtered out in the translation of precipitation to runoff, the  $\sigma_Q \leq \sigma_P$  criterion is not unreasonable. The points indicated in Figure 3c can therefore be deemed suspicious: The yearly variations in the ISLSCP-2 precipitation and runoff data sets are likely to be inconsistent in these regions, and thus the annual means derived from these yearly values are also questionable. Some of the analyses in section 4 are performed with and without the  $\sigma_Q \leq \sigma_P$  criterion applied.

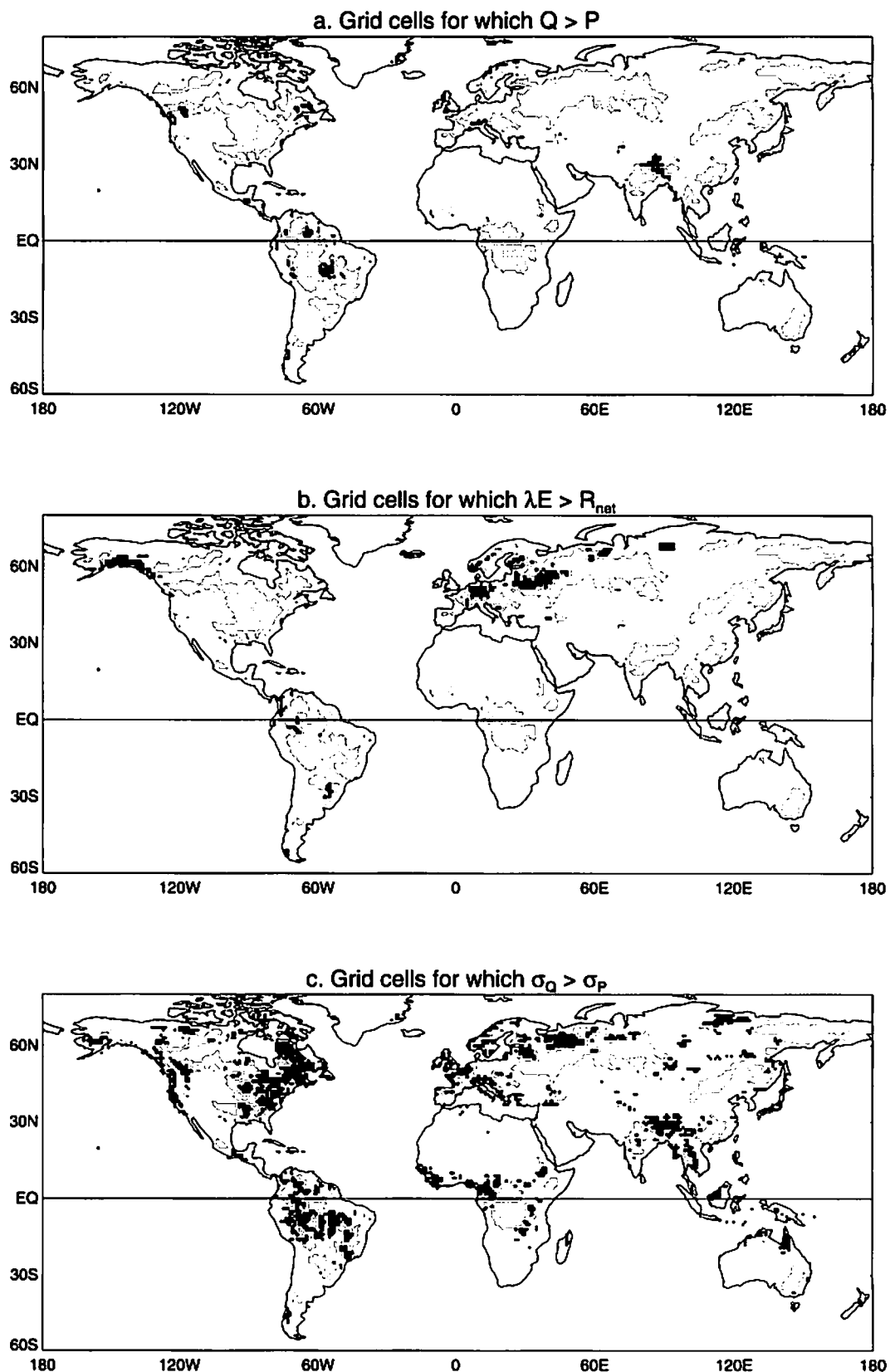
### 3.4. Time Period Considered

[28] The ISLSCP-2 data set nominally covers the period 1986–1995. The final year of the data set, however, included some missing data. For this reason, all calculations were limited to the time period 1986–1994.

## 4. Relevance of Budyko’s Relationships for Modern, Global Data

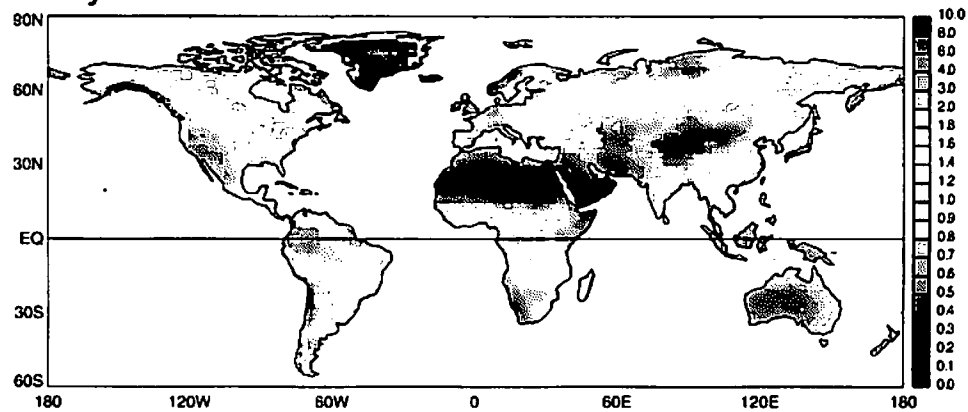
### 4.1. Global Fields of Dryness Index

[29] Figure 4a shows the global distribution of Budyko’s dryness index,  $D$ , as computed from ISLSCP-2 precipitation

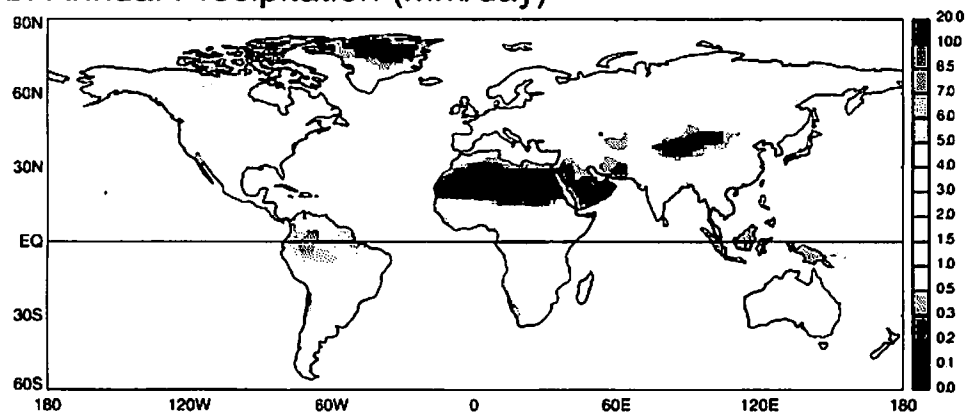


**Figure 3.** (a) Locations (dots) for which ISLSCP-2 annual runoff values exceed ISLSCP-2 annual precipitation values, implying hydrological inconsistency in the data. (b) Locations for which the annual evaporation rate (as derived from ISLSCP-2 annual precipitation and runoff rates) exceeds the amount allowed by the net radiation,  $R_{net}/\lambda$ . (c) Locations for which the standard deviation of annual runoff in the ISLSCP-2 data set exceeds that of annual precipitation in that data set. In all panels, the location of the “GRDC area” from Figure 2 is shaded.

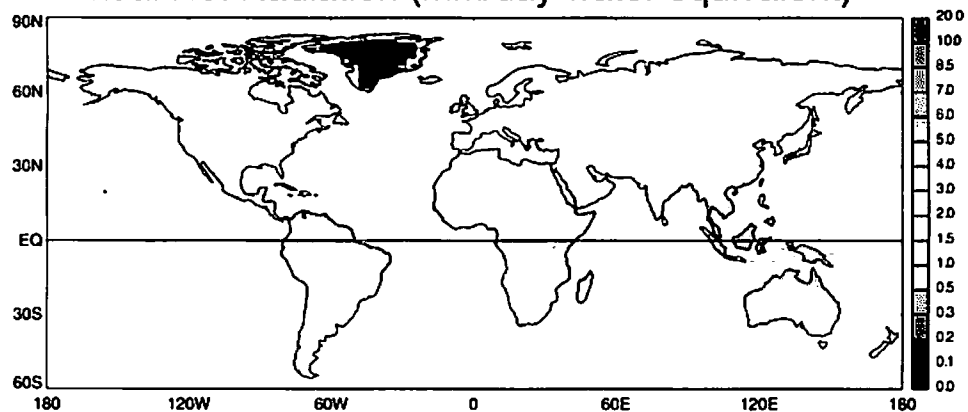
### a. Dryness Index



### b. Annual Precipitation (mm/day)



### c. Annual Net Radiation (mm/day water equivalent)

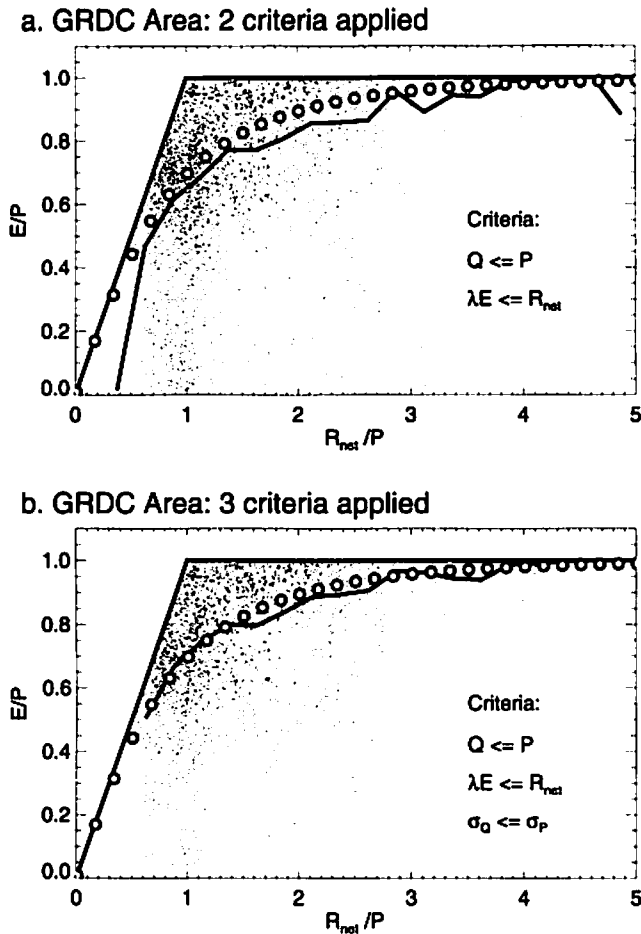


**Figure 4.** (a) Dryness index, as computed from the long-term annual means of precipitation and net radiation in the ISLSCP-2 data set. (b) Long-term annual mean of precipitation, from ISLSCP-2. (c) Long-term annual mean of net radiation (in terms of water equivalent), from ISLSCP-2. Color appears in back of the print issue.

and net radiation fields.  $D$  exceeds 1 over a majority of the Earth's land surface, implying (according to Budyko's treatment) that water availability rather than energy availability controls most land surface evaporation. Locations, however, for which energy availability is most important include the Amazon, Southeast Asia and Indonesia, and the northern edge of Europe and central Asia. Intermediate

regions, for which energy and water availability play similar roles, include the eastern half of North America, subtropical South America, tropical Africa, and much of Europe.

[30] Figure 4a provides a unique, observations-based global map of a first-order hydroclimatological control over evaporation (and, by extension, runoff), a map that should prove valuable for understanding the global water cycle and



**Figure 5.** (a) Scatterplot showing, for the ISLSCP-2 data, how the ratio of annual evaporation to precipitation varies with dryness index. Each point in the plot represents a  $1^\circ \times 1^\circ$  grid cell lying within the GRDC-corrected area in Figure 2 and satisfying the  $Q < P$  and  $E < R_{\text{net}}/\lambda$  criteria. The solid bin curve shows the average relationship between the data, and the open circles show the original Budyko relationship, from Figure 1. (b) Same as Figure 5a but removing from the analysis the grid cells for which  $\sigma_Q > \sigma_P$ .

for evaluating its simulation by climate models. The annual water balance is, after all, effectively controlled by precipitation and net radiation acting jointly, so that a look at either field alone would prove less instructive. *Koster et al.* [2000a, Figure 11] show a map of climate model-simulated dryness index that differs from Figure 4a in several areas, including over northeastern Asia, where the simulated values appear much too low. These simulated values are for a longer time period (several decades rather than the nine years considered here), so the differences may be due in part to decadal-scale climate fluctuations that would have a larger impact on the signal in the ISLSCP data. We note, however, that a supplemental analysis of GPCP data for the period 1979–2000 suggests that the ISLSCP period is not strongly anomalous: Outside of desert regions, the differences between rainfall in the ISLSCP period and that of the full 22-year period are generally below 5%. The differences are more likely explained by biases in the model. The model's simulation of evaporation, and even the ratio of evaporation to precipitation, in regions such as northeastern

Asia would likely be poor even if the land model employed were perfect. The hydroclimatology of any global model can be similarly evaluated.

[31] Figures 4b and 4c show the corresponding ISLSCP-2 mean annual precipitation and net radiation fields (the latter scaled by the latent heat of vaporization,  $\lambda$ , to produce water-equivalent units). As might be expected, the  $D$  field's patterns largely match those of the precipitation field, since net radiation has fewer spatial variations. The net radiation, however, does have a strong latitudinal gradient, a gradient that is folded into  $D$ .

#### 4.2. Reproduction of Budyko's Curve

[32] The two scatterplots in Figure 5 show the relationship between  $E/P$  and  $D$  under two different sets of criteria. In Figure 5a, the relationship is shown for GRDC-corrected data that satisfy the  $Q \leq P$  and  $\lambda E \leq R_{\text{net}}$  criteria discussed in section 3. As before,  $E$  is estimated as  $P - Q$ . Overlain on the scatterplot is a bin curve (the solid curve) showing the average relationship between the variables; the  $E/P$  values for all  $D$  values within a given range (bin width = 0.25) were averaged to compute the corresponding point in the bin curve. The open circles show the expected relationship from Budyko, computed with (1). Note that when Budyko developed his curve using observations, he also used a binning procedure to filter out scatter.

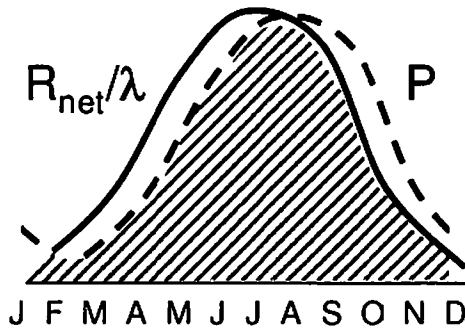
[33] The bin curve and the Budyko curve have the same basic shape, suggesting that Budyko's reasoning regarding rainfall and net radiation controls on annual evaporation, and their different levels of importance in different hydroclimatological regimes, is fully supported by the data. Note, though, that the bin curve falls slightly below the Budyko curve. This implies that either Budyko's simple semiempirical framework is slightly deficient or that the ISLSCP-2 precipitation, net radiation, and runoff data sets are slightly inconsistent. Curiously, *Milly and Dunne* [2002] found the opposite result for their larger-scale basins; in their analysis, the Budyko curve underestimated the inferred observational evaporation. They suggested the possibility that precipitation underestimates (due, for example, to gauge undercatch) may have hampered Budyko's ability to produce an accurate fit. The ISLSCP precipitation estimates, however, account for undercatch and should not show significant negative biases.

[34] Figure 5b shows the same analysis, but with some additional points excluded: the 20% of the points for which  $\sigma_Q > \sigma_P$  (see Figure 3c). As discussed in section 2, because year-to-year variability in precipitation is the chief driver of year-to-year variability in runoff, the condition  $\sigma_Q > \sigma_P$  while not rigorously incorrect, nevertheless suggests that the underlying data sets may be inconsistent. When the  $\sigma_Q < \sigma_P$  criterion is applied, i.e., when these suspicious points are excluded, the agreement between the Budyko and ISLSCP curves becomes stronger.

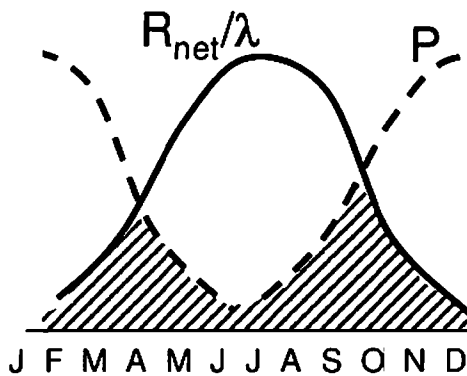
[35] Naturally, any conclusion regarding the agreement between the curves requires some qualification. Budyko did not have access to the wealth of data provided by ISLSCP-2, and thus sampling error alone on his part may be responsible for the differences in Figure 5a. The removal of additional ISLSCP-2 points in Figure 5b, many of which appear to be outliers, may be compensating for his sampling



### a. $P$ and $R_{net}$ mostly in phase



### b. $P$ and $R_{net}$ out of phase



**Figure 6.** Idealized schematic of two contrasting regions: (a) one for which the precipitation and net radiation cycles are mostly in phase and (b) the other for which they are mostly out of phase. The crosshatched area represents the integral over time of the minimum (at any given month) of the two cycles.

error. In other words, it is possible that the Earth's true underlying relationship between  $E/P$  and dryness index does not match the Budyko curve. (Perhaps it looks more like the ISLSCP-2 curve in Figure 5a.) Still, the agreement in Figure 5b is strong, and the criterion for excluding the additional points in Figure 5b is reasonable. Thus, with the above caveat, Figure 5b suggests that, according to present-day, state-of-the-art global observations, Budyko's semiempirical fit to water balance behavior in selected basins does hold true on the global scale.

#### 4.3. Phase Differences in the Seasonal Cycles of the Forcing

[36] The Budyko relationship provides the “average” value of  $E/P$  for a wide range of basin-scale regions with a given dryness index. It is not meant to give an accurate

value of the ratio at a single point or region. Figure 5 shows a tremendous amount of scatter about the Budyko function. Reasons for the scatter, for why two locations with the same dryness index partition rainfall in different ways, are many. For example, even with the same  $D$ , regions may differ in their vegetation structure, which affects the effective resistance applied to evaporation; in their topography or soil character, which affect their ability to generate runoff; and in the high-frequency structure of the precipitation, which also affects runoff.

[37] The ISLSCP-2 precipitation and radiation data afford a look at an additional, potentially important source for the scatter, namely, the relative phasing of the seasonal cycles of the precipitation and net radiation forcing. Consider the two sets of seasonal cycles in Figure 6. In Figure 6a, the  $P$  and  $R_{net}$  cycles are largely in phase, whereas in Figure 6b, the two cycles are out of phase. According to an overly strict interpretation of the Budyko relationship, the evaporation rate for the two cases would be the same, since the annual precipitation and net radiation, and thus the dryness index, for the two cases are exactly the same. Recall, however, that evaporation requires the presence of both water (through  $P$ ) and energy (through  $R_{net}$ ). The crosshatched area in each panel lies below the minimum of  $P$  and  $R_{net}$  at each month and thus represents the evaporation rate possible assuming no interseasonal storage of moisture or energy. The evaporation is larger (and thus runoff is smaller) when the two seasonal cycles are more in phase, despite the equivalent dryness index.

[38] Of course, this argument is flawed because the interseasonal storage of moisture (through snowpack and soil moisture reservoirs) is known to be very important [e.g., Milly and Dunne, 1994]. Still, the relative phasing of the cycles may have some impact on runoff, and we can search for this impact in the ISLSCP data. Table 1 shows the degree to which phasing affects the observed runoff within narrow ranges of the dryness index. The phasing diagnostic used here, computed separately at each grid cell, is the area of the crosshatched region in Figure 6 divided by whichever is smaller: the total annual precipitation or the total annual net radiation. It thus varies from 0 to 1, with lower values implying that the precipitation and net radiation cycles are more out of phase. The runoff diagnostic is simply the ratio of observed annual runoff to observed annual precipitation. The second column of the table shows the correlation ( $r$ ) between the phasing diagnostic and the runoff diagnostic. The third column shows the number of points ( $N$ ) that went into the calculation (i.e., lying within the dryness index range) and the probability, in percent, that the underlying true correlation is not zero (based on Monte Carlo analysis, in which 30,000 pairs of independent, randomly generated data sets of length  $N$  are compared to each other). The final column shows the mean of the phasing diagnostic, which hovers around 0.8 for the different bins. (Note that the higher mean values for the lower bins are consistent with the construct of the diagnostic; when either  $P \gg R_{net}/\lambda$  or  $R_{net}/\lambda \gg P$ , the phasing diagnostic tends to go to 1.) The analysis is limited to points falling within GRDC basins and to points between  $40^\circ\text{N}$  and  $40^\circ\text{S}$ , to avoid known storage issues associated with snow in higher latitudes. Only ranges containing at least 50 points are included in the table. To maximize the number of points considered, the analysis

**Table 1.** Correlations Between Phase Index and Runoff Index for Various Dryness Index Intervals<sup>a</sup>

Dryness Index Range	Correlation Between Phase Index and Runoff Index	Number of Points (and Statistical Significance)	Average of Phase Index
0.7 < D < 0.8	-0.077	127 (80%)	0.87
0.8 < D < 0.9	-0.706	136 (99.99+%)	0.84
0.9 < D < 1.0	-0.285	161 (99.98%)	0.80
1.0 < D < 1.1	-0.563	220 (99.99+%)	0.80
1.1 < D < 1.2	-0.474	144 (99.99+%)	0.81
1.2 < D < 1.3	-0.261	82 (99.14%)	0.81
1.3 < D < 1.4	-0.241	66 (97.4%)	0.82

<sup>a</sup>See text for details.

applied the  $Q < P$  and  $E < R_{\text{net}}$  criteria but did not apply the  $\sigma_Q < \sigma_P$  criterion.

[39] As expected, the correlations between the phasing diagnostic and the runoff ratio are negative, with a large degree of statistical significance. In other words, despite their disparate sources, the observational data do provide evidence that annual runoff is affected by the relative phasing of the cycles of water supply and energy supply at the land surface. Note, however, that the correlations in Table 1, though significant, are not large. There are two likely reasons for this. First, the underlying correlations in nature may be small; the other factors contributing to the scatter, as noted above, may be dominant, and in any case the interseasonal storage of water, even in snow-free regions, may ameliorate the effect of the phase differences. Second, of course, the ISLSCP-2 data sets may be inconsistent in places.

[40] For perspective, we can address the phasing question with a global modeling study, one that uses parameterized land surface physics to generate runoff and evaporation fields in response to observed precipitation and net radiation forcing. Such models, of course, are notorious for their biases and their suboptimal parameterizations, but they do provide consistent estimates of  $P$ ,  $R_{\text{net}}$ ,  $E$ , and  $Q$ . As part of GSWP-2 (the Global Soil Wetness Project, Phase 2 [see Dirmeyer *et al.*, 2005]), the catchment model [Koster *et al.*, 2000b] of the NASA Goddard Space Flight Center's Global Modeling and Assimilation Office was driven globally offline with realistic meteorological forcing. The correlations between the phasing diagnostic and the simulated runoff ratio were found to be similar to those in Table 1. The GSWP-2 model result suggests that the ISLSCP-2 observations are consistent enough to capture most of the impact of phasing on the water budget, small as it might be.

#### 4.4. Interannual Variability

[41] To examine the consistency between the year-to-year variations in the ISLSCP-2 precipitation and runoff data, we could compute correlations between monthly or annual values. This approach, however, cannot account for the fact that runoff tends to lag precipitation. Precipitation falling at the end of one calendar year may run off at the beginning of the next, sometimes throwing off the correlation significantly. Attempts to correct for this would require a subjective grid cell by grid cell analysis, which is beyond the scope of this paper.

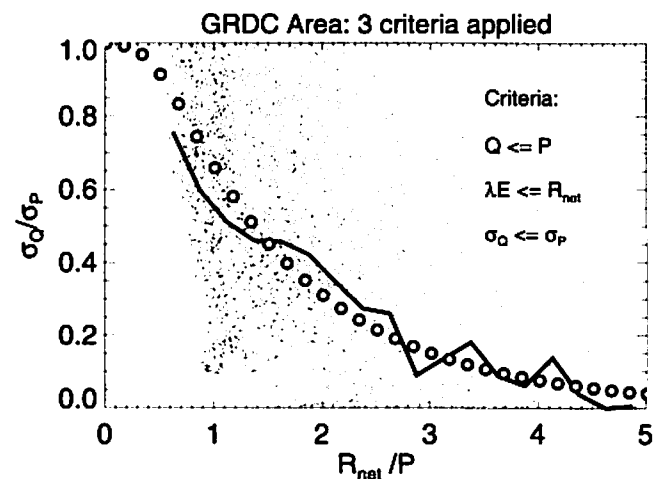
[42] The approach of Koster and Suarez [1999], however, is suitable for the present study, since it can be reformulated to rely strictly on the standard deviations of rainfall and

runoff ( $\sigma_P$  and  $\sigma_Q$ , respectively) without consideration of their correlation. A reformulation of (2), the equation used by Koster and Suarez [1999], is necessary because estimates of  $\sigma_E$  require yearly estimates of evaporation, estimates that are prevented by the lagged behavior of the runoff. Equation (2) is thus rewritten as:

$$\sigma_Q/\sigma_P = 1 - \mathcal{F}(D) + D\mathcal{F}'(D) \quad (3)$$

[43] Underlying (3) is the idea that a given year's total precipitation and net radiation are enough to determine the partitioning of that year's precipitation into runoff and evaporation. As a result, the framework is inconsistent with the idea that  $\sigma_Q$  can exceed  $\sigma_P$ ; indeed, all three criteria used in Figure 5b ( $Q \leq P$ ,  $\lambda E \leq R_{\text{net}}$ ,  $\sigma_Q \leq \sigma_P$ ) must be applied to the data before evaluating their consistency with (3). Figure 7 shows the scatterplot of  $\sigma_Q/\sigma_P$  versus  $D$  for the ISLSCP-2 data in the GRDC area, applying the three criteria. The bin curve representing the average relationship is also shown, as is the curve representing (3).

[44] Once again, the agreement, while not perfect, is strong. Our ability to examine the consistency of year-to-



**Figure 7.** Scatterplot showing, for the ISLSCP-2 data, how the ratio of the standard deviation of annual runoff to that of annual precipitation varies with dryness index. Each point in the plot represents a  $1^\circ \times 1^\circ$  grid cell lying within the GRDC-corrected area in Figure 2 and satisfying the  $Q < P$ ,  $E < R_{\text{net}}/\lambda$ , and  $\sigma_Q < \sigma_P$  criteria. The solid bin curve shows the average relationship between the data, and the open circles show the original Budyko-based relationship, from Figure 1.

year variations in the ISLSCP-2 data sets is very limited, but to the extent that we can test it, and outside of the indicated areas for which  $\sigma_Q > \sigma_P$ , the data sets appear generally consistent in the context of the Budyko framework.

## 5. Summary

[45] The ISLSCP-2 data set provides coregistered global fields of three fundamental elements of the global hydrological cycle: precipitation ( $P$ ), net radiation ( $R_{net}$ ), and runoff ( $Q$ ). From the mean annual  $P$  and  $R_{net}$  fields, we can compute an observations-based global distribution (Figure 4a) of Budyko's dryness index ( $D$ ), known to be a first-order control on global evaporation and runoff fields. This  $D$  field by itself should prove invaluable to global modelers wishing to evaluate strengths and deficiencies in their simulation of the global hydrological cycle.

[46] The  $P$ ,  $R_{net}$ , and  $Q$  fields in the ISLSCP data set, however, come from independent measurement systems, particularly the portion of the  $Q$  fields lying within GRDC-corrected basins (Figure 2). Thus we might expect to see some inconsistency between the fields. In some grid cells, runoff exceeds precipitation, a clear violation of the long-term water balance (Figure 3a). Inferred evaporation rates exceed net radiation rates in a number of grid cells (Figure 3b), and while this does not strictly violate the long-term energy budget, since external, advected energy can be transferred to the land surface through a downward sensible heat flux, the data sets' consistency in the tropical and subtropical subset of these regions is strongly suspect. Consistency is also suspect at those locations for which the standard deviation of runoff exceeds that of precipitation (Figure 3c), since annual rainfall variability is the chief driver of annual runoff variability.

[47] Focusing on the GRDC-corrected regions in Figure 2, and with all suspicious points removed from the analysis, the observations reproduce quite well Budyko's semiempirical relationship between dryness index and the ratio of annual evaporation to annual precipitation (Figure 5b). Thus, in terms of the Budyko relationship, the precipitation, net radiation, and runoff data in the ISLSCP-2 data set, though obtained from different sources, are generally consistent.

[48] The ISLSCP-2 data also allow a rare global look at a potentially important control on annual runoff, namely, the relative phasing of the mean seasonal cycles of precipitation and net radiation. The data indeed show (Table 1) that when the cycles are more out of phase, annual runoff tends to increase. The interannual variations of the ISLSCP-2 precipitation and runoff data were examined using the Budyko-based framework of Koster and Suarez [1999]. The results (Figure 7) show that the disparate data sets are also generally consistent in this context.

[49] Of course, the Budyko curves in Figure 1 are semiempirical and thus do not stem strictly from theory. Also, the scatter around the curves is substantial, diminishing the relevance of the relationships at any individual point or region. Thus it is incorrect to conclude that the agreement with the Budyko curves in Figures 5b and 7 quantitatively validates the ISLSCP-2 fields. Rather, the agreement shows that together, the ISLSCP-2 fields successfully capture the basic (and reasonable) hydroclimatological constraints on

evaporation outlined by Budyko: the fact that evaporation is controlled by precipitation in drier climates, by net radiation in wetter climates, and by a more complex set of controls (including the relative phasing of  $P$  and  $R_{net}$ ) in intermediate climates. In short, outside of the identified inconsistent and suspicious points, the independently derived ISLSCP-2 fields appear hydroclimatologically consistent. Conversely, the agreement shows that the decades-old Budyko framework holds up very well when tested with the most up-to-date, spatially complete and highly resolved global observations available. To the extent that the ISLSCP-2 data are valid, the Budyko framework is seen to remain a useful approach for characterizing first-order controls on the global hydrological cycle.

[50] **Acknowledgments.** The authors thank Ping Liu for help with the processing of the ISLSCP data. Forrest Hall and Eric Brown de Colstoun are thanked for their leadership in producing the ISLSCP-2 data set.

## References

- Adam, J. C., E. A. Clark, D. P. Lettenmaier, and E. F. Wood (2006), Correction of global precipitation products for orographic effects, *J. Clim.*, **19**, 15–38.
- Adler, R. F., G. J. Huffman, and P. R. Keehn (1994), Global tropical rain estimates from microwave-adjusted geosynchronous IR data, *Remote Sens. Rev.*, **11**, 125–152.
- Adler, R. F., et al. (2003), The version 2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present), *J. Hydrometeorol.*, **4**, 1147–1167.
- Budyko, M. I. (1958), *The Heat Balance of the Earth's Surface*, translated by N. A. Stepanova, 259 pp., U.S. Dep. of Commer., Washington, D. C.
- Budyko, M. I. (1974), *Climate and Life*, 508 pp., Elsevier, New York.
- Charlock, T. P., and T. L. Alberta (1996), The CERES/ARM/GEWEX Experiment (CAGEX) for the retrieval of radiative fluxes with satellite data, *Bull. Am. Meteorol. Soc.*, **77**, 2673–2683.
- Cox, S. J., P. W. Stackhouse Jr., S. K. Gupta, J. C. Mikovitz, M. Chiacchio, and T. Zhang (2004), The NASA/GEWEX Surface Radiation Budget Project, results and analysis, paper presented at International Radiation Symposium, Int. Radiat. Comm., Busan, South Korea, 23–27 Aug.
- Dirmeyer, P. A., X. Gao, M. Zhao, Z. Guo, T. Oki, and N. Hanasaki (2005), The second Global Soil Wetness Project (GSWP-2): Multi-model analysis and implications for our perception of the land surface, *COLA Tech. Rep. 185*, 46 pp., Cent. for Ocean-Land-Atmos. Stud., Calverton, Md. (Available at [http://grads.iges.org/pub/ctr/ctr\\_185.pdf](http://grads.iges.org/pub/ctr/ctr_185.pdf))
- 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.
- Gilgen, H., and A. Ohmura (1999), The global energy balance archive, *Bull. Am. Meteorol. Soc.*, **80**(5), 831–850.
- Gupta, S. K., N. A. Ritchey, A. C. Wilber, C. H. Whitlock, G. G. Gibson, and P. W. Stackhouse Jr. (1999), A climatology of surface radiation budget derived from satellite data, *J. Clim.*, **12**, 2691–2710.
- Hagemann, S., and L. Dümenil (1998), A parameterization of the lateral waterflow for the global scale, *Clim. Dyn.*, **14**, 17–31.
- Hall, F. G., E. B. de Colstoun, G. J. Collatz, D. Landis, P. Dirmeyer, A. Betts, G. J. Huffman, L. Bounoua, M. Bosilovich, and B. Meeson (2006), The ISLSCP Initiative II global data sets: Surface boundary conditions and atmospheric forcings for land-atmosphere studies, *J. Geophys. Res.*, doi:10.1029/2006JD007366, in press.
- Koster, R. D., and M. J. Suarez (1999), A simple framework for examining the interannual variability of land surface moisture fluxes, *J. Clim.*, **12**, 1911–1917.
- Koster, R. D., M. J. Suarez, and M. Heiser (2000a), Variance and predictability of precipitation at seasonal-to-interannual timescales, *J. Hydrometeorol.*, **1**, 26–46.
- Koster, R. D., M. J. Suarez, A. Ducharme, M. Stieglitz, and P. Kumar (2000b), A catchment-based approach to modeling land surface processes in a general circulation model: I. Model structure, *J. Geophys. Res.*, **105**, 24,809–24,822.
- Koster, R. D., P. A. Dirmeyer, P. C. D. Milly, and G. L. Russell (2001), Comparing GCM-generated land surface water budgets using a simple common framework, in *Land Surface Hydrology, Meteorology, and Climate: Observations and Modeling, Water Sci. and Appl. Ser.*, vol. 3, edited by V. Lakshmi, pp. 95–105, AGU, Washington, D. C.

- Milly, P. C. D., and K. A. Dunne (1994), Sensitivity of the global water cycle to the water-holding capacity of land, *J. Clim.*, **7**, 506–526.
- Milly, P. C. D., and K. A. Dunne (2002), Macroscale water fluxes: 2. Water and energy supply control of their interannual variability, *Water Resour. Res.*, **38**(10), 1206, doi:10.1029/2001WR000760.
- New, M., M. Hulme, and P. Jones (1999), Representing twentieth century space-time climate variability. Part I: Development of a 1961–1990 mean monthly terrestrial climatology, *J. Clim.*, **12**, 829–856.
- New, M., M. Hulme, and P. Jones (2000), Representing twentieth century space-time climate variability. Part II: Development of 1901–1996 monthly grids of terrestrial surface climate, *J. Clim.*, **13**, 2217–2238.
- Ohmura, A., et al. (1998), Baseline Surface Radiation Network (BSRN/WCRP): New precision radiometry for climate research, *Bull. Am. Meteorol. Soc.*, **79**(10), 2115–2136.
- Pinker, R., and J. A. Ewing (1985), Modeling of surface solar radiation: Model formulation and validation, *J. Clim. Appl. Meteorol.*, **24**, 389–401.
- Rantz, S. E., et al. (1982), Measurement and computation of streamflow, vol. 2, Computation of discharge, *U. S. Geol. Surv. Water Supply Pap.* **2175**, 347 pp.
- Rossow, W. B., and R. Schiffer (1999), Advances in understanding clouds from ISCCP, *Bull. Am. Meteorol. Soc.*, **80**, 2261–2287.
- Rossow, W. B., A. W. Walker, D. E. Beuscher, and M. D. Roiter (1996), International Satellite Cloud Climatology Project (ISCCP) documentation of new cloud datasets, *WMO/TD 737*, 115 pp., World Clim. Res. Programme, Geneva, Switzerland.
- Schubert, S., C.-K. Park, C.-Y. Wu, W. Higgins, Y. Kondratyeva, A. Molod, L. Takacs, M. Seaborn, and R. Rood (1995), A multi-year assimilation with the GEOS-I system: Overview and results, *NASA Tech. Memo.* **104606**, vol. 6.
- Stackhouse, P. W., Jr., S. K. Gupta, S. J. Cox, J. C. Mikovitz, T. Zhang, and M. Chiacchio (2004), 12-year surface radiation budget data set, *GEWEX News*, **14**(4), 10–12.
- Wilber, A. C., D. P. Kratz, and S. K. Gupta (1999), Surface emissivity maps for use of satellite retrievals of longwave radiation, *NASA Tech. Publ.*, **TP-99-209362**, 35 pp.
- 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.
- Zhang, Y., W. B. Rossow, and P. W. Stackhouse Jr. (2006), Comparison of different global information sources used in surface radiative flux calculation: Radiative properties of the near-surface atmosphere, *J. Geophys. Res.*, **111**, D13106, doi:10.1029/2005JD006873.

---

B. M. Fekete, Water Systems Analysis Group, Complex Systems Research Center, Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, Morse Hall, Durham, NH 03824, USA.

G. J. Huffman, Science Systems and Applications, Inc., Code 613, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA.

R. D. Koster, Global Modeling and Assimilation Office, Code 610.1, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA. (randal.d.koster@nasa.gov)

P. W. Stackhouse Jr., NASA Langley Research Center, 21 Langley Boulevard, M.S. 420, Hampton, VA 23681, USA.



## Evaluation of ISLSCP Initiative II satellite-based land cover data sets and assessment of progress in land cover data for global modeling

Eric C. Brown de Colstoun,<sup>1</sup> Ruth S. DeFries,<sup>2</sup> and John R. G. Townshend<sup>2</sup>

Received 28 April 2006; revised 31 August 2006; accepted 10 October 2006; published 15 November 2006.

[1] As an important component of the International Satellite Land Surface Climatology Project (ISLSCP) Initiative II data collection, eight state-of-the-art land cover/use data sets have been compiled and made consistent with the ISLSCP Initiative II land/water mask in support of global modeling efforts. These data sets contain new and improved global data sets at coarse resolutions (1/4, 1/2 and 1°) describing historical, recent and present land cover conditions and are a testament to the tremendous progress made in this area over the past decade. In addition to the historical data, data describing the subcell heterogeneity in land cover are also provided, both in terms of subcell proportions of land cover classes and vegetation continuous fields such as % tree, grass and bare cover. Here we present the various ISLSCP II land cover data sets and compare the principal satellite-derived data sets and the effect of their respective aggregation methods. We find that despite some notable disagreements among similar classes, the satellite-based data sets agree remarkably well over large portions of the Earth's surface (over 50% for all resolutions). We also find that the methods of aggregation, whether done by a strictly dominant type, or using more information on subcell tree cover, can have an important impact on the final output and need to be considered by the user. Finally, by integrating the vegetation continuous fields data into our analyses we are able to show that the principal differences in terms of discrete land cover classes are in fact transition zones between similar classes.

**Citation:** 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.*, **111**, D22S07, doi:10.1029/2006JD007453.

### 1. Introduction

[2] Contemporary, accurate, and consistently repeatable global land cover characterizations such as land cover/use classifications and/or vegetation continuous fields (e.g., % tree cover) play an important role in various aspects of global change studies. Present land cover conditions are needed to generate land cover dependent biophysical parameter fields used in many current General Circulation Models (GCMs) and Numerical Weather Prediction Models. These models can simulate atmospheric circulation and climatic variables such as temperature, rainfall, humidity and wind at a fairly coarse spatial scale and under various global warming scenarios [Dickinson *et al.*, 1986; Sellers *et al.*, 1996a]. The inclusion of land into these computerized models has progressed greatly since the first description of land as a “bucket” by Manabe [1969]. Most current models are now coupled with Land Surface Parameterization (LSPs) models which have depended on digital compila-

tions of global land cover such as those of Olson *et al.* [1983], Matthews [1983], Wilson and Henderson-Sellers [1985], or more recent satellite-derived land cover maps such as those of DeFries and Townshend [1994a], Loveland and Belward [1997], Hansen *et al.* [2000], Friedl *et al.* [2002], and Bartholomé and Belward [2005], among others. The LSPs come from a realization that vegetation and soils play an important role, both in space and time, in regulating the exchange of energy, gases and water vapor between the biosphere and the atmosphere and, as such, should be included in global simulations [Charney *et al.*, 1975; Dickinson, 1983]. The LSPs serve to produce databases or look-up tables of land cover dependent albedo, surface roughness, and evapotranspiration and respiration, parameters that control, respectively, the transfer of energy, momentum, mass, and latent and sensible heat between the biosphere and the lower layers of the atmosphere [Dorman and Sellers, 1989; Bonan, 2002]. These digital land cover maps also provide the means by which to include the fine-scale heterogeneity of land processes within the coarser grid of the GCMs.

[3] Land cover information is also an important input to biogeochemical, ecosystem, and hydrological models which track the cycling of carbon, nutrients, energy and water between the biosphere and the atmosphere [Melillo *et al.*, 1993; Running and Hunt, 1993; Melillo, 1994; Nemani and

<sup>1</sup>Science Systems and Applications, Inc., Biospheric Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.

<sup>2</sup>Department of Geography, University of Maryland, College Park, Maryland, USA.

*Running*, 1996]. Some of these models can simulate the response of terrestrial ecosystems to elevated CO<sub>2</sub> concentrations and/or climate change, for example. By quantifying the gross and Net Primary Production (NPP) of these ecosystems they can help to identify the geographical location of the principal sources and sinks of carbon, and their temporal and spatial variability, as well as providing improved estimates of the size of various global carbon pools [e.g., *Kicklighter et al.*, 1999]. Vegetation type information is important to these models because various plant and tree species have varied mechanisms for photosynthesis and carbon assimilation which can be affected by different stresses, all factors which can in turn significantly alter estimates obtained from the models [*Bonan*, 1995; *Bondeau et al.*, 1999].

[4] Finally, there is still disagreement between estimates of the land cover conversion which has occurred in the past, and is occurring now, but also on the *rates* of change in land cover conversion [e.g., *Skole and Tucker*, 1993] and the impact of such changes on the global carbon cycle [*Plattner et al.*, 2002; *Houghton*, 2003]. Global land cover characterizations can simplify the monitoring of natural or human-induced changes of land cover/use and are important in modeling the consequences of these changes on local and global processes [e.g., *Bonan*, 1997; *Bounoua et al.*, 2002]. Clearly, if land cover can be accurately measured and consistently monitored globally for a period of several years, significant changes over time, and the rates of these changes, could be quantitatively evaluated and some of the above uncertainties potentially reduced. In fact, recent global land cover classifications have been used as a baseline from which land cover change models can be applied to determine the historical land cover change and rates of change, as demonstrated by *Ramankutty and Foley* [1999] for croplands.

[5] While it remains a highly desirable goal to parameterize global models directly from the remotely sensed observations, the use of static or semistatic land cover data sets for model parameterization is still anticipated for the near future. In alignment with the goals of the International Satellite Land Surface Climatology Project (ISLSCP) Initiative II Data Collection, a collection of state-of-the-art land cover data sets has been compiled in collaboration with the data set producers, reprocessed to common spatial resolutions of 1/4, 1/2 and 1° in both latitude and longitude, with common land/ocean boundaries, to support global modeling efforts. The purpose of this paper is to present these ISLSCP Initiative II land cover data sets, compare and contrast several of the satellite-derived data sets, discuss issues of consistency between the various data sets, and provide an assessment of the progress in these data since the publication of the first coarse resolution satellite-based land cover data sets as a part of the first ISLSCP data collection [*Sellers et al.*, 1996b].

## 2. Background

[6] The single land cover data set provided in the ISLSCP Initiative I data collection was a 1° by 1° land cover map based on satellite data from the Advanced Very High Resolution Radiometer (AVHRR) [*DeFries and Townshend*, 1994a] and was the first such data set to be generated from

remotely sensed data at a global scale. The impetus for the production of that data set was a comparison carried out by *DeFries and Townshend* [1994b] of the three most widely available digital global land cover classifications at the time, those of *Olson et al.* [1983], *Matthews* [1983], and *Wilson and Henderson-Sellers* [1985]. *DeFries and Townshend* [1994b] found that only 26% of the total land area was classified as the same land cover type in all three maps and also noted large discrepancies in terms of the spatial distribution of different major land cover types as well as their actual areal extent over the globe. Because satellite remote sensing provides a synoptic view of the Earth and is able to perform consistent and repetitive quantitative measurements of many terrestrial processes at a variety of spatial scales, they argued that remotely sensed data sets could potentially provide the means by which to generate more consistent and accurate global-scale land cover data sets.

### 2.1. Global Land Cover Maps Compiled From Ground-Based Sources

[7] Historically, land cover classifications have been performed from ground surveys and/or previous maps and the mapping or delineation of land cover types has typically been made by reference to climate, physiognomic characteristics, floristic composition, or geographical location [*Mueller-Dombois*, 1984; *Prentice*, 1990]. Several important points can be made about these classifications. First, they are subjective in that they reflect the biases of the compilers and the variety of sources they depend on. Second, they offer only qualitative information that is not very useful for input to computerized models of global change.

[8] Several digital maps of global vegetation [e.g., *Olson et al.*, 1983; *Matthews*, 1983; *Wilson and Henderson-Sellers*, 1985] have been compiled from a variety of ground-based sources such as paper maps and atlases, and limited satellite data. While the above databases have been used extensively to support climate change studies, they also are influenced by the decisions and choices of the compilers as well as the quality of their sources. As previously noted, these have disagreed both in terms of the land cover present as well as the areal extent of particular biomes [*Matthews*, 1983; *DeFries and Townshend*, 1994b] but, in all fairness, the differences may also reflect the different purposes of each database. Another difficulty in comparing these data is that, because of the varied methods, classification schemes, and age of sources used, it is not always entirely clear whether the maps reflect the potential or actual vegetation cover, except in the case of bioclimatic classifications. Finally, because they have relied on ground-based sources, any updates or changes have been difficult to implement.

### 2.2. Global Land Cover Characterizations From Remotely Sensed Data

[9] Satellite remote sensing has been, and is currently being, explored as an attractive alternative for *actual* continental to global-scale land cover classifications [*Tucker et al.*, 1985; *Townshend et al.*, 1987; *Loveland et al.*, 1991; *DeFries and Townshend*, 1994a; *Running et al.*, 1995; *Loveland and Belward*, 1997; *DeFries et al.*, 1995, 1998; *Hansen et al.*, 2000; *Friedl et al.*, 2002; *Bartholomé and Belward*, 2005]. These studies used remotely sensed spectral data acquired from instruments such as the AVHRR, the MODerate

Resolution Imaging Spectroradiometer (MODIS) or the Système Probatoire Pour l'Observation de la Terre (SPOT4)-VEGETATION, coupled with their temporal evolution, to separate land cover classes at the continental and/or global scales. These classifications have typically relied on the variability as a function of cover type of the Normalized Difference Vegetation Index (NDVI). This index, defined as the difference of the solar energy reflected from surfaces in the near-infrared and red portions of the electromagnetic spectrum divided by their sum, is recognized as a broad indicator of surface "greenness," photosynthetic activity, and canopy phenology [Asrar et al., 1984; Justice et al., 1985; Daughtry et al., 1992].

[10] The approach of Loveland et al. [1991], Loveland and Belward [1997], and Bartholomé and Belward [2005] is essentially based on utilizing 12 months of NDVI data with an unsupervised classification algorithm. A large database of ancillary information is used as an aid for the human interpretation of the results. Because of the laborious nature of the postprocessing of these unsupervised classification data sets, they have been produced irregularly and are difficult to implement completely objectively and repeatedly for data sets of multiple years. Other techniques, such as those of DeFries et al. [1995, 1998], and Hansen et al. [2000], are supervised classification approaches which rely on a data set of carefully screened global training data derived from Landsat data; they also used the NDVI, but in conjunction with information from the individual spectral bands of AVHRR, including those in the thermal wavelength region, to improve the efficacy of remotely sensed global land cover classifications. The current MODIS land cover algorithm [Friedl et al., 2002] follows the heritage of supervised classification from AVHRR but its inputs are 16-day composites for the individual MODIS land bands and the Enhanced Vegetation Index (EVI) for an entire year. The current land cover algorithm for the Visible/Infrared Imager/Radiometer Suite (VIIRS) scheduled to fly on the future National Polar Orbiting Environmental Satellite System (NPOESS) [Brown de Colstoun et al., 2000] follows from the AVHRR and MODIS heritage as a supervised classification and also uses a decision tree classifier but is closer to the approach used by Hansen et al. [2000] in terms of data inputs. Finally, the same approach used by DeFries et al. [1995, 1998] and Hansen et al. [2000] to generate global land cover products has also been used with linear mixture models and regression tree algorithms to generate global fields of continuous vegetation characteristics such as tree, herbaceous and bare cover [DeFries et al., 1999; Hansen et al., 2002]. These products are found to more closely represent natural gradients and ecotones in vegetation characteristics, as opposed to the classification of cover types into discrete values and in fact may potentially be more useful to global modelers than stratifications by land cover because they scale linearly to coarser resolutions.

[11] The production of these global land cover data sets would simply not have been possible without the production of the input data necessary for the classifications as well as the production of global training data sets from Landsat data [DeFries et al., 1998]. The first data sets of DeFries et al. [1995, 1998] were produced using the Pathfinder AVHRR Land (PAL) data sets at 8 km spatial resolution [James and Kalluri, 1994]. In parallel, and under the auspices of the

Data and Information System of the International Geosphere Biosphere Programme (IGBP-DIS), a global 1 km data set from AVHRR data was produced spanning the years 1992–1996 to address the needs of several of the IGBP's programs [Townshend et al., 1994]. This 1 km data set [Eidenshink and Faundeen, 1994] forms the core input data for several land cover data sets provided in this ISLSCP Initiative II data collection (Table 1), including the University of Maryland (UMD) land cover data set [Hansen et al., 2000], the IGBP-DIScover vegetation classification [Loveland and Belward, 1997], and the UMD continuous fields of vegetation cover [DeFries et al., 2000]. A recent (2000–2001) MODIS land cover product from MODIS collection 4 [Friedl et al., 2002] has also been added to the ISLSCP Initiative II land cover "suite" to provide a linkage to future data sets that will become available with MODIS and VIIRS. We note that a recent global land cover data set based on SPOT-VGT data named GLC-2000 has been produced under the coordination of the European Commission's Joint Research Centre [Bartholomé and Belward, 2005] but was not available within the time constraints for publication in the ISLSCP Initiative II collection.

### 2.3. ISLSCP Initiative II Land Cover Data Sets

[12] While the ISLSCP Initiative I collection contained a single global land cover data set, the Initiative II collection now contains 8 different state-of-the-art data sets dealing with various aspects of land cover and/or land use (Table 1), including two historical land cover data sets: the historical croplands fractional cover data set of Ramankutty and Foley [1999], covering the period from 1700–1992, and a related historical land cover and land use (1700–1990) data set from the National Institute of Public Health and the Environment (RIVM) in the Netherlands [Klein Goldewijk, 2001]. Ramankutty and Foley [1998] derived a spatially explicit data set of croplands for the year 1992 by synthesizing remotely sensed land cover data (IGBP-DIScover data set in Table 1) with contemporary land inventory data. Furthermore, Ramankutty and Foley [1999] extended this data set back to 1700 using historical land inventory data. By extending their data set back in time, they were also able to produce a land cover map of "potential" vegetation, or the natural vegetation before human alteration or other types of disturbance, which is also included in this collection. Klein Goldewijk [2001] used historical statistical inventories on agricultural land (census data, tax records, land surveys, etc.) and different spatial analysis techniques to create a geographically explicit data set of land use change, with a regular time interval (see Table 1). These two new global data sets of historical land cover change compare fairly well over most of the Earth despite the different modeling approaches and input data used [Klein Goldewijk and Ramankutty, 2004].

[13] Another interesting addition to the ISLSCP II collection is the data set of Still et al. [2003] which identifies the fraction of each cell with a C<sub>4</sub> dominant photosynthetic pathway. This data set was actually produced from various data sets which are also included in this collection: vegetation continuous fields data [DeFries et al., 2000, see Table 1], that describe the percent of a grid cell covered by herbaceous and/or woody vegetation; the historical croplands data set of Ramankutty and Foley [1999]; climate data from



**Table 1.** Listing of Land Cover Related Data Sets Provided in the International Satellite Land Surface Climatology Project (ISLSCP) Initiative II Data Collection<sup>a</sup>

Data Category and Data Set Title	Author(s) and Originating Institution	Input Data Temporal Coverage	Spatial Scale	Data Set Comments
<b>Vegetation</b>				
C4 vegetation percentage	Chris Still, University of California at Santa Barbara	1996–1998	1°	% of each cell which possesses the C4 photosynthetic pathway
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
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)
Historical land cover and land use	Kees Klein Goldewijk, National Institute of Public Health and the Environment (RIVM), The Netherlands	1700–1990	1 and 0.5°	every 50 years (1700–1950); every 10 years (1950–1990);
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
Potential vegetation	Navin Ramankutty and Jonathan Foley, University of Wisconsin	N/A	1 and 0.5°	represents natural vegetation before human alteration
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
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)
<b>Ancillary</b>				
Land/water masks, land outline overlays, latitude and longitude grids	Tom Logan, Jet Propulsion Laboratory; ISLSCP II Staff	N/A	1, 0.5, and 0.25°	binary water masks and fractional water/land cover in each cell

<sup>a</sup>The data collection can be accessed online at <http://www.daac.ornl.gov>.

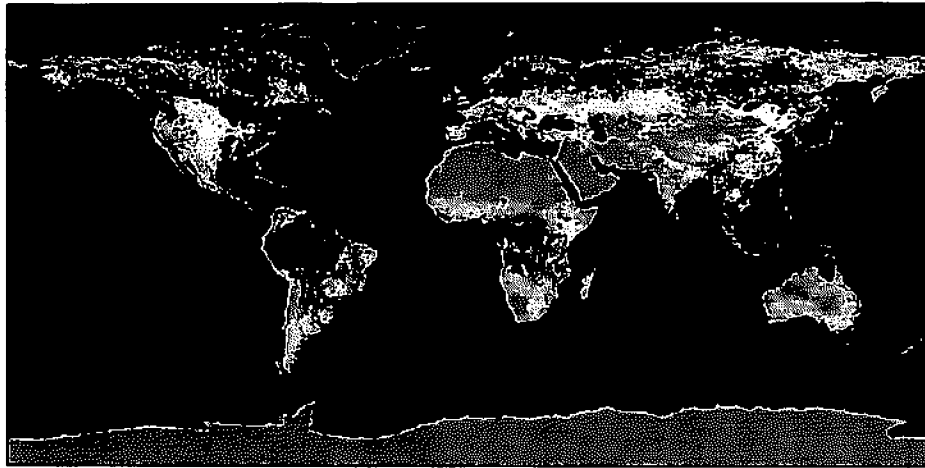
the Climate Research Unit (CRU) at the University of East Anglia in the United Kingdom [New *et al.*, 2000]; and national crop type harvest area statistics from the Food and Agricultural Organization (FAO) of the United Nations (UN-FAO) and the United States Department of Agriculture (USDA). We refer the reader and future users to the data set documentation and/or the above references for more in-depth and specific information on the production of the data sets in Table 1.

[14] The AVHRR-based data sets on Table 1 (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]. In the process of aggregating from 1 km to 1/4, 1/2 and 1° spatial resolutions the percentages of each class within the coarser cell are calculated and allow the dominant land cover type to be determined. Thus each data set contains one global layer with the dominant type and one layer each per cover type showing the percentage of that cover type in each cell (Figure 1). So in addition to improved classification algorithms, improved input data and spatial resolution, the ISLSCP Initiative II land cover “suite” now provides the user with a thorough description of the subcell variability in land cover that was not available in ISLSCP I data. Therefore the users may now also use

their own rules for aggregation using the layers for each land cover class and potentially produce products that better suit their needs.

[15] The IGBP-DIScover product is the only currently available global product that has been validated against a truly global, and statistically valid, independent set of high-resolution data (~70% accuracy), although it is difficult to say what those numbers may correspond to at the 1/4, 1/2, or 1° spatial scales since the validation was done at 1 km resolution [Scepan, 1999]. In fact, an analysis by DeFries and Los [1999] suggested that the implications of the classification errors for model parameterizations were substantially lessened at coarser resolutions when compared to the native 1 km resolution of the IGBP-DIScover product. Both the UMD and MODIS data sets have been evaluated against subsets of the same data use to train the classifiers and also have global accuracies near 70% [Hansen *et al.*, 2000; <http://geography.bu.edu/landcover/userguide/c/consistent.htm>]. In addition, the MODIS land cover product provides gridded estimates of classifier confidences for each cell, addressing an additional user-stated need for accuracy estimates of the products. Finally, in this collection, the IGBP-DIScover data set is provided in three different classification schemes (IGBP, SiB, BATS) to better support the needs of the modeling community.





**Figure 2.** UMD vegetation continuous fields product at  $1/4^\circ$  spatial resolution. The % bare, herbaceous, and woody cover for each cell has been coded as red, green, and blue, respectively, to create this global representation. Other vegetation continuous fields provided in ISLSCP Initiative II include leaf type (needleleaf, broadleaf) and longevity (deciduous, broadleaf) for tree cover. Color appears in back of the print issue.

[16] New types of previously unavailable land cover information are also available at the coarse scale in the form of continuous fields of vegetation cover which describe the % tree, grass and bare cover of each cell (Figure 2), and the % leaf type and/or leaf longevity for tree canopies. All these data sets have been made consistent with the ISLSCP Initiative II land/water mask, which is also based on 1 km original data, and which also contains subcell fractions of water and land at each resolution (see Table 1). This was done by first adjusting the percentage of land and water of each product to correspond to the percentages in the ISLSCP II mask, then recalculating the proportion of each land cover type in the cell on the basis of the new percentage of land, and then producing the dominant cover type maps using aggregation rules specific to each product.

#### 2.4. Previous Land Cover Comparisons

[17] In addition to the product comparisons described by Matthews [1983] and DeFries and Townshend [1994b], Hansen and Reed [2000] have compared the UMD and IGBP-DIScover classifications which are derived from the common 1992–1993 AVHRR data set. They found that the overall per-cell agreement of the two data sets at their original 1 km resolution for all common classes was 48%. For aggregated classes such as forest/woodland, grass/shrubs, crops, this increased to 74% and to 84% when considering even broader classes such as tall woody land cover versus short and/or sparsely vegetated lands. While they found that in general the IGBP-DIScover had more areas of all forest types and 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 principal differences were along transition zones between large core areas. Another difficulty in the comparison was the lack of natural vegetation/croplands mosaic classes in the UMD map, and to a lesser extent, permanent wetlands and ice classes. Finally, Hansen and Reed [2000] determined that the overall agreement between IGBP-DIScover and UMD was much greater at  $0.5^\circ$  reso-

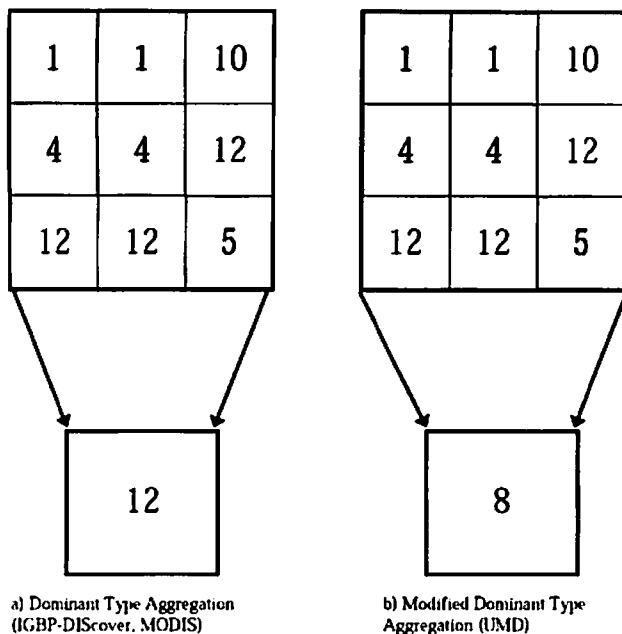
lution than the agreement of the well-known digital land cover maps of Olson *et al.* [1983] and Matthews [1983]. Their results show a significant decrease of 46% in the amount of disagreement between the remotely sensed data sets as opposed to the digital maps for four broad land cover categories. In this study we have performed similar comparisons using the coarse-scale data in ISLSCP Initiative II at  $1/4$ ,  $1/2$  and  $1^\circ$  resolutions, but we have also analyzed the effects of the aggregation methods on the agreement of the two data sets and used the new data layers available in ISLSCP II (subcell proportion of classes and continuous fields products) to assess the areas of disagreement.

#### 3. Data and Methods

[18] The data sets used in the land cover comparison here were the IGBP-DIScover data set using the 17-class IGBP legend [Loveland and Belward, 1997], and the UMD land cover classification using 15 classes [Hansen *et al.*, 2000], at  $1/4$ ,  $1/2$  and  $1^\circ$  spatial resolutions. The MODIS product was not considered in this comparison because of the nearly 10-year gap between the products. The  $1/4^\circ$  UMD vegetation continuous fields product was also evaluated against the  $1/4^\circ$  UMD land cover product to check for data set consistency. As discussed by Hall *et al.* [2006], the incompatibilities in land cover legends between the UMD and IGBP schemes meant that we could only compare similar classes and not classes such as the IGBP natural vegetation/croplands mosaic, which is not included in the UMD product, or the permanent wetlands category. We did however examine the IGBP mosaic class in terms of its subcell makeup in terms of the UMD classes. Also, even though the UMD product did not contain a permanent ice category, we did compare the IGBP-DIScover ice category against the UMD bare category. In total, we were able to compare the two products over 94% of the land surface.

[19] The two data sets have been aggregated to coarser resolution using somewhat different rules that can influence the dominant type found on the final land cover map. The IGBP-DIScover product was aggregated using a strictly





**Figure 3.** Illustration of two methods of aggregation of finer-resolution land cover products to coarser resolutions. (a) For the IGBP-DIScover and MODIS products, the dominant type is selected from the maximum percentage of any cover type in the cell, irrespective of type. (b) In the UMD “modified” approach, the presence of multiple forest types within the cell is accounted for as well as an estimation of the amount of woody cover. The resulting aggregation is more robust for Figure 3b because the cell is at least 56% forest and only 33% cropland. IGBP classes shown are evergreen needleleaf forest (1), deciduous broadleaf forest (4), mixed forest (5), wooded savannas (8), grasslands (10), and croplands (12).

dominant land cover type, whereby the land cover type with the largest percentage in the cell was selected as the dominant type, regardless of any other information included the cell (see Figure 3a). As can be seen in Figure 3a, however, a purely dominant type approach can create results that may emphasize the importance of single land cover types (e.g., croplands in Figure 3a) over multiple land cover types such as forests. In Figure 3a, croplands is selected as the dominant type even though the three forest types account for 56% of the cell coverage. We should note that Figure 3 shows just an illustration of the aggregation with only nine cells and that for the ISLSCP Initiative II data sets, windows of  $120 \times 120$ ,  $60 \times 60$  and  $30 \times 30$  cells of 1 km (i.e., 30 arc-seconds) were averaged to obtain the percentages in each 1,  $1/2$  and  $1/4^\circ$  cell, respectively.

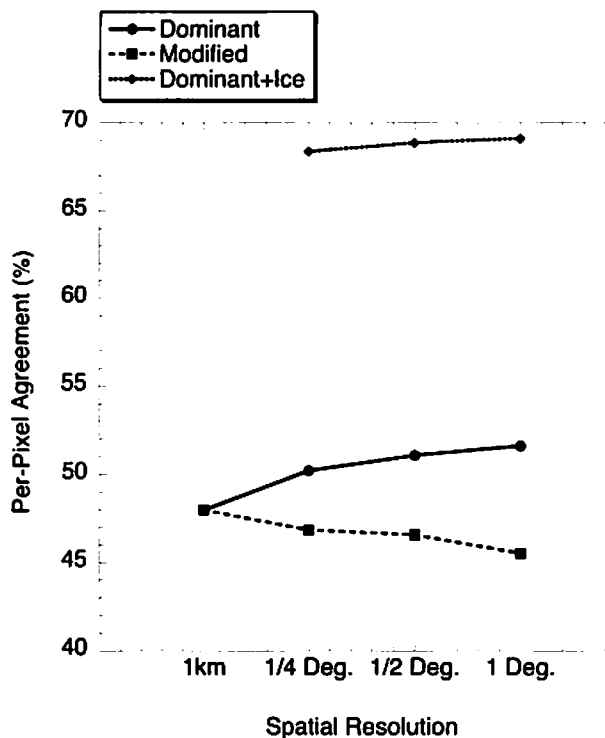
[20] The UMD product was aggregated by the data providers using what we have termed a “modified” dominant type approach (Figure 3b). This approach uses the UMD class definitions, particularly in terms of woody cover, and voting rules that account for the woody cover of the aggregated cell, to assign a dominant type and attempts to overcome some of the issues seen with a strictly dominant approach. For example, this approach assigns forest land cover types when the forest cover of the cell is

greater than or equal to 60%, wooded grasslands for forest covers between 40% and 60%, and so on. From the results shown in Figure 3b, it appears that the modified approach may account for the subcell variability of cover types in a more robust fashion than the dominant approach, although the results globally are not substantially different for most core areas.

[21] We have compared the IGBP-DIScover product with the UMD data set first using a dominant approach for both, and then the modified approach for UMD as is currently provided in the collection, and analyzed the results in terms of overall agreement at the various resolutions, global land cover proportions for each land cover type as well as per class agreements. For the principal areas of disagreement between the two data sets, we have used the subpixel proportions to better understand the nature of the differences, but also checked the makeup of the IGBP natural vegetation/croplands mosaic against the UMD proportions. Finally, we have also used the UMD vegetation continuous fields data sets to both check the areas of disagreement between the two land cover maps but also to check the correspondence and consistency of this product and the UMD land cover product. Clearly, these types of analyses would not have been possible if the data sets did not consistently overlay each other in terms of land/water areas so that the work of making these data sets consistent with the ISLSCP II land/water mask was essential.

#### 4. Results and Discussion

[22] Figure 4 shows the results of the per-pixel comparisons of the two land cover data sets at multiple resolutions. It is important to note that these comparisons provide a good indication as to the level of consistency between the data products and are not meant to imply that one product is necessarily superior. The results show that, when using a strictly dominant criterion for both data sets, the agreement increases as the resolution gets coarser, from 48% at 1 km [Hansen and Reed, 2000], to 50.23% at  $1/4^\circ$ , 51.10% at  $1/2^\circ$  and 51.63% at  $1^\circ$ . When adding the agreement of the IGBP permanent ice and the UMD bare categories, the two data sets agree over 68% of the global land surface, a remarkable agreement given the significant algorithmic differences in generating the data sets. In contrast, when comparing the IGBP-DIScover with the modified dominant type UMD map, we find that the agreement decreases with coarser resolution, from 46.87% at  $1/4^\circ$  to 45.55% at  $1^\circ$ . Clearly, it appears that the methods of aggregation to coarser resolution do have an important effect and that the agreement of the data sets increases with coarser resolution for both data sets aggregated using the same dominant method. The modified approach tends to create more and more “mixed” pixels classes such as woodlands as the resolution gets coarser and thus begins to diverge with the strictly modified version which will emphasize the dominance of single cover types. Nonetheless, when looking at the spatial differences between the two maps, our findings are in agreement with those of Hansen and Reed [2000] in that the large core areas of land cover are mapped similarly in both products, and the larger differences are found in transition zones between similar cover types. Again the overall agreement of nearly 50% is a clear improvement

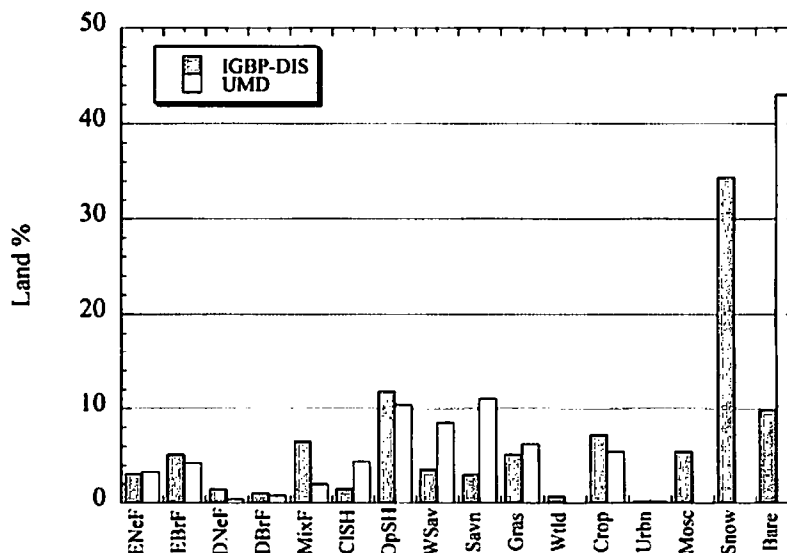


**Figure 4.** Per-cell agreement of IGBP-DIScover and UMD land cover products at several spatial resolutions and using either a dominant or “modified” dominant aggregation scheme. Results of a comparison including the IGBP-DIScover permanent ice category against the UMD bare category are also shown. The agreement value of 48% at 1 km resolution is from *Hansen and Reed* [2000].

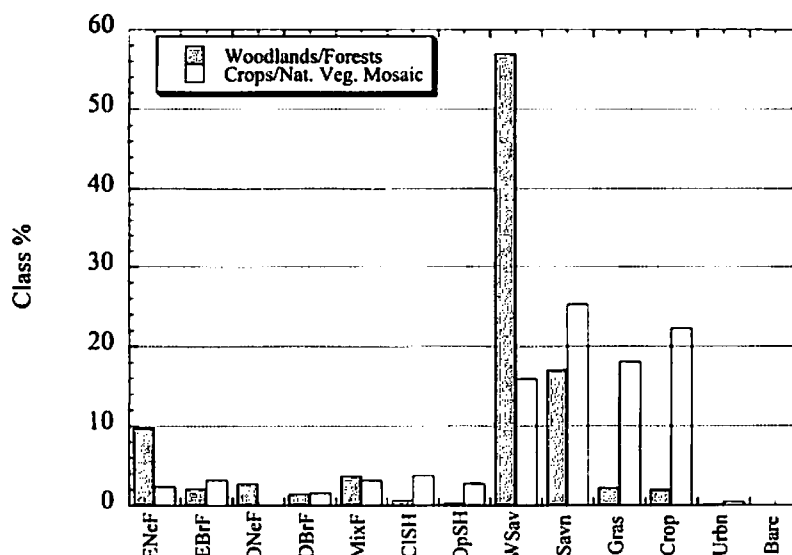
over the 26% agreement found by *DeFries and Townshend* [1994a] using the *Olson et al.* [1983], *Matthews* [1983] and *Wilson and Henderson-Sellers* [1985] maps.

[23] We averaged the subcell proportion layers associated with each map to obtain the percentage of each land cover type on the surface of the Earth estimated by each map, as shown on Figure 5 for the 1° resolution data. Similar to what *Hansen and Reed* [2000] did with global land area totals for each class but with proportions here, we find that at this resolution the IGBP-DIScover map contain more forests than the UMD map (+6.51%) while the UMD has substantially more woody savannas and savannas (+13%), corresponding to the woodlands and wooded grasslands UMD classes. Proportions for the shrublands classes, grasslands and croplands are quite close, within 2%, while the proportion of the IGBP ice and bare categories are 1% higher than the UMD bare categories. We also found that the proportions for the urban classes of the two data sets did not match, even though both data sets used the same Digital Chart of the World as the source for this class. Differences in the water proportions for both data sets were also found but addressed by making the data consistent with the ISLSCP II land/water mask.

[24] An examination of the typical areas of disagreement for each land cover class showed substantial disagreement between the five different forest types, although the evergreen broadleaf forest compared quite well between the two data sets. Over 90% of this class in the UMD map was mapped accordingly in the IGBP-DIS while 72.28% of all the IGBP-DIScover evergreen broadleaf forest was mapped as such in the UMD data set. In general the UMD forest types, except for deciduous broadleaf forest, were fairly well mapped in the IGBP-DIScover product. Reasonable agreement was found for the open shrublands (>50%), croplands, grasslands and bare categories. Fairly large disagreement was found in the woody savanna, savanna, and closed shrublands classes but very often the source of this disagreement was between similar classes such as



**Figure 5.** Global land cover proportions for the 1° IGBP-DIScover and UMD land cover products derived from an average of the subcell proportion layers at the same resolution.



**Figure 6.** Subcell composition of areas of disagreement between the IGBP-DIScover five forest classes and the UMD woodlands and woody grasslands classes. The graph also shows the subcell composition of the IGBP-DIScover natural vegetation/croplands mosaic class in terms of UMD cover types. These are the types of analyses and comparisons that can be made with the subcell proportion data available in this collection.

woody savanna and savanna, or between closed and open shrublands, for example, and not between core classes like forests and bare soil or croplands. In fact, 15% of the total disagreement could be attributed to the UMD woodlands (i.e., woody savanna) class alone, and together with the wooded grasslands (i.e., savanna) class, accounted for 22% of all the total disagreement. Likewise, the IGBP-DIScover mixed forest class was confused across almost all the UMD classes. When grouping all forest classes into one, both savanna classes into another and both shrublands classes into yet another, the agreement was approximately 60%, indicating that core classes compared well among the two products, as shown by *Hansen and Reed* [2000].

[25] We used the tools provided in the ISLSCP Initiative II collection to assess the typical areas of disagreement between the data sets and also as a way to explore the “severity” of these disagreements. As an example, we used the per-pixel proportions to check the composition of the most common areas of disagreement which were between the UMD woodlands and wooded grasslands classes and the IGBP-DIScover forest classes. Results are shown in Figure 6. What Figure 6 shows is that for the most part the areas of disagreement are made up by a majority of the woody savanna class, about 17% of the savanna and 10% evergreen needleleaf forest. What this analysis confirms is that the disagreements are indeed between classes with similar tree cover and are found in areas where forests transition into more open canopies and then into savannas, such as the ecotones of the boreal forest, or the Miombo woodlands of Africa. In Figure 6 we also show the subcell composition of the IGBP-DIScover natural vegetation/croplands mosaic class in terms of the UMD cover types which show that this class is made up of a mixture of croplands with the other cover types, but principally with the savanna (i.e., woody grasslands), grasslands and woody

savannas (i.e., woodlands). This is entirely consistent with the definition of this IGBP mosaic class and show how these new types of data available in this collection can be used for meaningful analyses.

[26] The areas of disagreement have also been explored with the UMD vegetation continuous fields data. Although this data set was generated separately from the UMD land cover product and the approaches to generate each product are different, we should note that the input AVHRR data and the training data are the same for both data sets. Figures 7a–7c show histograms for % grass, % tree and % bare cover for each cell in the 1/4° data sets where we found disagreement between the IGBP-DIScover forest types and the UMD woodlands classes. For all of these cells the mean value of % bare, grass and tree cover was 2.76, 51.21, and 44.75%, respectively. These figures show that, while there are some apparent inconsistencies between the two UMD products with unusually high and/or low values of tree cover or grass cover, and also while there are some substantial disagreements between the two land cover maps, this disagreement is not as significant when considered in terms of tree cover. Figure 7b shows that indeed many of the confused areas are likely to be transition zones between forest types, with greater than 60% tree cover, to woody savannas with 40 to 60% tree cover, as seen in Figure 6 as well. Figure 7a also confirms the confusion with savanna classes with tree cover values between 10 and 40% and Figure 7c confirms that there are few areas of confusion with areas of low tree cover. These percentages of tree cover are consistent with the IGBP-DIScover and UMD definitions of woody savanna and savanna classes yet the smaller differences seen here in terms of tree cover are amplified when comparing discrete classes. We would like to reemphasize to potential users that while the maps may not always agree on a category by category basis and cell by

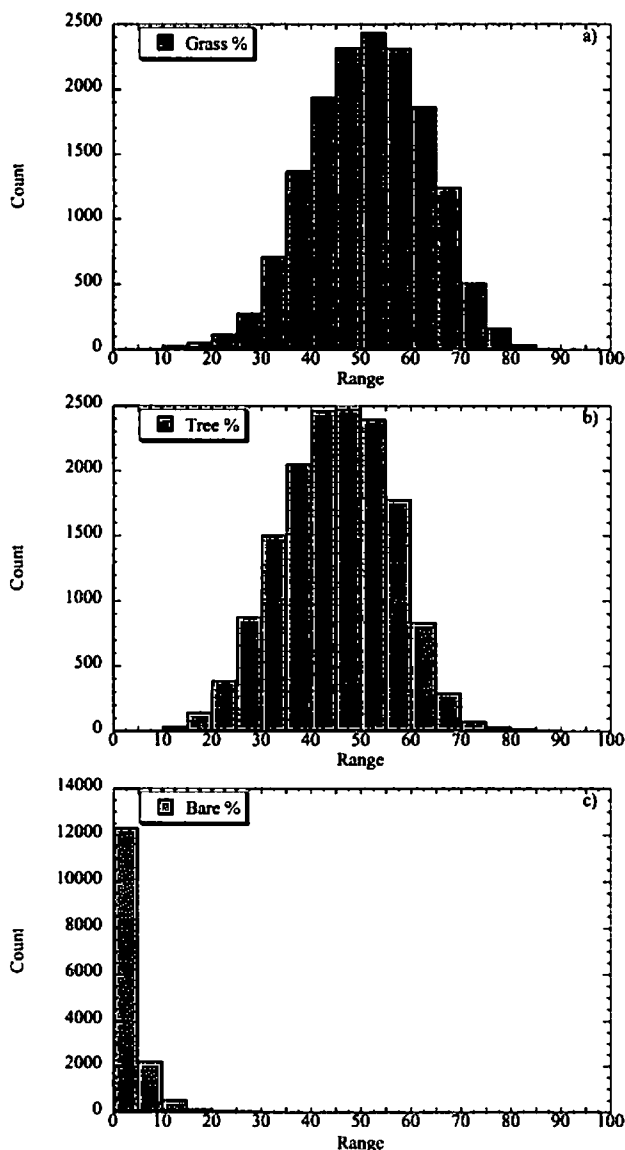


cell basis, the differences in terms of the actual canopy cover types considered are usually not large.

[27] As a final step in our evaluation, we have compared the UMD continuous fields and land cover products to check for internal consistency and also as an evaluation of the UMD continuous fields data provided here. Table 2 shows the mean continuous fields values for all cells in each UMD land cover type from the  $1/4^\circ$  data sets. Overall, the mean proportions of bare, grass and tree cover are entirely consistent with the UMD land cover definitions for each land cover type, with forest types above 60% tree cover, for example, woodlands with >40% tree cover and woody grasslands with tree cover greater than 10%. There are also some outliers where we find bare pixels with high tree cover

**Table 2.** Mean Subpixel Proportions for Each UMD Land Cover Type in Terms of the UMD Vegetation Continuous Fields Data

Land Cover Class	% Bare	% Grass	% Tree
ENeF	1.79	35.4	60.98
EBrF	2.60	22.11	73.64
DNeF	1.78	35.63	61.74
DBrF	1.11	30.01	67.91
MixF	1.73	31.02	65.91
CISh	15.58	79.68	3.42
OpSh	38.15	59.65	0.69
Wdld	2.91	53.86	41.95
WoGr	4.12	72.32	22.09
Gras	4.12	87.50	4.61
Crop	1.87	87.61	9.5
Urbn	2.41	76.36	15.77
Bare	95.47	3.83	0.02



**Figure 7.** Histograms showing the subcell proportions of (a) % grass, (b) % tree, and (c) % bare cover from the UMD vegetation continuous fields data set for all cells where the IGBP-DIScover forest types and the UMD woodlands categories disagreed at a  $1/4^\circ$  resolution.

and/or forest types with high bare cover, but these are uncommon. These inconsistencies do point to the need for a potentially common, or at least internally consistent, processing method(s) for these types of data sets. Finally, the mean values in Table 2 for both the evergreen needleleaf and deciduous needleleaf forest classes are lower than for the other classes, and close to 60% tree cover, which may explain some of the disagreements seen with the IGBP-DIScover maps for these classes. Likewise the mean value of 41.95% tree cover for the woody savannas class is somewhat low and may be indicative of the source of disagreements with this class. Also, it is interesting to note that the mean composition of the urban class according to these data contains very little bare areas but again this class has not been provided by the remotely sensed data but rather has been superimposed from a static database.

## 5. Conclusions

[28] The various land cover data sets provided in the ISLSCP Initiative II data collection to support global modeling efforts represent the tremendous progress made in this area over the past decade or so. From the first global data set from remotely sensed data provided in ISLSCP I [DeFries and Townshend, 1994a], we have now progressed to multiple and improved data sets that accurately describe the past, present and future land cover conditions on the Earth. The algorithms to generate such data sets have progressed to include machine learning classifiers such as decision trees [Hansen et al., 2000; Friedl et al., 2002] which can efficiently handle nonnormal distributions in the training data; new global training data sets generated from high-resolution data and improved input data at higher spatial resolutions are now available and being used for validation as well; the number and types of classes that can be provided has been greatly expanded, and finally, particularly for modeling applications, the subcell variability in terms of land cover types is provided with each data set. New types of data sets with subcell information have also become available such as the vegetation continuous fields data and the  $C_4$  fraction data set of Still et al. [2003], giving the user great flexibility in land cover class definitions but also providing a better representation of landscape continuity across land cover types and ecotones. All of these new and improved data sets should in turn provide improved

estimates for those models that are using them for land surface parameterizations.

[29] Our comparison of the two most widely used land cover data sets for the 1990s shows that, despite some real differences at the level of individual classes, the two data sets agree over nearly two thirds of the Earth's surface and for large core classes. The comparison also shows that clearly, the methods used for aggregation of the products can have a significant impact on the final land cover product. However, while the "modified dominant" approach does appear to produce more robust results, there are to date no guidelines or "best practices" for users to follow in the aggregation of land cover from moderate to coarse resolutions. Also, more analyses are needed to assess the impact of these aggregation methods on the final results, and their subsequent impacts upon the models that use them.

[30] The areas of disagreement, when considered in terms of tree cover, are shown to be transition zones between similar classes and as such their impact on modeling studies may not be as severe as the disagreement between discrete classes seems to indicate. The level of agreement between the data sets also shows a marked improvement over the agreement of previously available digital data sets. However, the data set inconsistencies seen here do point to the need for better integration and harmonization of efforts and more consistent approaches aimed at reducing interproduct differences and thus facilitate the use of the data. It is also critical to note that the increased agreement between two data sets does not necessarily make either one correct, since they could agree to 100% and yet still be both wrong. This points to the need for a sustained and continued effort of independent validation of these data which will allow absolute accuracies to be derived for each product and facilitate intercomparisons. Of particular interest to the modeling community will be the validation at coarser resolutions such as  $\frac{1}{4}$ ,  $\frac{1}{2}$  or  $1^\circ$ .

[31] There are a number of issues that remain to be resolved from a continuing dialogue between the users and producers of these global land cover data sets. The first is the need for consistent land cover legends to support a majority of users. This can be facilitated by the development of a global Land Cover Classification System (LCCS) [DiGregorio and Jansen, 2000] by the FAO which is a standardized, hierarchical and flexible classification scheme that can be applied irrespective of the source or spatial resolution of the input data. The appeal of the LCCS for global land cover data sets is that it can also be easily collapsed and/or expanded into more or less classes (i.e., cross-walked) to support a wide variety of users. Alternatively, continuous fields approaches that completely bypass the classification scheme may provide a more flexible and accurate approach for land cover product generation. However, this approach will demand some parallel model development so that these data can be used more effectively than they are in current global models. Likewise, classes such as urban areas and wetlands need to be better integrated into future global models.

[32] Finally, the one critical land cover data set that is missing from ISLSCP Initiative II is land cover/use change or disturbance. With new and improved algorithms and data sets, it is now possible to generate time series of land cover

products from remotely sensed data, instead of the static, 1992–1993 data set provided here. The challenge remains the integration of "historical" data sets such as those of the AVHRR with those of MODIS and VIIRS, and more importantly, the development of an approach that can consistently separate interannual vegetation changes from actual land cover/use change. It also remains a significant challenge to archive, monitor and generally upkeep the global, high-resolution training data used for the classifications as these change and/or are updated over time. As we look to the future and the NPOESS systems it will be critical that the long-term land cover/use record be established and maintained.

[33] **Acknowledgments.** A large data collection such as ISLSCP Initiative II would simply not have been possible without the substantial in-kind contributions of the data set producers, evaluators and reviewers. The authors would like to express their sincere gratitude to all those who so graciously contributed their time and efforts to the success of this collection. We would like to especially acknowledge the efforts and perseverance of Forrest Hall and Blanche Meeson in leading the charge for the production of the collection.

## References

- Asrar, G., M. Fuchs, E. T. Kanemasu, and J. L. Hatfield (1984), Estimating absorbed photosynthetically active radiation and leaf area index from spectral reflectance in wheat, *Agron. J.*, **76**, 300–306.
- Bartholomé, E., and A. S. Belward (2005), GLC2000: A new approach to global land cover mapping from Earth observation data, *Int. J. Remote Sens.*, **26**, 1959–1977.
- Bonan, G. B. (1995), Land-atmosphere interactions for climate system models: Coupling biophysical, biogeochemical, and ecosystem dynamical processes, *Remote Sens. Environ.*, **51**, 57–73.
- Bonan, G. (1997), Effects of land use on the climate of the United States, *Clim. Change*, **37**, 449–486.
- Bonan, G. B. (2002), *Ecological Climatology: Concepts and Applications*, Cambridge Univ. Press, New York.
- Bondeau, A., D. W. Kicklighter, J. Kaduk, and the Participants of the Potsdam NPP Model Intercomparison (1999), Comparing global models of terrestrial net primary productivity (NPP): Importance of vegetation structure on seasonal NPP estimates, *Global Change Biol.*, **5**, 35–45.
- Bounoua, L., R. S. DeFries, G. J. Collatz, P. Sellers, and H. Khan (2002), Effects of land cover conversion on surface climate, *Clim. Change*, **52**, 29–64.
- Brown de Colstoun, E. C., W. Yang, R. DeFries, M. Hansen, and J. Townshend (2000), Surface type visible/infrared imager radiometer suite algorithm theoretical basis document, version 3.0. Natl. Polar-orbiting Oper. Environ. Satell. Syst., Silver Spring, Md. (Available at [http://npoesslib.ipo.noaa.gov/u\\_listcategory\\_v3.php?35](http://npoesslib.ipo.noaa.gov/u_listcategory_v3.php?35)).
- Chamey, J., P. H. Stone, and W. J. Quirk (1975), Drought in the Sahara: A biogeophysical feedback mechanism, *Science*, **187**, 434–435.
- Daughtry, C. S. T., K. P. Gallo, S. N. Goward, S. D. Prince, and W. P. Kustas (1992), Spectral estimates of absorbed radiation and phytomass production in corn and soybean canopies, *Remote Sens. Environ.*, **39**, 141–152.
- DeFries, R. S., and S. O. Los (1999), Implications for land-cover misclassification for parameter estimates in global land-surface models: An example from the Simple Biosphere Model (SiB2), *Photogramm. Eng. Remote Sens.*, **65**, 1083–1088.
- DeFries, R. S., and J. R. G. Townshend (1994a), NDVI-derived land cover classifications at a global scale, *Int. J. Remote Sens.*, **15**, 3567–3586.
- DeFries, R. S., and J. R. G. Townshend (1994b), Global land cover: Comparison of ground-based data sets to classifications with AVHRR data, in *Environmental Remote Sensing From Regional to Global Scales*, edited by G. Foody and P. Curran, pp. 84–110, John Wiley, Hoboken, N. J.
- DeFries, R. S., M. Hansen, and J. Townshend (1995), Global discrimination of land cover types from metrics derived from AVHRR pathfinder data, *Remote Sens. Environ.*, **54**, 209–222.
- DeFries, R. S., M. Hansen, J. R. G. Townshend, and R. Sohlberg (1998), Global land cover classifications at 8 km spatial resolution: the use of training data derived from Landsat imagery in decision tree classifiers, *Int. J. Remote Sens.*, **19**, 3141–3168.
- DeFries, R. S., J. R. G. Townshend, and M. Hansen (1999), Continuous fields of vegetation characteristics at the global scale at 1 km resolution, *J. Geophys. Res.*, **104**, 16,911–16,925.

- DeFries, R. S., M. C. Hansen, J. R. G. Townshend, A. C. Janetos, and T. R. Loveland (2000), A new global 1-km dataset of percentage tree cover derived from remote sensing, *Global Change Biol.*, **6**, 247–254.
- Dickinson, R. E. (1983), Land surface processes and climate-surface albedo and energy balance, *Adv. Geophys.*, **25**, 305–350.
- 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.
- DiGregorio, A., and L. J. M. Jansen (2000), Land Cover Classification System (LCCS): Classification concepts and user manual, U. N. Food and Agric. Organ., Rome.
- 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.
- Friedl, M. A., et al. (2002), Global land cover mapping from MODIS: Algorithms and early results, *Remote Sens. Environ.*, **83**, 287–302.
- Hall, F. G., E. Brown de Colstoun, G. J. Collatz, D. Landis, P. Dirmeyer, A. Betts, G. 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.*, doi:10.1029/2006JD007366, 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.
- Hansen, M. C., R. S. DeFries, J. R. G. Townshend, R. Sohlberg, C. Dimiceli, and M. Carroll (2002), Towards an operational MODIS continuous field of percent tree cover algorithm: examples using AVHRR and MODIS data, *Remote Sens. Environ.*, **83**, 303–319.
- Houghton, R. A. (2003), Revised estimates of the annual net flux of carbon to the atmosphere from changes in land use and land management 1850–2000, *Tellus, Ser. B*, **55**, 378–390.
- James, M. E., and S. N. V. Kalluri (1994), The pathfinder AVHRR land data set: An improved coarse resolution data set for terrestrial monitoring, *Int. J. Remote Sens.*, **15**, 3347–3364.
- Justice, C. O., J. R. G. Townshend, B. N. Holben, and C. J. Tucker (1985), Analysis of the phenology of global vegetation using meteorological satellite data, *Int. J. Remote Sens.*, **6**, 1271–1318.
- Kicklighter, D. W., A. Bondeau, A. L. Schloss, J. Kaduk, A. D. McGuire and the Participants of the Potsdam NPP Model Intercomparison (1999), Comparing global models of net primary productivity (NPP): Global pattern and differentiation by major biomes, *Global Change Biol.*, **5**, 16–24.
- 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.
- 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.
- Loveland, T. R., J. W. Merchant, D. O. Ohlen, and J. F. Brown (1991), Development of a land-cover characteristics database for the conterminous U.S., *Photogramm. Eng. Remote Sens.*, **57**, 1453–1463.
- Manabe, S. (1969), Climate and the ocean circulation: I. The atmospheric circulation and the hydrology of the Earth's surface, *Mon. Weather Rev.*, **97**, 739–805.
- Matthews, E. (1983), Global vegetation and land use: New high-resolution data bases for climate studies, *J. Clim. Appl. Meteorol.*, **22**, 474–487.
- Melillo, J. M. (1994), Modeling land-atmosphere interactions: A short review, in *Changes in Land Use and Land Cover: A Global Perspective*, edited by W. B. Meyer and B. L. Turner III, pp. 387–410, Cambridge Univ. Press, New York.
- Melillo, J. M., et al. (1993), Global climate change and terrestrial net primary production, *Nature*, **363**, 234–240.
- Mueller-Dombois, D. (1984), Classification and mapping of plant communities: A review with emphasis on tropical vegetation, in *The Role of Terrestrial Vegetation in the Global Carbon Cycle: Measurement by Remote Sensing*, edited by G. M. Woodwell, pp. 19–88, John Wiley, Hoboken, N. J.
- Nemani, R., and S. W. Running (1996), Implementation of a hierarchical global vegetation classification in ecosystem function models, *J. Veg. Sci.*, **7**, 337–346.
- New, M., M. Hulme, and P. Jones (2000), Representing twentieth-century space-time climate variability. Part II: Development of 1901–1996 monthly grids of terrestrial surface climate, *J. Clim.*, **13**(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*, U.S. Dep. of Energy, Oak Ridge Natl. Lab., Oak Ridge, Tenn.
- Plattner, G., F. Joos, and T. F. Stocker (2002), Revision of the global carbon budget due to changing air-sea oxygen fluxes, *Global Biogeochem. Cycles*, **16**(4), 1096, doi:10.1029/2001GB001746.
- Prentice, K. C. (1990), Bioclimatic distribution of vegetation for general circulation models, *J. Geophys. Res.*, **95**, 11,811–11,830.
- Ramankutty, N., and J. A. Foley (1998), Characterizing patterns of global land use: An analysis of global croplands data, *Global Biogeochem. Cycles*, **12**, 667–685.
- Ramankutty, N., and J. A. Foley (1999), Estimating historical changes in global land cover: Croplands from 1700 to 1992, *Global Biogeochem. Cycles*, **13**, 997–1027.
- Running, S. W., and E. R. Hunt Jr. (1993), Generalization of a forest ecosystem process model for other biomes, BIOME-BGC, and an application for global-scale models, in *Scaling Processes Between Leaf and Landscape Levels*, edited by J. R. Ehleringer and C. Field, pp. 141–158, Elsevier, New York.
- Running, S. W., T. R. Loveland, L. L. Pierce, R. R. Nemani, and E. R. Hunt Jr. (1995), A remote sensing based vegetation classification logic for global land cover analysis, *Remote Sens. Environ.*, **51**, 39–48.
- Scepan, J. (1999), Thematic validation of high-resolution global land-cover data sets, *Photogramm. Eng. Remote Sens.*, **65**, 1051–1060.
- Sellers, P. J., S. O. Los, C. J. Tucker, C. O. Justice, D. A. Dazlich, G. J. Collatz, and D. A. Randall (1996a), A revised land surface parameterization (SiB2) for atmospheric GCMs. Part II: The generation of global fields of terrestrial biophysical parameters from satellite data, *J. Clim.*, **9**, 706–737.
- Sellers, P. J., B. W. Meeson, J. Closs, J. Collatz, F. Corpnew, D. Dazlich, and F. G. Hall (1996b), The ISLSCP Initiative I global datasets: Surface boundary conditions and atmospheric forcings for land-atmosphere studies, *Bull. Am. Meteorol. Soc.*, **77**, 1987–2005.
- Skole, D. L., and C. J. Tucker (1993), Tropical deforestation and habitat fragmentation in the Amazon: Satellite data from 1978 to 1988, *Science*, **260**, 1905–1910.
- 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.
- Townshend, J. R. G., C. O. Justice, and V. T. Kalb (1987), Characterization and classification of South American land cover types using satellite data, *Int. J. Remote Sens.*, **8**, 1189–1207.
- 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.
- Tucker, C. J., J. R. G. Townshend, and T. E. Goff (1985), African land-cover classification using satellite data, *Science*, **227**, 369–375.
- Wilson, M. F., and A. Henderson-Sellers (1985), A global archive of land cover and soils data for use in general circulation models, *J. Clim.*, **5**, 119–143.

E. C. Brown de Colstoun, Science Systems and Applications, Inc., Code 614.4, Biospheric Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA. (ericbdc@ltpmail.gsfc.nasa.gov)

R. S. DeFries and J. R. G. Townshend, Department of Geography, University of Maryland, College Park, MD 20742, USA.

## Evaluation of ISLSCP Initiative II FASIR and GIMMS NDVI products and implications for carbon cycle science

Forrest Hall,<sup>1,2</sup> Jeffrey G. Masek,<sup>1</sup> and G. James Collatz<sup>1</sup>

Received 25 April 2006; revised 15 August 2006; accepted 9 October 2006; published 23 November 2006.

[1] Integration of NDVI data into ecological and biogeochemical modeling has placed more stringent requirements on the accuracy and stability of the measurement.

We compare two recent AVHRR NDVI data sets included as part of ISLSCP Initiative II: (1) the Fourier-Adjusted, Sensor and Solar zenith angle corrected, Interpolated, Reconstructed (FASIR) monthly time series and (2) the Global Inventory Modeling and Mapping Studies (GIMMS) monthly time series. Although both started with nearly identical composited AVHRR GAC data sets, each data set has been processed differently to reduce sensor, atmospheric, and illumination effects that vary over time. We find that the resulting absolute NDVI data records differ substantially and consistently for large parts of the globe. These differences also propagate into the NDVI anomaly record (e.g., deviations from monthly or annual means) particularly in the 1984–1985, 1994 periods. To assess the effect of these differences on predictions of land surface CO<sub>2</sub> fluxes, the fraction of absorbed photosynthetically active radiation (fPAR) was calculated from each record, and used to drive a biogeochemical model (CASA). On a global basis, calculated net ecosystem exchange shows large variability inherited from the NDVI records. However, these variations do not match global CO<sub>2</sub> fluxes derived from atmospheric inversion of CO<sub>2</sub> concentration measurements. We conclude that other processes (burning, physiologic response to stress) are likely responsible for major anomalies in the observed global land net carbon fluxes to the atmosphere during the period 1982–1998.

**Citation:** Hall, F., J. G. Masek, and G. J. Collatz (2006), Evaluation of ISLSCP Initiative II FASIR and GIMMS NDVI products and implications for carbon cycle science, *J. Geophys. Res.*, **111**, D22S08, doi:10.1029/2006JD007438.

### 1. Introduction

[2] A primary accomplishment of land remote sensing over the last thirty years has been the development of global vegetation indices for monitoring the terrestrial environment. The contrast between near-infrared and visible reflectance for characterizing vegetation “greenness” was originally recognized using field radiometers [Birth and McVey, 1968] and early Landsat imagery [Rouse *et al.*, 1974]. Gradually, the Normalized Difference Vegetation Index (NDVI) became the standard formulation, in large part because of its insensitivity to variability in illumination conditions [Tucker, 1979]. NDVI is formally defined as  $(\rho_{\text{nir}} - \rho_{\text{vis}})/(\rho_{\text{nir}} + \rho_{\text{vis}})$  where  $\rho_{\text{nir}}$  is either the surface or top of atmosphere reflectance in the near-infrared wavelength range and  $\rho_{\text{vis}}$  is for the visible wavelength range. For narrowband sensors, reflectance in the red wavelength is usually substituted for  $\rho_{\text{vis}}$ .

[3] Increased concern with global climate and environmental changes in the 1980s, coupled with the cost of

Landsat data as a result of privatization of the Landsat program, spurred researchers to use data from the NOAA Advanced Very High Resolution Radiometer (AVHRR) instruments. Although these instruments were designed for operational meteorology, their capacity to acquire daily, global data in visible and near-infrared wavelengths made them useful for tracking vegetation conditions. NOAA began producing the first vegetation index products from AVHRR in 1982 (the Global Vegetation Index or GVI product). Early analyses of the GVI products focused on basic biogeography, seasonality of vegetation patterns at continental scales, and correlations with atmospheric carbon dioxide measurements [Justice *et al.*, 1985; Goward *et al.*, 1985; Tucker *et al.*, 1985, 1986]. Although critical for establishing the utility of global satellite observations for monitoring vegetation, these early studies primarily focused on qualitative relations between the NDVI and vegetation properties. Gradually, researchers developed quantitative relations between NDVI and biophysical variables controlling vegetation productivity and land/atmosphere fluxes [Asrar *et al.*, 1984; Sellers, 1985; Nemani and Running, 1989]. Principal among these are leaf area index (LAI) and the fraction of absorbed photosynthetically active radiation (fPAR). Hall *et al.* [1992] found that NDVI responded almost linearly to the fraction of incident photosynthetically active radiation absorbed by the photosynthetically active

<sup>1</sup>NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.

<sup>2</sup>Also at Joint Center for Earth Systems Technology, University of Maryland Baltimore County, Baltimore, Maryland, USA.



tissue in the canopy (fPAR) and was insensitive to that absorbed by nonphotosynthetic tissue.

[4] These pioneering studies led to a current generation of ecosystem models that use satellite-based vegetation indices for predicting carbon, energy, and water fluxes in response to climate variability and ecosystem disturbance [e.g., *Randerson et al.*, 1996; *Sellers et al.*, 1996; *Kaminski et al.*, 2002; *van der Werf et al.*, 2004]. Ideally, these models should provide understanding of how trends in climate since the early 1980s have affected vegetation patterns, and how these feed back to the climate system [e.g., *Nemani et al.*, 2001; *Bounoua et al.*, 2000]. Several carbon and climate studies have shown that prescribing seasonal phenology of vegetation using NDVI based products produces more realistic model simulations than a phenology based on climatology [Kang et al., 2006]. NDVI products are also used to validate phenology and interannual variability in vegetation grown in so-called dynamic vegetation models [Dickinson et al., 1998; Zeng et al., 1999; Lucht et al., 2002].

[5] The integration of NDVI observations into physical models placed more stringent requirements on the accuracy and stability of the measurement. From the perspective of long-term ecosystem monitoring, the AVHRR observation record suffers from numerous deficiencies including poor sensor calibration, poor intercalibration between successive NOAA platforms, wide variations in solar-view geometry, atmospheric contamination, cloud contamination, and the systematic drift in acquisition time during the life of individual missions [Gutman, 1991; Goward et al., 1993; Los et al., 1994; Privette et al., 1995; Gutman, 1999]. As a result, the AVHRR Global Area Coverage (GAC) record has been reprocessed several times since the mid-1990s in an effort to mitigate these deficiencies [Townshend, 1994; Tucker et al., 1994; James and Kalluri, 1994; Goward et al., 1994; Los et al., 2000; Tucker et al., 2005].

[6] Because NDVI is a ratio of differences between two adjacent bands, it is largely insensitive to variations in illumination intensity. However, NDVI is sensitive to sensor, atmospheric and illumination effects that differ between bands. Band calibrations, for example, have changed frequently between the five NOAA AVHRR instruments that acquired the 22-year NDVI record for Initiative II. In addition, natural variability in atmospheric aerosols and column water vapor have created surface-independent variations in the NDVI record. Finally, over the period of record there were two major volcanic eruptions, El Chichon in 1982 and Mt. Pinatubo in 1991 [Rosen et al., 1994] that injected large quantities of aerosols into the Earth's stratosphere. These aerosols, along with smoke from biomass burning and dust from soil erosion and other factors, introduce significant variability in the AVHRR NDVI record. These constituents have significantly different effects on AVHRR channel 1 (visible) and channel 2 (near infrared). Variations in the illumination conditions over the period of record also introduce spurious variation into the NDVI signal. The AVHRR sensors flew aboard the afternoon NOAA platforms, beginning with the NOAA 7 satellite launched in January of 1980, continuing on NOAA 9, 11 and 14. The NOAA satellite overpass times drifted from the nominal 1:30 pm overpass time by as much as

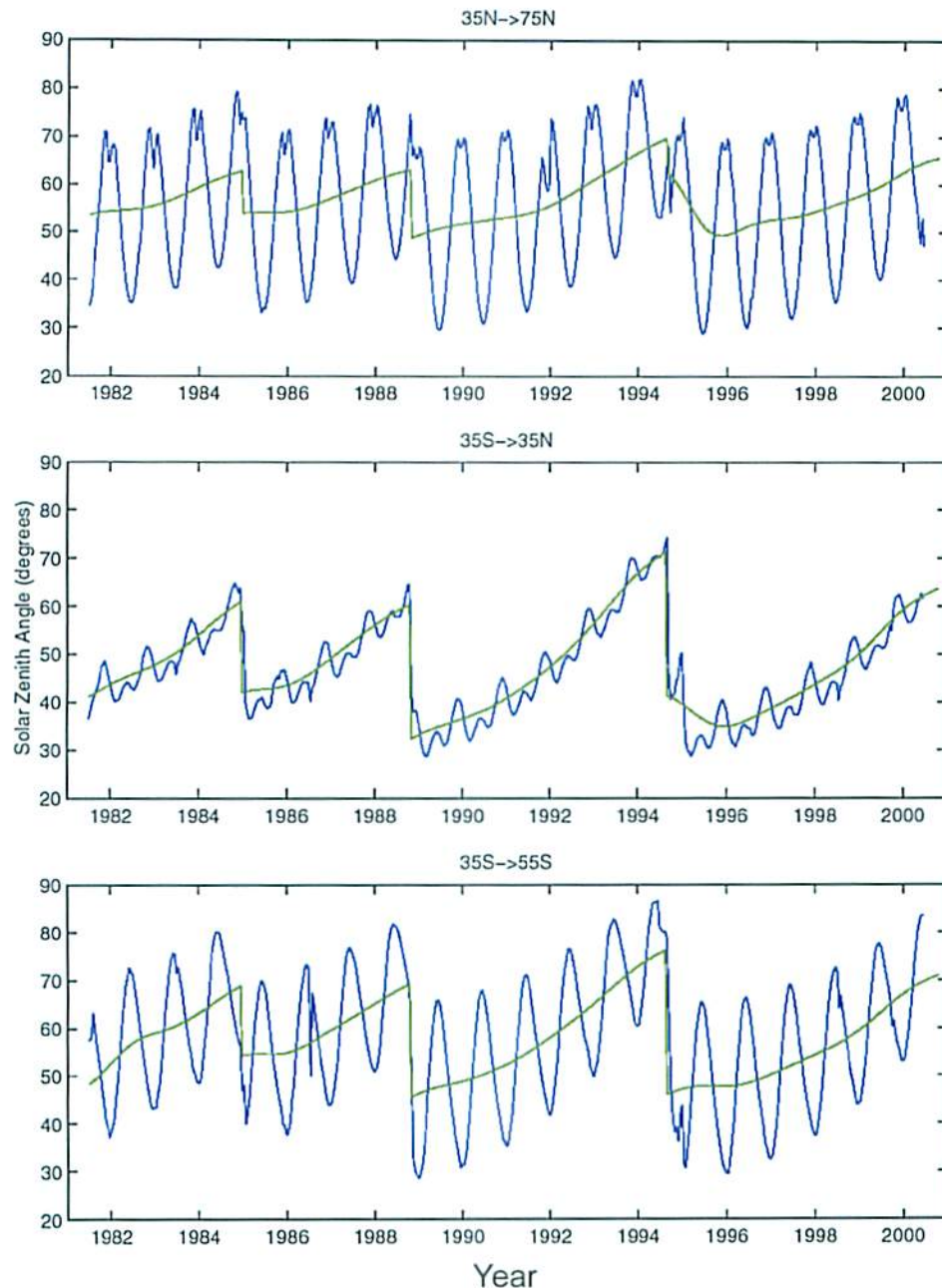
4 1/2 hours toward evening, creating variable illumination and view angles (see Figure 1).

[7] The ISLSCP Initiative II data collection contains two AVHRR NDVI time series, processed using two different algorithms, both aimed at reducing the above mentioned effects on the NDVI signal. These are (1) The Fourier-Adjusted, Sensor and Solar zenith angle corrected, Interpolated, Reconstructed (FASIR) monthly time series 1981–1998 [Los et al., 2005; Hall et al., 2006] and (2) The Global Inventory Modeling and Mapping Studies (GIMMS) monthly time series, 1981 to 2000 [Tucker et al., 2005]. The inclusion of these data sets provides the land science community with a consistently gridded, set of vegetation records for driving process models. Although both data sets start with the NOAA AVHRR 4 km resolution GAC data, each has chosen unique processing approaches for producing a consistent NDVI record. Both the GIMMS and FASIR records were originally produced at 8 km resolution, but are represented within the International Satellite Land Surface Climatology Project (ISLSCP) Initiative II collection using the standard 0.25, 0.50, and 1.0 degree resolution latitude-longitude grids. The ISLSCP Initiative II minimum 0.25 degree equal-angle grid, a fourfold increase in spatial resolution over the Initiative I collection, was chosen to correspond to data input requirements for carbon, water and energy models.

[8] This paper presents a comparison between the ISLSCP FASIR and GIMMS NDVI data sets, with a specific focus on the applicability of the ISLSCP data sets to carbon cycle modeling. We first explore differences between the data sets in terms of absolute NDVI, including comparisons with Landsat-derived NDVI from single clear-sky observations. We then examine differences in terms of interannual anomalies. Finally, we address the implications of observed NDVI variability for global carbon cycle modeling.

## 2. Description of ISLSCP NDVI Data Sets

[9] Key aspects of the FASIR and GIMMS algorithms and their differences are highlighted below in Table 1. 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 (09/94 to 01/95) 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 cloud contamination. The GIMMS processing approach did not utilize atmospheric correction, except during the El Chichon and Mt. Pinatubo volcanic stratospheric aerosol periods, and applied corrections to



**Figure 1.** Variation of latitude-averaged solar zenith angle (blue) and trends (green) from NOAA 7 through NOAA 14 for (top) 35 to 75 north latitude; (middle) 35 north to 35 south latitude; and (bottom) 35 to 55 south latitude.

NDVI directly (i.e., does 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]. These differences between the GIMMS and FASIR products are discussed more fully in the following sections.

## 2.1. Data Input

[10] Both FASIR and GIMMS used data composites from the full AVHRR data record but the compositing periods used are somewhat different. Frequent cloud cover eliminates roughly 2/3 of the daily AVHRR record. In order to construct periodic cloud-free views of the Earth, composite monthly images were constructed by selecting for each pixel the maximum NDVI during 10-day (FASIR) or 15-day (GIMMS) intervals. Choosing each pixel's maximum NDVI during an interval of a few days reduces aerosol, cloud cover and water vapor effects since NDVI

**Table 1.** FASIR and GIMMS Processing Similarities/Differences

	Data Input	Calibration	View/Illum Angle Corr	Atmospheric Corr.	Aerosol Corr.	Other
FASIR Monthly Band 1, Band 2, NDVI	NOAA Pathfinder data set [ <i>James and Kalluri</i> , 1994]; 10-day composites; corrected for mol scattering. NOAA 11 gap 09/94 to 01/95 filled by interpolation.	NOAA Pathfinder data set after <i>James and Kalluri</i> [1994]; additional adjustments using desert calibration targets.	Li Sparse and Ross Thick Kernel [ <i>Wanner et al.</i> , 1995].	molecular and ozone as per <i>James and Kalluri</i> [1994]; not corrected for water vapor; no correction for tropospheric aerosols.	volcanic aerosol corrected Pinatubo 04/82 to 12/84; El Chichon 06/91 to 12/94 [ <i>Sato et al.</i> , 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 09/94 to 01/95 filled with NOAA 9.	<i>Vermote and Kaufman</i> [1995] desert calibration; AVHRR NDVI is adjusted higher to match 1997–2003 SPOT NDVI less affected by atmospheric water vapor absorption.	empirical mode decomp.	no explicit atmospheric corrections; however, compositing reduces atmospheric variability; the match to SPOT NDVI reduces the effect of atmospheric water vapor absorption.	volcanic aerosol corrected Pinatubo 04/82 to 12/84; El Chichon 06/91 to 12/94 [ <i>Sato et al.</i> , 1993; <i>Vermote et al.</i> , 1997].	cloud screen using AVHRR thermal band.

is maximum on the clearest days. It should be noted however, that maximum NDVI compositing does not completely remove these atmospheric effects. Compositing can be done over any time interval, but 9 days is generally selected as the minimum period since the NOAA orbit repeats at that frequency. The primary FASIR data record was constructed from 10-day composites. Since ISLSCP contains monthly data, a monthly FASIR data record was generated for the Initiative II collection based on the composite value for days 11–20 because this middle period best represented the entire month. The GIMMS monthly data record is based the average maximum NDVI for each of the 15-day periods in a month.

[11] FASIR NDVI data sets were compiled using band 1, band 2, solar zenith angle, scan angle, and relative azimuth angle values from the Pathfinder AVHRR Land (PAL) data set [*James and Kalluri*, 1994] for the period of 1982–1998. Pathfinder radiances were determined to be more appropriate for FASIR BRF adjustments than top of the atmosphere measurements from the AVHRR sensors because the Pathfinder data set contained band 1 and band 2 radiances corrected for intersensor calibration differences, atmospheric molecular scattering and ozone absorption (from TOMS). Corrections accounted for atmospheric path length changes induced by topographic variations. Pathfinder also cloud screened the AVHRR data using the CLAVR algorithm.

[12] The input data for the GIMMS 1981–2002 time series were the top of the atmosphere NOAA AVHRR GAC 1B data, obtained from NOAA and from National Center for Atmospheric Research (NCAR). GIMMS augmented this data set with GAC 1B data available from NOAA's Satellite Active Archive. Cloud screening was provided by a channel 5 thermal mask of 0°C for all continents except Africa, where a cloud mask of 10°C was used. In addition bimonthly compositing significantly reduced cloud contamination.

[13] The AVHRR acquisitions used for the 1981 to 1998 time period for FASIR and GIMMS are similar, with an important exception. GIMMS utilized NOAA 9 data in the October 1994 to January 1995 period to fill a gap in the Pathfinder record during this period (NOAA 11 started to malfunction and its replacement, NOAA 13, failed shortly after launch; NOAA 14 was not yet launched). During this period, FASIR estimated AVHRR using a climatological mean and the Fourier Adjustment. Specifically, two low-pass filters ( $\pm 220$  day and  $\pm 50$  day moving windows, respectively) were used to interpolate the missing data for late 1994.

## 2.2. Data Processing

### 2.2.1. Calibration

[14] Both GIMMS and FASIR products recalibrated AVHRR band 1 and 2 reflectance measurements to reduce intersensor and intrasensor errors. FASIR started with the calibration procedure of *Rao and Chen* [1994]. This was improved by using a thousand reflectance-invariant sites globally (e.g., deserts) to examine residual variations in Pathfinder band 1, band 2 and NDVI. On the basis of these observations adjustments were made to the *Rao and Chen* [1994] sensor gains to render the band 1 and 2 reflectance more stable over these targets. Following this, the relative RMS error as a result of sensor degradation and intercali-

bration differences in the channel 1 and 2 gains were estimated to be about 1%.

[15] GIMMS used the technique of *Vermote and Kaufman* [1995] to adjust the calibration of NOAA-7 through NOAA-14 AVHRR channel 1 and 2 data. This approach uses a different calibration standard from *Rao and Chen* [1994] and results in slightly higher NDVI values (S. Los, personal communication, 2005). Like FASIR, the calibration was refined using invariant desert sites. The GIMMS data set extends beyond 1998 to include data from NOAA-16. In order to tie together the NOAA-14 and NOAA-16 time series, GIMMS also adjusted the historical (NOAA-14 and earlier) and NOAA-16 NDVI by a constant offset to match up with a coincident and spatially aggregated 8-km SPOT Vegetation NDVI time series. This was necessary because the bilinear gain for channel 1 and channel 2 of NOAA-16's AVHRR instrument complicates *ex post facto* calibration. It should be noted that calibrating the AVHRR NDVI values to SPOT effectively narrows the instrument band pass, and results in overall higher NDVI values.

### 2.2.2. NDVI Variations Resulting From Variations in Solar and View Zenith Angle

[16] The FASIR algorithm adjusted for illumination and viewing angle effects in the AVHRR NDVI by estimating the BRF for each pixel from its 17-year variation. This is accomplished by employing the Li Sparse and Ross Thick kernel approach used to estimate MODIS BRF [*Wanner et al.*, 1995]. To obtain sufficient numbers of observations, the BRF for each pixel was assumed constant over the 17-year period so that all monthly NDVI values over the 17 years could be used to estimate the kernels. Using the estimated kernel weights, NDVI was normalized to standard viewing geometry (30 degrees solar zenith angle and 0 degrees view angle). In contrast, GIMMS utilized the empirical mode decomposition (EMD) [*Huang et al.*, 1998, 1999], an empirical approach to correct for view and illumination angle effects. EMD extracts NDVI trends that are more than 80% correlated to the solar zenith angle. Areas with trends that have a lower correlation were not corrected.

### 2.2.3. Atmospheric Correction

[17] GIMMS applied no atmospheric correction, except during the El Chichon and Mt. Pinatubo volcanic stratospheric aerosol periods. A stratospheric aerosol correction was applied as proposed by *Vermote et al.* [1997] from April 1982 through December 1984 and from June 1991 through December 1994. GIMMS formed composite stratospheric aerosol optical depth fields by combining the work of *Sato et al.* [1993], and *Vermote et al.* [1997]. The work of *Rosen et al.* [1994], *Russell et al.* [1993] and *Dutton* [1994] were used to compare specific optical depth measurements to GIMMS blended global fields. GIMMS optical depth field varied by month and degree of latitude. The Pathfinder data set on which FASIR is based is corrected for atmospheric molecular scattering and ozone absorption, but not water vapor. FASIR also corrected for aerosols injected into the stratosphere by Mt. Pinatubo and El Chichon, but used the aerosol optical depth data from *Sato et al.* [1993]. The *Sato et al.* [1993] data extends further north and south (90 S to 90 N) than the data set by *Vermote* (50 S to 50 N). Comparison of the NDVI corrected with the Sato and the *Vermote* optical depth data showed close agreement between 50 degrees South and 50 degrees North. Stratospheric

optical depths prior to and 2 years after the eruptions were set to zero for both data sets.

## 3. Geographic Characteristics

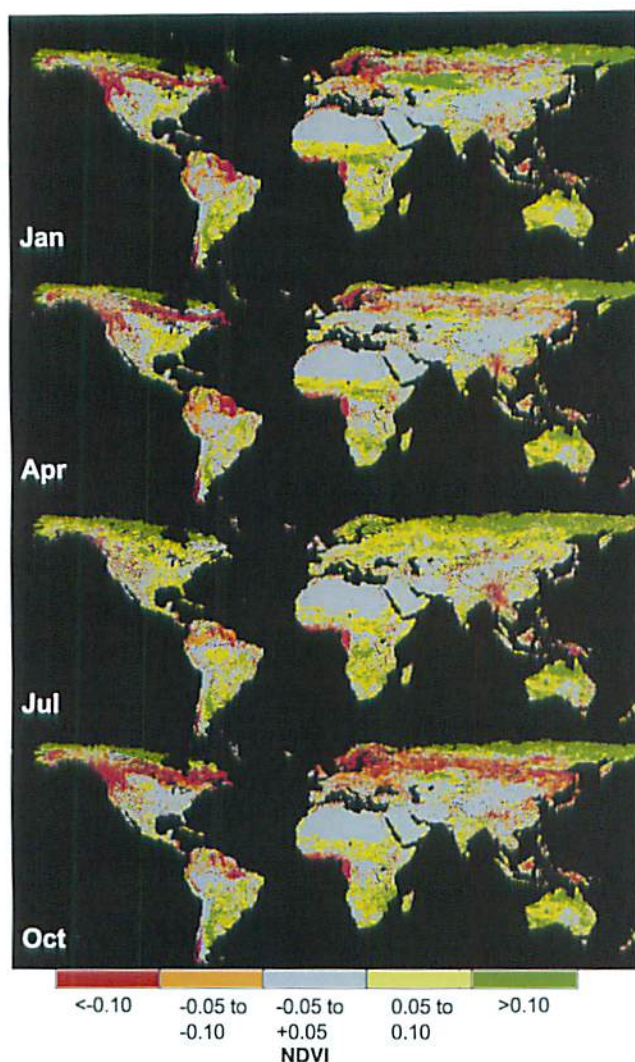
[18] As part of the ISLSCP processing, NDVI data sets were produced at 0.25, 0.50, and 1.0 degree resolution. Original input data for both data sets was the ~4 km resolution NOAA GAC data. The GIMMS NDVI data set was derived directly from this record. The FASIR NDVI data set was derived from the AVHRR Land Pathfinder (PAL) data set, which was resampled to 8 km resolution.

[19] Since NDVI is a nonlinear transformation of the original reflectance data, averaging NDVI from several independent observations across a region will give a different result compared to averaging the original spectral data and then calculating a single, aggregate NDVI value [*Hall et al.*, 1992]. The original NOAA GAC data represent a subsample of the original 1.1 km LAC sensor stream; at nadir only 27% of each 4 km GAC pixel was actually imaged by the instrument [*Townshend*, 1994]. Within the GIMMS processing, these ~4 km NDVI subsamples were mapped to output 8 km grid cells, and the single observation with the maximum NDVI value during the 15-day compositing period was retained. Thus each 64 km<sup>2</sup> 15-day GIMMS NDVI value actually corresponds to an irregularly shaped 4.4 km<sup>2</sup> observation from the compositing period; up to 93% of the 8 km cell was not observed by the instrument. For ISLSCP Initiative II, all 8 km GIMMS NDVI values were averaged within a single 0.25 degree grid cell. The FASIR geographic processing approach was similar, except that the compositing period is the middle 10 days of each month, then the 8 km FASIR NDVI values are averaged to form the 0.25 degree output. The differences in compositing procedure imply that each monthly GIMMS 0.25 degree cell at the equator includes information from roughly twice as many AVHRR 1.1 km LAC observations compared to, since two values for each month are incorporated into the GIMMS product, but only one value for the FASIR.

[20] The nature of the FASIR data processing introduced two types of smoothing that are visible in the NDVI data. First, to reduce tropical cloud contamination, FASIR retained the maximum NDVI value from a moving 3 × 3 pixel window for tropical regions. As a result, the nominal resolution of the FASIR data in the tropics is ~24 × 24 km, about the same as the 0.25 degree resolution of the finest ISLSCP grid. Second, the time series from each pixel was fitted with a Fourier representation to identify and remove outliers and create a temporally smooth curve. The removal of outliers also give the FASIR NDVI imagery a smoother, less noisy appearance compared to GIMMS data.

[21] Overlaying the Initiative II GIMMS and FASIR data sets at 0.25 degree resolution indicates a systematic misregistration of about 1/2 pixel (~13 km at the equator), with the FASIR data offset to the southwest compared to the GIMMS. To evaluate which data set is correctly registered, orthorectified (<60 m 1 $\sigma$  geodetic accuracy) Landsat TM imagery from the Nile river was compared with the two NDVI data sets within ArcGIS. The Nile river offers a useful registration target given the sharp boundary between the vegetated valley and delta, and the unvegetated desert





**Figure 2.** Absolute difference in average monthly NDVI between GIMMS and FASIR data sets for January, April, July, and October. The length of record in each case was 1983–1998.

surroundings. Comparing the boundaries of the vegetated zone mapped from the Landsat suggests that the GIMMS data set is correctly registered, while the FASIR data is misregistered (to the southwest).

#### 4. Comparisons of GIMMS, FASIR NDVI Absolute Values

[22] As noted above, the processing approaches of the FASIR and GIMMS NDVI data sets are quite different in philosophy and approach. In comparing absolute NDVI values between the products it is useful to consider how these processing choices have affected the NDVI measurement.

##### 4.1. Calibration and Scaling

[23] Although both FASIR and GIMMS used desert calibration targets, they used slightly different calibration methodologies. The GIMMS calibration approach results in a small offset (bias) of NDVI by  $\sim 0.05$  (S. Los, personal

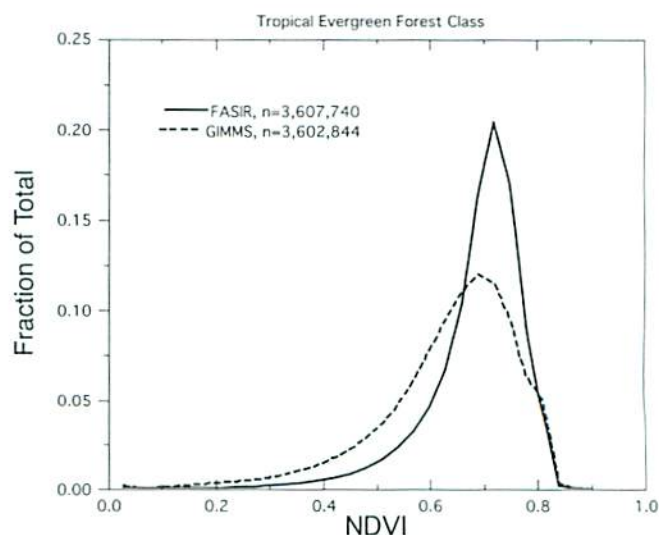
communication, 2005). In addition, the GIMMS applies a scaling factor to match the NDVI derived from the SPOT Vegetation sensor data during the 1997–2003 period of observational overlap. The much narrower band pass of SPOT Vegetation compared to AVHRR was designed to avoid atmospheric water vapor absorption in the NIR band, hence SPOT NDVI will be larger than that calculated from AVHRR; hence scaling the GIMMS with the SPOT product increases the GIMMS NDVI amplitude.

##### 4.2. Atmospheric Correction

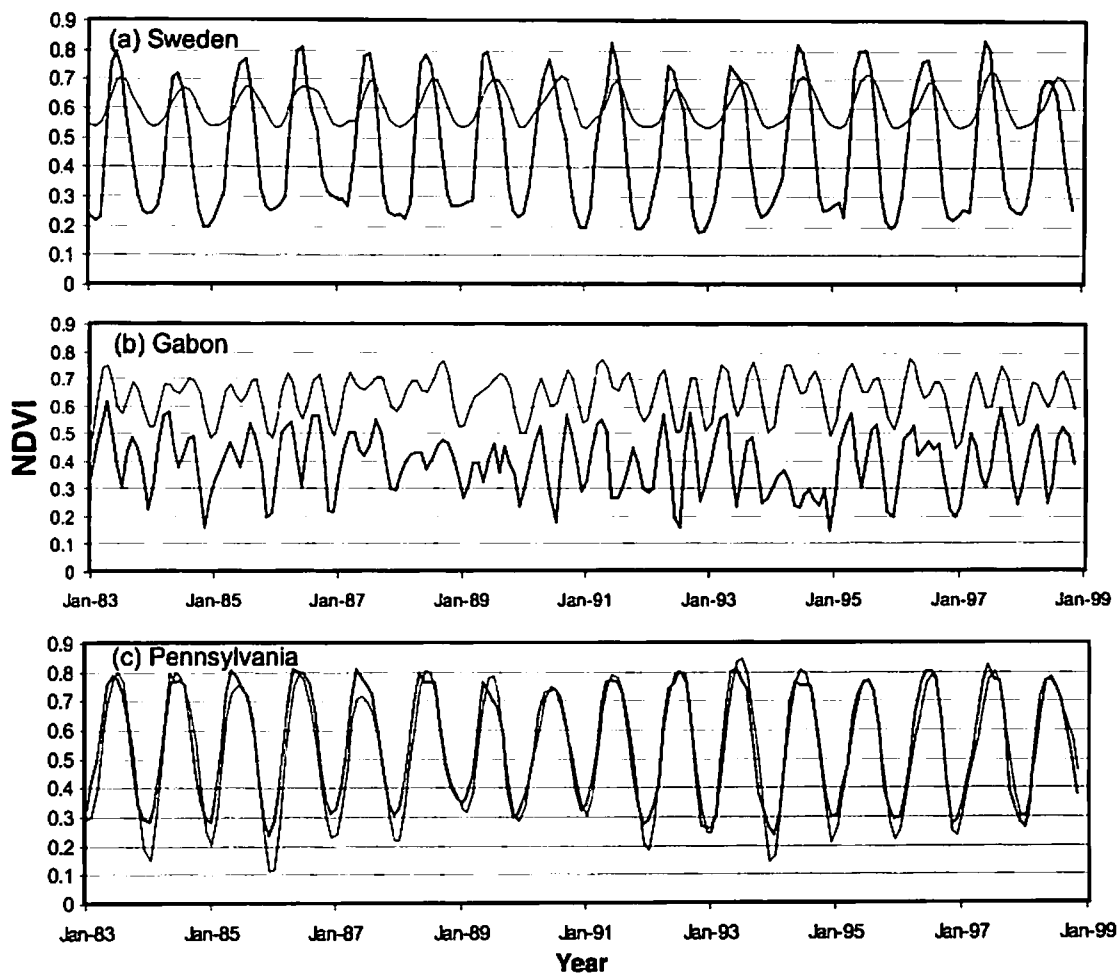
[24] GIMMS and FASIR NDVI represent fundamentally different measurements of the land surface. Although NDVI is commonly defined in terms of surface or top of atmosphere reflectance, neither the GIMMS nor FASIR records implemented a full atmospheric correction. GIMMS was corrected only for stratospheric aerosols. FASIR NDVI included ozone and Rayleigh corrections, but not tropospheric aerosols or water vapor. In general, the Rayleigh scattering correction implemented by FASIR should result in lower reflectance in the AVHRR visible band for dark vegetated targets, and hence higher NDVI values compared to the TOA GIMMS product.

##### 4.3. Surface BRDF Correction

[25] The FASIR data have been BRDF corrected to nadir look angle, and 60 degree solar elevation [Los *et al.*, 2005] whereas GIMMS NDVI values were measured at a range of view angles selected by compositing. The BRDF adjustment may increase NDVI calculated from TOA or surface reflectance, particularly in conditions of high aerosol optical thickness [Los *et al.*, 2005]. The effects of the BRDF correction are more pronounced in forests (owing to multiple scattering of near-infrared radiation in the canopy), and at high latitudes in the winter when the observed solar elevation is low.



**Figure 3.** Distribution of NDVI values for tropical evergreen forest for all months in the 1982–1998 period, derived from ISLSCP II 0.25 degree EDC Landcover classification and the 0.25 degree FASIR and GIMMS data sets.



**Figure 4.** Time series of NDVI values from FASIR (thin black line) and GIMMS (thick grey line) for (a) Sweden, (b) Gabon, and (c) Pennsylvania.

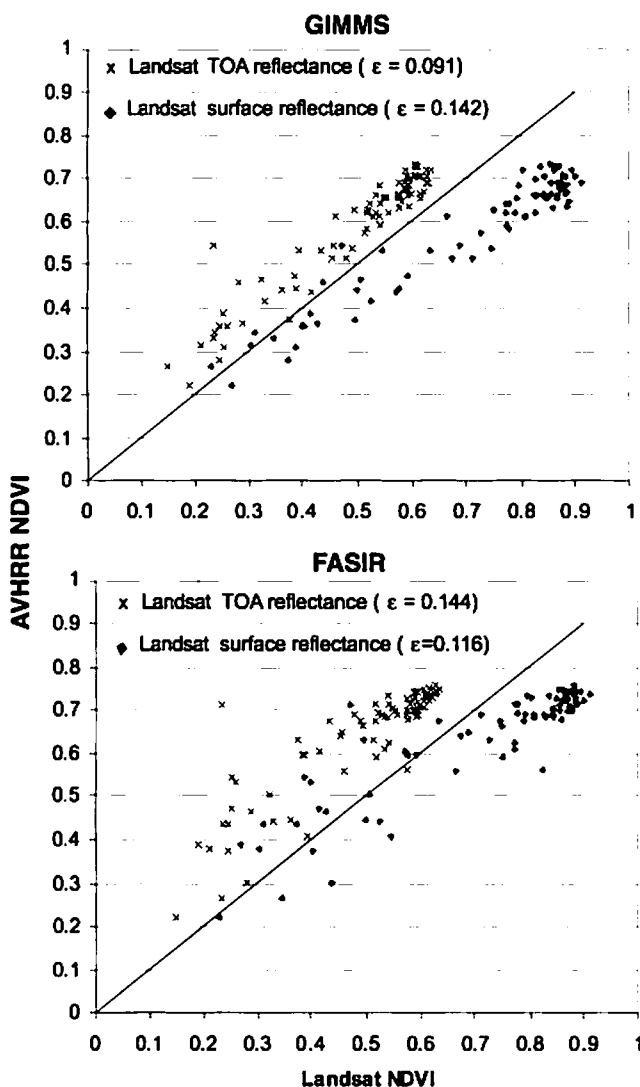
#### 4.4. Filtering and Interpolation

[26] As noted above, the FASIR used filtering and interpolation methods not employed in GIMMS. FASIR replaced NDVI values in the tropics with the maximum of a  $3 \times 3$  pixel moving window to actively suppress cloud contamination, thus tending to raise NDVI values. In the Boreal forests (northern conifers) FASIR replaced low winter NDVI values with the median October value to avoid snow contamination, thus producing a higher winter NDVI in this biome. Finally, the FASIR Fourier filtering will tend to remove individual outliers, whether very high or very low.

[27] Given these processing differences, it is not surprising that absolute NDVI values of the two data sets show strong and consistent mismatches. The differences between the FASIR and GIMMS monthly average NDVI are shown in Figure 2. Slightly over half of the globe's land area shows good agreement ( $<0.05$  absolute NDVI difference) between the two data sets. These areas tend to be those with limited NDVI values (e.g., hyperarid regions in North Africa and Arabia) or range (Western Amazon and Central Africa). In general, FASIR NDVI is considerably higher in the cloudy tropics (e.g., West Africa, Guyana, Myanmar, New Guinea), reflecting the effect of the FASIR  $3 \times 3$  maximum NDVI filter. This filter was implemented to reduce cloud contam-

ination but will also result in higher NDVI in clear areas. As a result, NDVI values in these areas range from 0.5–0.7 in the FASIR data set, but just 0.3–0.5 in the GIMMS data set (Figures 3 and 4b). On the basis of comparisons with Landsat TOA and surface reflectance imagery, NDVI values for tropical forests should be considerably higher than those represented in the GIMMS record, despite the lack of atmospheric correction in GIMMS. Though the maximum NDVI values for evergreen tropical forest type for all months of the data record are similar between the two data sets, FASIR has more observations at high NDVI values while GIMMS shows a broader distribution of NDVI values (Figure 3).

[28] Except for the midsummer, northern temperate and boreal forests tend to exhibit considerably higher NDVI in FASIR than GIMMS (Figures 2 and 4a). This can be explained through three contributing factors. First, the Rayleigh and ozone corrections applied to the FASIR data tend to increase NDVI, by reducing path radiance in the AVHRR visible band. Second, the FASIR BRDF correction tends to increase NDVI in Winter, Fall, and Spring. Finally, the filtering applied to the FASIR record replicates the October NDVI value in boreal regions during the winter to eliminate snow contamination, resulting in higher values compared to GIMMS. During the midsummer the situation



**Figure 5.** Comparison of Landsat-derived NDVI values (averaged over  $0.25^\circ$  ISLSCP Initiative II grid) with (top) GIMMS and (bottom) FASIR values, for Montana/Idaho region. Landsat data are from WRS-2 path 42, rows 27–28 acquired 10 September 1990. GIMMS and FASIR data are from September 1990 maps. Landsat NDVI calculated from both top-of-atmosphere reflectance (crosses) and atmospherically corrected surface reflectance (diamonds) are shown.

reverses, with GIMMS NDVI data trending significantly higher at latitudes north of  $\sim 50$  degrees N. At  $60^\circ\text{N}$ , the GIMMS NDVI values are up to 0.20 greater than FASIR. Since FASIR is not corrected for water vapor effects in band 2, this probably reflects the postcalibration scaling of GIMMS to the higher TOA values of SPOT Vegetation NDVI (its IR band largely avoids the water vapor absorption line), which tends to amplify the NDVI difference over the seasonal NDVI cycle.

[29] Southern hemisphere grasslands, shrublands, and semiarid environments tend to show higher values in the GIMMS data set than in FASIR, at all times of year. For example, values in Western Australia and Queensland are up to 0.20 higher in the GIMMS record, and values in the Argentine Pampas are up to 0.16 higher. Results of *Los et*

*al.* [2005] suggest that BRf correction for grasslands (in North America) tend to reduce FASIR NDVI values during peak greenness, which could explain some of the offset. However, it is not clear why this difference should persist throughout the year.

[30] Despite these differences, in many parts of the world the NDVI values in the two data sets do agree. For example in North Africa (where both data sets were calibrated prior to atmospheric correction) NDVI differences are typically less than 0.03. In temperate regions of the Northern Hemisphere (e.g., central and eastern United States, central Europe, southern Russia, China) differences are usually less than 0.05 (Figures 2 and 4c).

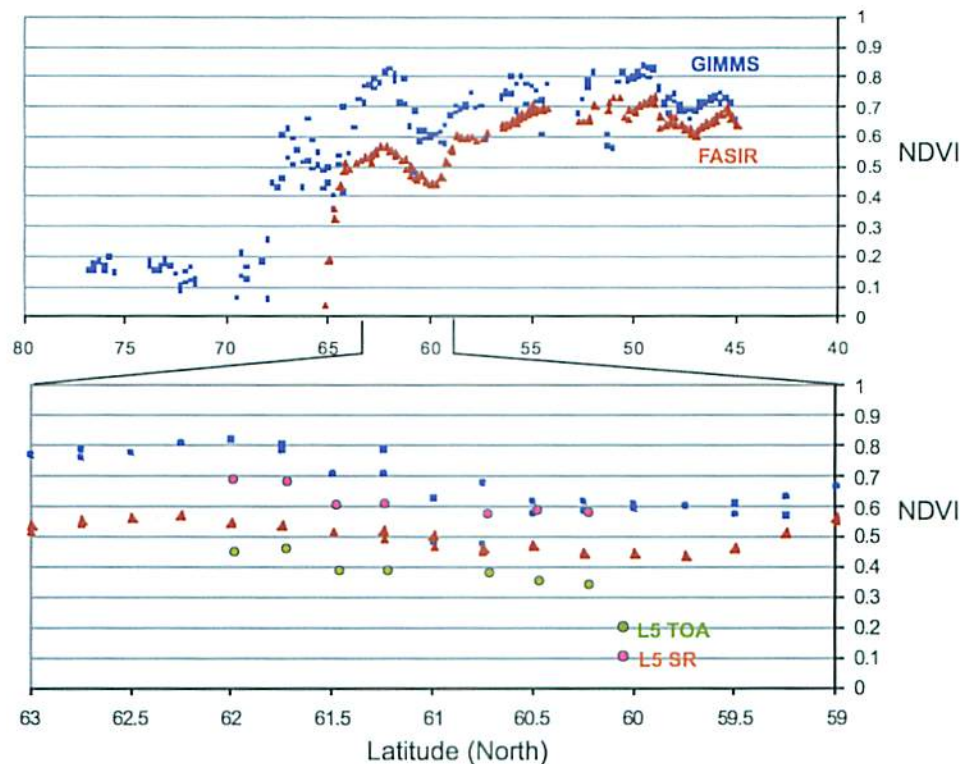
[31] To investigate differences between the data sets more fully, we have processed several Landsat TM scenes from the 1980s and 1990s to both surface and TOA reflectance using the 6S radiative transfer model [Vermote *et al.*, 1997; Masek *et al.*, 2006]. Because the Landsat data represent instantaneous, cloud-free views, NDVI retrieved from Landsat avoids errors associated with compositing and cloud clearing. Landsat NDVI was calculated from red and near-infrared (band 3 and 4) reflectance values, and then averaged to the  $0.25^\circ$  resolution of the ISLSCP Initiative II grid. The Landsat TM band passes are substantially narrower compared to the AVHRR band pass, and one could in principle adjust the Landsat measurements to “match” the AVHRR NDVI. Here we have not elected to do so, in part because the GIMMS data is already normalized to the narrow band pass of the SPOT Vegetation sensor.

[32] A transect across western Montana and Idaho includes a range of vegetation conditions at midlatitudes, including dense conifer forest, sparse woodlands, semiarid shrublands, and agriculture. Two successive Landsat acquisitions from 10 September 1990 provide 79 quarter degree observations for comparison with the ISLSCP data sets (Figure 5). The GIMMS record shows a strong correlation with the Landsat NDVI values. As expected, the GIMMS data (derived from AVHRR TOA reflectances) most closely matches the Landsat TOA NDVI, on the basis of the lower average absolute difference between values (Figure 5). The correlation with FASIR data over the same region is slightly weaker, with greater dispersion of low NDVI values, but a closer association with the Landsat surface reflectance NDVI.

[33] A second, north-south transect was analyzed to examine differences between FASIR and GIMMS in Boreal midsummer conditions (Figure 6). This transect (July 1991) includes data from  $98.75$ – $99.00$  degrees west, and extends from  $80$  degrees north (Northwest Territories, Canada) to  $45$  degrees north (South Dakota, United States). As noted above, GIMMS and FASIR values appear consistent in the midcontinent but diverge in Boreal regions, with GIMMS having consistently higher values in midsummer. A set of Landsat NDVI values from July 1991 are also plotted (path 35, rows 17–18, acquired July 26, 1991). In contrast to the Montana example, the GIMMS data in this case show far higher values than that found for Landsat TOA NDVI, and are even higher than Landsat surface reflectance NDVI.

[34] It is clear from these analyses and comparisons between the FASIR and GIMMS NDVI records, that neither can be used in an absolute sense, since neither completely





**Figure 6.** (top) North-south transect near 100 degrees W longitude, from Northwest Territories, Canada, to South Dakota, United States. Both GIMMS (blue square) and FASIR (red triangle) NDVI data are shown. (bottom) Expanded version GIMMS (blue square) and FASIR (red triangle) extending from 59 to 63 degrees North, with averaged Landsat NDVI observations superposed, based on top-of-atmosphere reflectance (green) and surface reflectance (red).

corrects for spatial or seasonal atmospheric variations, particularly tropospheric aerosols and water vapor, which themselves can have a strong seasonal signal. Alternatives to using the absolute values are discussed and evaluated further in the next section.

## 5. Interannual Variability of NDVI Anomalies

### 5.1. Interannual NDVI Patterns

[35] Both FASIR and GIMMS NDVI data records exhibit variability resulting from calibration, viewing geometry and atmospheric conditions that alter the land surface reflectance signal inferred from the satellite observations. Recognizing this some analyses of long-term trends in vegetation greenness [Myneni *et al.*, 1997; Zhou *et al.*, 2003] or drought response [Anyamba *et al.*, 2001; Ji and Peters, 2003] rely on the use of NDVI anomalies, i.e., the deviation of NDVI measurements in a record from the average monthly or annual NDVI computed from that record. Ideally, interannual anomalies should match across the FASIR and GIMMS records, even if absolute NDVI values differ.

[36] In Figure 7 we show the NDVI anomalies for continental regions of the globe. Not only are there differences in the absolute NDVI records shown in Figures 2 through 6, there are also significant differences in the two anomaly records that are of similar magnitude to the annual anomalies themselves. The anomalies in the FASIR record

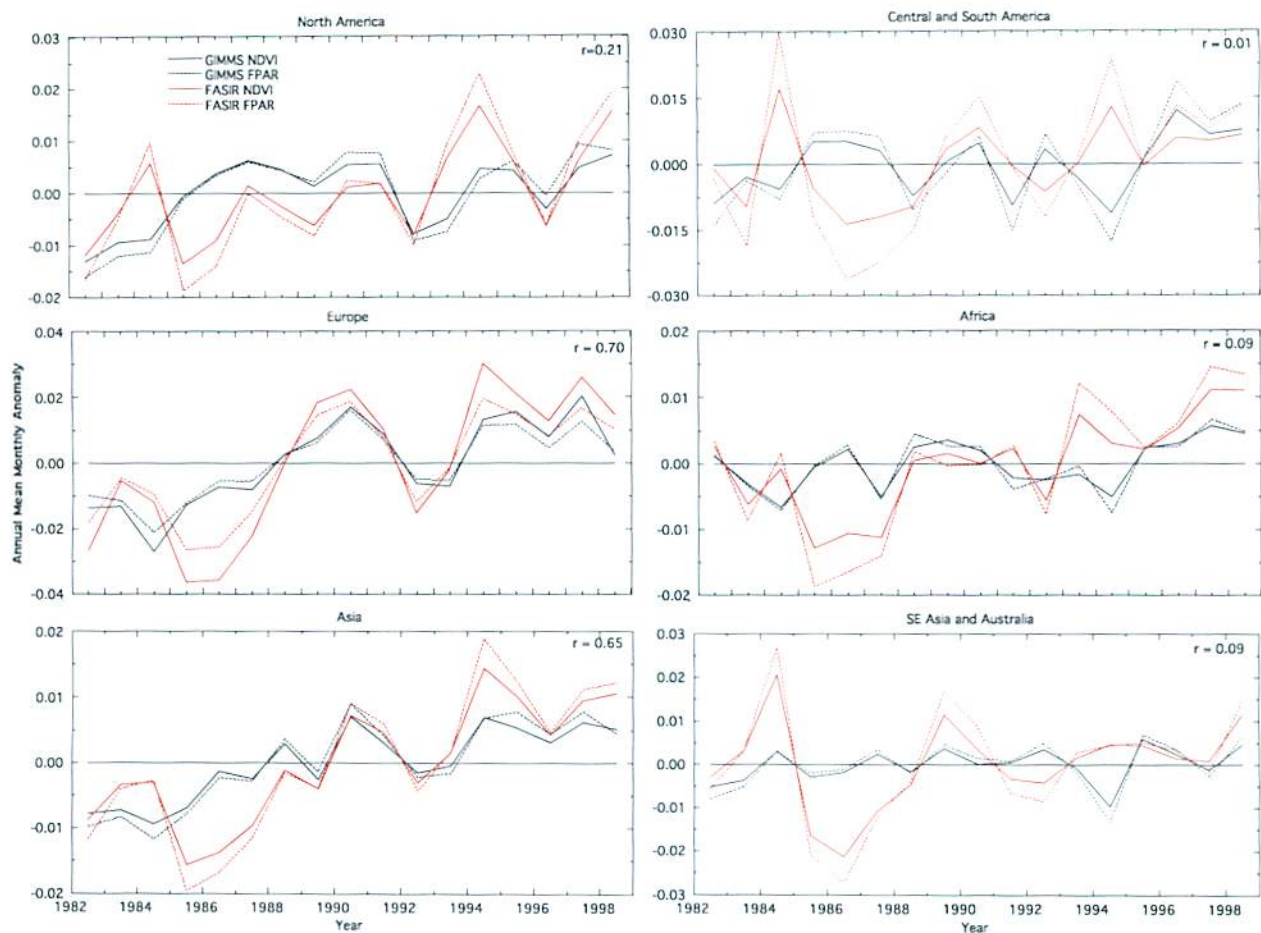
and the difference between FASIR and GIMMS are especially large in the period 1984–1986 and in 1994 for all continental regions. In addition, FASIR NDVI data tend to show much stronger negative anomalies compared to GIMMS for 1992. This difference is possibly a result of different stratospheric aerosol optical thickness data sets used in GIMMS and FASIR. Both data sets show positive trends for North America, Europe and Asia, while only FASIR shows a positive trend for Africa and both data sets show no trends for Central/South America and SE Asia/Australia. The regional correlation coefficients for the NDVI anomalies are shown in the figure. Correlations are generally poor (and not statistically significant) except for Europe and Asia (>99%). The correlation between the anomalies for boreal North America is high but poor for temperate North American (not shown) producing a low correlation for the continent.

[37] The 1994 disparity is largely a result of the different FASIR, GIMMS approaches dealing with the NOAA 13 failure between NOAA-11 and NOAA-14 as discussed in section 2.1. The anomalies are more similar during the rest of the time series for northern latitude continental regions.

### 5.2. Evaluation and Comparison of the Initiative II NDVI Series in the Context of Carbon Cycle Modeling

[38] To determine which NDVI time series might be more realistic, we use both the GIMMS and FASIR NDVI time





**Figure 7.** Time series plots showing annual GIMMS NDVI and fPAR anomalies (black solid, dashed lines, respectively) and FASIR NDVI and fPAR anomalies (red solid, dashed lines, respectively) for continental regions of the globe. Correlation coefficients ( $r$ ) are given for NDVI anomalies.

series as input to a global carbon cycle model, CASA, that simulates the surface-atmosphere exchange of  $\text{CO}_2$  [Potter *et al.*, 1993; Randerson *et al.*, 1996] (and described briefly in the next section) to explore these implications. CASA uses fPAR, derived from the NDVI record (as described in the next section) and surface incident solar irradiance to compute photosynthetically active radiation absorbed by the canopy, and uses temperature and precipitation data to compute light use efficiency and heterotrophic respiration. Using CASA we predicted global interannual variation in net land surface-atmosphere carbon exchange and compare these values to those estimated from an atmospheric  $\text{CO}_2$  flux inversion study.

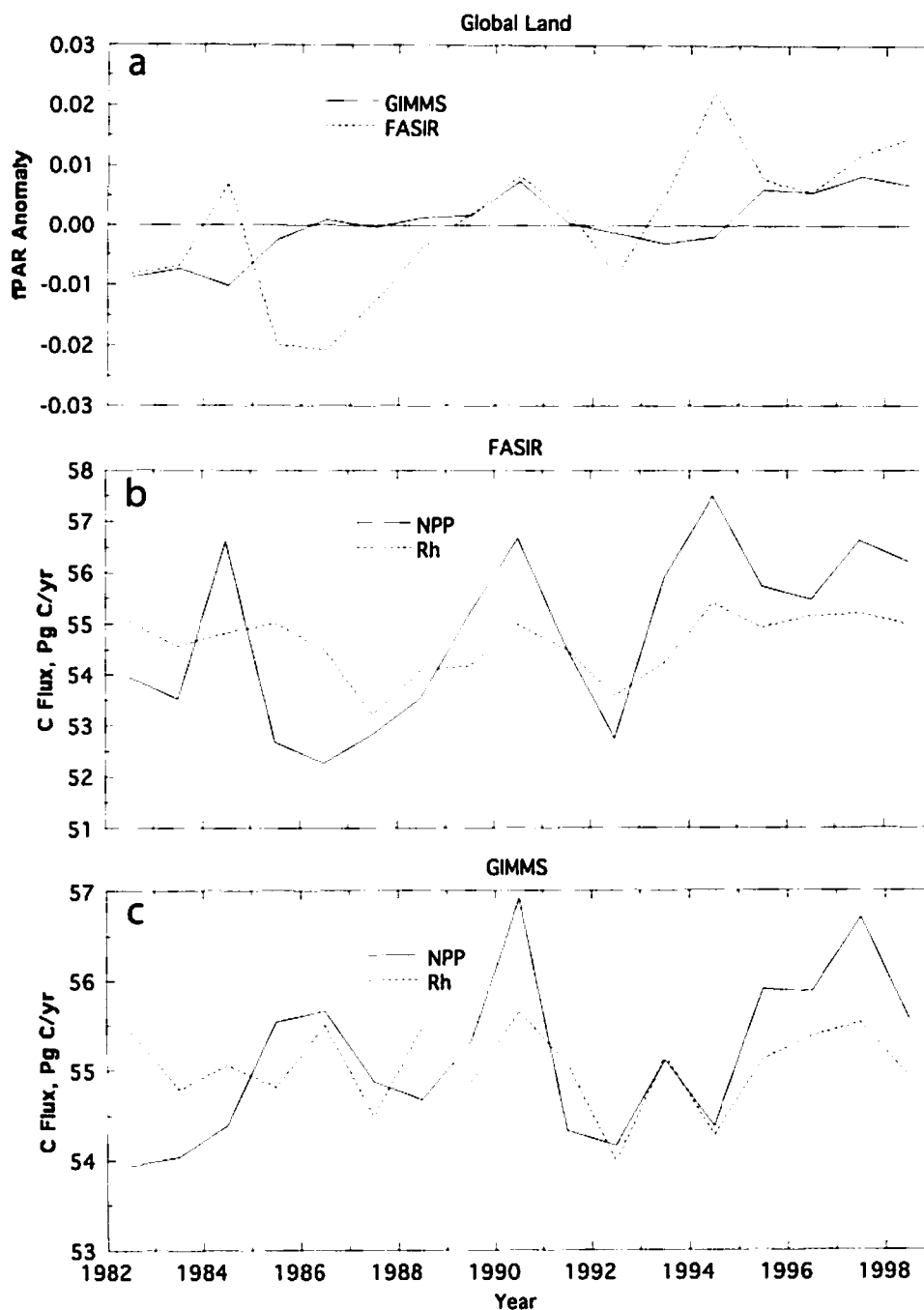
### 5.2.1. Approach

[39] CASA uses the light use efficiency approach [Monteith, 1997] to characterize climate-induced deviations in productivity capacity. Monthly global ( $1^\circ \times 1^\circ$ ) net primary productivity (NPP) is calculated as the product of incident photosynthetically active radiation (PAR), and its fraction absorbed by the canopy (fPAR) and a light use efficiency parameter, the latter dependent on temperature and precipitation (for temperature, precipitation and solar irradiance data, see Hansen *et al.* [1999], Adler *et al.* [2003], and Zhang *et al.* [2004], respectively). Heterotrophic respiration is computed using temperature and precipitation data

and also depends on modeled NPP. To convert NDVI into fPAR we implemented the algorithm reported by Los *et al.* [2000] in which NDVI is used to scale fPAR from some minimum value ( $<0.05$ ), corresponding to the minimum NDVI observed for vegetation, to a maximum (0.95) representing the maximum (e.g., 98th percentile) observed NDVI, hence full canopy closure. The scaling is dependent on maximum and minimum NDVI observed for various vegetation types (e.g., Figure 3). The derivation of fPAR “normalizes” the NDVI record allowing the FASIR and GIMMS anomaly records to be evaluated in a consistent way.

[40] Since the dynamic range of NDVI is higher for the FASIR than for GIMMS data sets the maximum/minimum scalars used for each data set reflected these differences, but were in fact small and showed consistent relationships among the vegetation types. Figure 7 shows the NDVI and fPAR annual anomalies for continental regions from 1982–1998. Each continental region consists of a number of vegetation types. The sign of the anomalies in NDVI correspond to those for fPAR, only the dynamic range of the fPAR is larger since NDVI ranges between  $\sim 0$ –0.8 while fPAR ranges from  $\sim 0.03$ –0.95.

[41] Using the calculated GIMMS and FASIR fPAR records and climate data, we ran CASA to simulate NPP, heterotrophic respiration (Rh) and the net flux to the

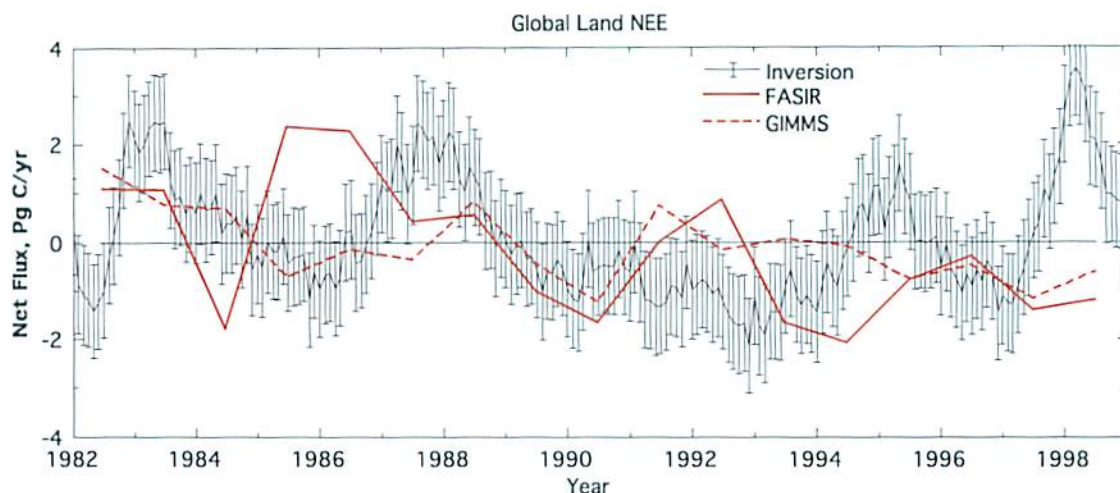


**Figure 8.** Global fPAR anomalies and carbon fluxes from CASA: (a) annual fPAR anomalies for the globe calculated from GIMMS (solid line) and FASIR (dashed line) NDVI records; (b) global net primary production (NPP) and heterotrophic respiration (Rh) fluxes calculated from observed climate record, using the CASA biogeochemical model driven with FASIR NDVI/fPAR anomalies; and (c) same as Figure 8b, except driven with GIMMS NDVI/fPAR anomalies.

atmosphere (Rh-NPP or NEE). We have aggregated the results to represent global land area. CASA was run to equilibrium (annual NEE = 0 for each grid cell) using the mean climate and fPAR for the 1982–1998 period and then the interannual simulation was started and run forward from January 1982 through December 1998 using observed time series of fPAR and climate inputs. Figure 8a shows the annual fPAR anomalies for the globe and Figures 8b and 8c the NPP anomalies for FASIR and GIMMS respectively.

The larger range of the FASIR NDVI, hence fPAR, is reflected in the range of NPP. NPP is largely driven by fPAR in the CASA model with weaker effects from lower variability in incident solar irradiance, and lower sensitivity to temperature and precipitation. A 0.01 change in the annual fPAR anomaly translates to about 1.5 Pg C/yr NPP anomaly which means that a 0.01 change in the original NDVI corresponds to about 2 Pg C/yr change in NPP. The largest year-to-year NPP anomalies using FASIR as the input





**Figure 9.** Global net ecosystem exchange (NEE) simulated by CASA using FASIR and GIMMS records plotted with observed global fluxes derived from atmospheric carbon flux inversion of Rödenbeck *et al.* [2003]. The mean land carbon sink of 1.5 PgC/yr for this period was added to the inversion results to emphasize the comparison of interannual variability between the inversion and CASA simulations. Positive NEE corresponds to carbon transfer from the land to the atmosphere (e.g., land CO<sub>2</sub> source); NEE is calculated as Rh-NPP.

to CASA were about 3 Pg C/yr and generally <2 Pg C/yr using GIMMS. Also plotted in Figures 8b and 8c are Rh anomalies. In CASA Rh is determined by temperature and precipitation controls on 9 carbon pools with varying turnover times. However, since the same climate was used for both CASA runs the differences in Rh are a result of NDVI/fPAR-driven NPP variability. A large NPP anomaly will result in a significant anomaly in the delivery of carbon to heterotrophic carbon pools some of which have turnover times of less than 1 year (e.g., leaf and root litter).

#### 5.2.2. Simulation Results and Comparisons

[42] Figure 9 shows the difference between Rh and NPP or NEE, the flux of carbon between the surface and atmosphere. When it is positive it represents a source to the atmosphere and negative a sink. Included in the figure are the results of an atmospheric inversion study [Rödenbeck *et al.*, 2003] where land sources and sinks were inferred from a model of atmospheric transport and surface measurements of atmospheric CO<sub>2</sub> partial pressure from a global sampling network. The error bars represent 1 standard deviation about the mean. The fossil fuel source has been “presubtracted” from the inversion results. The original inversion results showed that the global land surface to be on average for this period a carbon sink of about 1.5 Pg C per year for. In Figure 9 this constant 1.5 Pg C sink was added to the inversion results since the CASA simulations, were initialized to produce an equilibrium NEE (i.e., NEE = 0) at the 1982 start of the evaluation period. CASA simulations show a trend of increasing sink over the time period that is not evident in the inversion.

[43] The inversion produces maximum year-to-year variability of about 2–3 Pg C/yr. The Transcom inversion study [Baker *et al.*, 2006; K. Gurney, personal communication, 2005] inferred maximum variability of around 3 Pg C/yr as well, however, Bousquet *et al.* [2000] reported maximums of around 1.5 Pg C/yr. The CASA simulations

produced maximum variations of about 1 and 3 Pg C/yr for GIMMS and FASIR respectively. However, the CASA model results do not produce the same timing in the anomalies as the inversion. In fact for some periods the anomalies are of opposite sign. For instance, the FASIR driven simulations produced strong source anomalies in 1985–1986 and 1992 while the inversions show these to be periods of neutral or significant sink behavior. With the exception of 1983 neither of the NDVI driven simulations produced the strong source signals seen in the inversion during the El Niños of 1987, 1994–1995 and 1998.

#### 5.2.3. Discussion

[44] Our analysis in the previous section indicates little correspondence between either the FASIR NDVI or the GIMMS NDVI anomaly records and the global land source/sink variability inferred from atmospheric inversion; nor do the CASA simulations of NEE when driven by either FASIR or GIMMS NDVI. Assuming that the climate data used to drive the model and the inversion results for the global land are correct, this lack of agreement could be caused by errors in the NDVI records and/or errors and omissions in the modeling of NEE.

[45] In general, NEE can be described by the following equation:

$$NEE = Rh(T, \theta) + E - \int_t \{ \varepsilon(T, \theta) * fPAR * PAR \} dt.$$

The product of the last three terms (in brackets) describes NPP: PAR and fPAR defined in 5.2.1 (fPAR is derived observationally from NDVI) records the relative amount of radiation absorbed by the canopy, and  $\varepsilon$  is the light-use efficiency of vegetation as a function of temperature (T) and soil moisture ( $\theta$ ). The first two terms describe offsetting transfers of carbon to the atmosphere. Rh describes heterotrophic respiration of biomass debris (leaves, roots,

soil organic matter), and E corresponds to nonphysiological carbon emissions mostly due to biomass burning. The individual contributions of NPP, Rh or fire to the atmospheric CO<sub>2</sub> signal can be of opposite signs, and compensating, hence only the aggregate effect, not the individual contributions, is reflected in the atmospheric CO<sub>2</sub> inversion analyses. The version of CASA used here simulates only the first two processes (but see *van der Werf et al.* [2004]).

[46] At a minimum, it is clear that the Initiative II NDVI anomaly records alone cannot explain the global atmospheric CO<sub>2</sub> variability derived from the inversion analyses. In CASA, the observed variability in NDVI translates directly to variations in fPAR, which in turn cause large swings in NPP. Since variability in Rh is influenced by short-term pools of accumulated carbon as well as climate Rh will also respond in parallel with NPP to fPAR.

[47] One possibility is that the NDVI records are correct, and other factors not accounted for in CASA compensate for the effects of fPAR variability on global NEE. First, fire emissions may dominate variability of atmospheric CO<sub>2</sub> fluxes. A number of researchers have argued that fires associated with El Niño events contribute significantly to increased atmospheric CO<sub>2</sub> growth rate during these periods [*Langenfelds et al.*, 2002; *Schimel and Baker*, 2002; *van der Werf et al.*, 2004], and *van der Werf et al.* [2004] estimated that fires contributed a 2.1 PgC/yr growth rate anomaly during the 1997–1998 El Niño. If these fire emissions were added to the CASA simulations shown in Figure 9, most of the inversion fluxes would be captured for that period.

[48] In addition to fire, others have suggested that climate controls on light-use efficiency or heterotrophic respiration may act as a primary forcing on CO<sub>2</sub> fluxes. *Nemani et al.* [2003] showed a negative correlation between global atmospheric growth rate in CO<sub>2</sub> and modeled NPP during the large El Niño event of 1997–1998. NDVI tended to show positive anomalies for that period (see Figures 8 and 9), but in their model, physiological drought reduced NPP. Other analyses have shown similar drought limitations on productivity and/or temperature stimulation of respiration [*Tian et al.*, 1998; *Zeng et al.*, 2005]. *Lucht et al.* [2002] showed that negative anomalies in modeled LAI in the boreal latitudes in 1992 following the Mt. Pinatubo eruption corresponded to negative anomalies in NDVI yet their model produced a carbon sink in this region, which was in agreement with atmospheric inversion results. While in their model decreasing LAI did result in a decrease in modeled NPP, the cooler summer temperatures inhibited modeled Rh even more resulting in the net sink. It should be noted that the version of CASA used in this analysis includes temperature and moisture scalars designed to simulate climate effects on vegetation physiology and heterotrophic respiration. However, it is certainly possible that CASA is currently underestimating these effects.

[49] Finally, it is also possible that the NDVI anomalies themselves are wrong. Large interannual anomalies in fPAR, as suggested by the Initiative II data sets, imply significant anomalies in the seasonal duration of canopy LAI and/or its seasonal maximum, both of which would affect annual productivity anomalies [*Barr et al.*, 2004]. Unfortunately there are few in situ measurements of interannual variability of fPAR/LAI against which to assess

whether the fPAR variability inferred from the NDVI records is valid. *Barr et al.* [2004] reported strong links between interannual variability in LAI, GPP and NEE for a deciduous boreal forest (aspen), where LAI variations were driven primarily by spring temperature and by summer drought. The interannual variability in fPAR derived from their in situ-measured LAI were on the order of 0.03 and consistent with the anomalies derived from the Initiative II NDVI records. Seasonal and interannual variability of LAI in herbaceous vegetation types especially in response to water availability is well documented and these have been linked to productivity and NEE variability [e.g., *Flanagan et al.*, 2002], but contributions of nonforested ecosystems to interannual variations in carbon storage should be small because the biomass of these systems has short turnover times and most of the fixed carbon is returned to the atmosphere each year. We have not been able to find studies of interannual variability in fPAR/LAI measured in situ in evergreen forest types (e.g., boreal, tropical). These ecosystems are critical to global carbon cycling, but also problematic for NDVI estimation because of the short growing season for Boreal forests and persistent cloud cover in the tropics.

[50] We cannot conclude that our NDVI records, either FASIR or GIMMS, are incompatible with the inversion analyses. However, we can caution researchers that using observational NDVI alone to estimate carbon fluxes is unlikely to match atmospheric measurements. Either the NDVI records themselves are in error, or the contribution of fPAR variability to the interannual variation in NEE is small in comparison to climate-induced interannual variations in light use efficiency, Rh, or fire emissions. Our analyses of the Initiative II NDVI time series also indicate the need for continued improvement in the longer-term satellite derived surface vegetation records including in situ validation, along with improved sampling of atmospheric CO<sub>2</sub> variations and improved transport models supporting inversion analyses.

## 6. Conclusions

[51] In this paper we have compared the FASIR and GIMMS AVHRR NDVI data sets included in ISLSCP Initiative II [*Tucker et al.*, 2005]. Although both data sets start by compositing the AVHRR GAC record, each processing stream is unique. The FASIR data set attempts a series of physically based corrections to the original AVHRR data set, including atmospheric correction (Rayleigh and ozone), BRDF correction to standard sun-target-sensor geometry, and time-space filtering to remove snow and cloud contamination. The GIMMS data set adopts a more empirical approach, eschewing full atmospheric correction, and relying on empirical-mode decomposition to remove outliers, and normalizing NDVI values to the later SPOT Vegetation record.

[52] Our principal conclusions are as follows.

[53] 1. Absolute NDVI values differ considerably between the two data sets, both seasonally and geographically. Most of these differences are persistent across the 12 year record considered here, and can be understood as the result of specific processing choices. In particular, NDVI values in the humid tropics are systematically lower in the GIMMS record, while nonsummer values in northern latitudes are



systematically higher in the FASIR record. NDVI values in midlatitude deserts and agricultural areas generally agree between the two records.

[54] 2. As opposed to absolute values, NDVI anomalies (monthly, annual) generally agree between the two data sets for any given location. Agreement is better for later parts of the record (e.g., 1992–1998). Two notable exceptions are 1994 and 1984–1986. The former corresponds to the coverage gap between NOAA-11 and NOAA-14, during which the two records rely on different data sources. The origin of the 1984–1986 anomaly disagreement is not clear.

[55] 3. The fPAR absorbed by vegetation calculated from each record closely follows the NDVI anomalies, and shows large interannual variations. Using a standard light-use efficiency model for calculating net primary productivity (NPP), we find that global anomalies of  $\sim 0.010$  in NDVI correspond to anomalies of  $\sim 0.013$  in fPAR, and an NPP anomaly of  $\sim 2.0$  Pg C.

[56] 4. Using the NDVI records to drive the CASA biogeochemical model results in global carbon fluxes that show little or no correlation with sources and sinks derived from inverting atmospheric  $\text{CO}_2$  measurements [Rödenbeck et al., 2003]. Although the fPAR-driven fluxes show year-to-year variations comparable in magnitude to the inversion results, neither NDVI record produces a temporal pattern that matches the inversion results.

[57] Users should not rely on absolute NDVI values, particularly in the context of physical models. Although it would be difficult to find a simpler spectral index than NDVI, the processing chain required to convert AVHRR observations to a consistent surface NDVI record is complex and at this date, incomplete. As a result the FASIR NDVI and GIMMS NDVI records must be treated as different measurements. In part this situation reflects the inherent limitations of the AVHRR sensors, which were never intended for long-term terrestrial research. It can be hoped that future NDVI data sets, using carefully calibrated, screened, and atmospherically corrected data from MODIS and VIIRS, will result in a more consistent record of Earth's vegetation. Even with the current limitations of the AVHRR observational record, monthly and annual NDVI anomalies show reasonable agreement between the FASIR and GIMMS, suggesting that users should try to use normalized (anomaly) NDVI values rather than absolute values.

[58] The observation that neither NDVI data set can be correlated with observed carbon fluxes has significant implications for carbon modeling. Taken at face value, the CASA model results, driven by the NDVI anomaly record, predict large variations in vegetation fPAR and LAI that are not driving variability in global carbon fluxes. It is of course possible that CASA and/or the NDVI anomaly records are wrong. Other factors that may not be adequately represented in this version of CASA, such as the physiologic response to climate stress or fire emissions, could be compensating for real variations in NDVI/fPAR. Alternatively, large anomalies in NDVI/fPAR may not be real. Indeed, it is likely that modeling ecosystem productivity using a simple "climatological" mean fPAR record, supplemented with ENSO-driven fire emissions would produce a reasonable match to the inversion results shown in Figure 9. The results do indicate that researchers should be cautious when using

NDVI-driven models to predict net ecosystem exchange without independent comparison with observed fluxes.

[59] **Acknowledgments.** This study was supported by NASA's Hydrology and Terrestrial Ecology Programs as well as by a NASA REASON Cooperative Agreement. We thank Christian Rödenbeck for providing us with the atmospheric inversion data. Fred Huemmrich is thanked for analysis of the effects AVHRR and Landsat band passes.

## References

- Adler, R. F., et al. (2003), The Version 2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present), *J. Hydrometeorol.*, **4**, 1147–1167.
- Anyamba, A., C. J. Tucker, and J. R. Eastman (2001), NDVI anomaly patterns over Africa during the 1997/98 ENSO warm event, 2001, *Int. J. Remote Sens.*, **22**, 1847–1859.
- Asrar, G., M. Fuchs, E. T. Kanemasu, and J. L. Hatfield (1984), Estimating absorbed photosynthetic radiation and leaf area index from spectral reflectance in wheat, *J. Agron.*, **76**, 300–306.
- Baker, D. F., et al. (2006), TransCom 3 inversion intercomparison: Impact of transport model errors on the interannual variability of regional  $\text{CO}_2$  fluxes, 1988–2003, *Global Biogeochem. Cycles*, **20**, GB1002, doi:10.1029/2004GB002439.
- Barr, A. G., T. A. Black, E. H. Hogg, N. Kljun, M. Morgenstern, and Z. Nesic (2004), Inter-annual variability in the leaf area index of a boreal aspen-hazelnut forest in relation to net ecosystem production, *Agric. For. Meteorol.*, **126**, 237–255.
- Birth, G. S., and G. R. McVey (1968), Measuring the color of growing turf with a reflectance spectrometer, *Agron. J.*, **60**, 640–643.
- Bounoua, L., G. J. Collatz, S. O. Los, P. J. Sellers, D. A. Dazlich, C. J. Tucker, and D. A. Randall (2000), Sensitivity of climate to changes in NDVI, *J. Clim.*, **13**, 2277–2292.
- Bousquet, P., P. Peylin, P. Ciais, C. Le Quere, P. Friedlingstein, and P. P. Tans (2000), Regional changes in carbon dioxide fluxes of land and oceans since 1980, *Science*, **290**, 1342–1346.
- Dickinson, R. E., M. Shaikh, R. Bryant, and L. Graumlich (1998), Interactive canopies for a climate model, *J. Clim.*, **11**, 2823–2836.
- Dutton, E. (1994), Aerosol optical depth measurements from four NOAA/CMDL monitoring sites, in *Trends: A Compendium of Data on Global Change, ORNL/CDIAC-65*, pp. 484–494, Carbon Dioxide Inf. Anal. Cent., Oak Ridge Natl. Lab., Oak Ridge, Tenn.
- Flanagan, L. B., L. A. Wever, and P. J. Carlson (2002), Seasonal and interannual variation in carbon dioxide exchange and carbon balance in a northern temperate grassland, *Global Change Biol.*, **8**, 599–615.
- Goward, S. N., C. J. Tucker, and D. G. Dye (1985), North American vegetation patterns observed with the NOAA-7 Advanced Very High Resolution Radiometer, *Vegetatio*, **64**, 3–14.
- Goward, S. N., D. G. Dye, S. Turner, and J. Yang (1993), Objective assessment of the NOAA Global Vegetation Index data product, *Int. J. Remote Sens.*, **14**, 3365–3394.
- Goward, S. N., S. Turner, D. G. Dye, and S. Liang (1994), The University of Maryland improved Global Land Vegetation Index product, *Int. J. Remote Sens.*, **15**, 3365–3396.
- Gutman, G. G. (1991), Vegetation Indices from AVHRR: An update and future prospects, *Remote Sens. Environ.*, **35**, 121–136.
- Gutman, G. G. (1999), On the use of long-term global data of land reflectances and vegetation indices derived from the advanced very high resolution radiometer, *J. Geophys. Res.*, **104**(D6), 6241–6255.
- Hall, F. G., K. F. Huemmrich, S. J. Goetz, P. J. Sellers, and J. E. Nickeson (1992), Satellite remote-sensing of surface-energy balance: Success, failures, and unresolved issues in FIFE, *J. Geophys. Res.*, **97**(D17), 19,061–19,089.
- Hansen, J., R. Ruedy, J. Glascoe, and M. Sato (1999), GISS analysis of surface temperature change, *J. Geophys. Res.*, **104**(D24), 30,997–31,022.
- 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., Ser. A*, **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.
- 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**, 3347–3363.
- Ji, L., and A. J. Peters (2003), Assessing vegetation response to drought in the northern Great Plains using vegetation and drought indices, *Remote Sens. Environ.*, **87**, 85–98.
- Justice, C. O., J. R. G. Townshend, B. N. Holben, and C. J. Tucker (1985), Analysis of the phenology of global vegetation using meteorological satellite data, *Int. J. Remote Sens.*, **6**, 1271–1381.

- Kaminski, T., W. Knorr, P. J. Rayner, and M. Heimann (2002), Assimilating atmospheric data into a terrestrial biosphere model: A case study of the seasonal cycle, *Global Biogeochem. Cycles*, **16**(4), 1066, doi:10.1029/2001GB001463.
- Kang, H., Y. Xue, and G. J. Collatz (2006), Impact assessment of satellite-derived leaf area index datasets using a general circulation model, *J. Clim.*, in press.
- Langenfelds, R. L., R. J. Francey, B. C. Pak, L. P. Steele, J. Lloyd, C. M. Trudinger, and C. E. Allison (2002), Interannual growth rate variations of atmospheric CO<sub>2</sub> and its delta  $\delta^{13}\text{C}$ , H<sub>2</sub>, CH<sub>4</sub>, and CO between 1992 and 1999 linked to biomass burning (2002), *Global Biogeochem. Cycles*, **16**(3), 1048, doi:10.1029/2001GB001466.
- Los, S. O., C. O. Justice, and C. J. Tucker (1994), A global 1-degrees-by-1-degrees NDVI data set for climate studies derived from the GIMMS Continental NDVI data, *Int. J. Remote Sens.*, **15**, 3493–3518.
- Los, S. O., G. J. Collatz, P. J. Sellers, C. M. Malmstrom, N. H. Pollack, R. S. DeFries, L. Bounoua, M. T. Parriss, C. J. Tucker, and D. A. Dazlich (2000), A global 9-yr biophysical land surface dataset from NOAA AVHRR data, *J. Hydrometeorol.*, **1**, 183–199.
- Los, S. O., P. R. J. North, W. M. F. Grey, and M. J. Barnsley (2005), A method to convert AVHRR Normalized Difference Vegetation Index time series to a standard viewing and illumination geometry, *Remote Sens. Environ.*, **99**, 400–411.
- Lucht, W., I. C. Prentice, R. B. Myneni, S. Sitch, P. Friedlingstein, W. Cramer, P. Bousquet, W. Buermann, and B. Smith (2002), Climatic control of the high-latitude vegetation greening trend and Pinatubo effect, *Science*, **296**, 1687–1689.
- Masek, J. G., E. F. Vermote, N. Saleous, R. Wolfe, F. G. Hall, F. Huemmrich, F. Gao, J. Kutler, and T. K. Lim (2006), A Landsat surface reflectance data set for North America, 1990–2000, *Geosci. Remote Sens. Lett.*, **3**, 68–72.
- Monteith, J. L. (1997), Climate and efficiency of crop production in Britain, *Philos. Trans. R. Soc., Ser. B*, **281**, 271–294.
- Myneni, R. B., C. D. Keeling, C. J. Tucker, G. Asrar, and R. R. Nemani (1997), Increased plant growth in the northern high latitudes from 1981–1991, *Nature*, **386**, 698–701.
- Nemani, R. R., and S. W. Running (1989), Estimation of resistance to evapotranspiration from NDVI and Thermal-IR AVHRR data, *J. Appl. Meteorol.*, **28**, 276–284.
- Nemani, R. R., C. D. Keeling, H. Hashimoto, W. M. Jolly, S. C. Piper, C. J. Tucker, R. B. Myneni, and S. W. Running (2001), Climate-driven increases in global terrestrial net primary production from 1982 to 1999, *Science*, **300**, 1560–1563.
- Nemani, R. R., C. D. Keeling, H. Hashimoto, W. M. Jolly, S. C. Piper, C. J. Tucker, R. B. Myneni, and S. W. Running (2003), Climate-driven increases in global terrestrial net primary production from 1982 to 1999, *Science*, **300**, 1560.
- Potter, C. S., J. T. Randerson, C. B. Field, P. A. Matson, P. M. Vitousek, H. A. Mooney, and S. A. Klooster (1993), Terrestrial ecosystem production: A process model based on global satellite and surface data, *Global Biogeochem. Cycles*, **7**, 811–841.
- Privette, J. L., C. Fowler, G. A. Wick, D. Baldwin, and W. J. Emery (1995), Effects of orbital drift on advanced very high resolution radiometer products: Normalized difference vegetation index and sea surface temperature, *Remote Sens. Environ.*, **53**, 164–171.
- Randerson, J. T., M. V. Thompson, C. M. Malmstrom, C. B. Field, and I. Y. Fung (1996), Substrate limitations for heterotrophs: Implications for models that estimate the seasonal cycle of atmospheric CO<sub>2</sub>, *Global Biogeochem. Cycles*, **10**, 585–602.
- Rao, C. R. N., and J. Chen (1995), Inter-satellite calibration linkages for the visible and near-infrared channels of the Advanced Very High Resolution Radiometer on the NOAA-u, -9, and -11 spacecraft, *Int. J. Remote Sens.*, **16**, 1931–1942.
- Rödenbeck, C., S. Houweling, M. Gloor, and M. Heimann (2003), CO<sub>2</sub> flux history (1982–2001) inferred from atmospheric data using a global inversion of atmospheric transport, *Atmos. Chem. Phys.*, **3**, 1919–1964.
- Rosen, J. M., et al. (1994), Decay of Mount Pinatubo aerosol at midlatitudes in the Northern and Southern Hemispheres, *J. Geophys. Res.*, **99**(D12), 25,733–25,739.
- Rouse, J. W., R. H. Haas, J. A. Schell, D. W. Deering, and J. C. Harlan (1974), Monitoring the vernal advancement and retrogradation (green-wave effect) of natural vegetation, *Rep. RSC 1978-4*, Remote Sens. Cent., Texas A&M Univ., College Station, Tex.
- Russell, P. B., et al. (1993), Pinatubo and pre-Pinatubo optical-depth spectra: Mauna Loa measurements, comparisons, inferred particle size distributions, radiative effects, and relationships to lidar data, *J. Geophys. Res.*, **98**(D12), 22,969–22,985.
- Sato, M., et al. (1993), Stratospheric aerosol optical depths, 1850–1990, *J. Geophys. Res.*, **98**(D12), 22,987–22,994.
- Schimmel, D., and D. Baker (2002), Carbon cycle: The wildfire factor, *Nature*, **420**, 29–30.
- Sellers, P. J. (1985), Canopy reflectance, photosynthesis, and transpiration, *Int. J. Remote Sens.*, **6**, 1335–1371.
- 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 (1996), A revised land surface parameterization (SiB2) for atmospheric GCMs: Part 1. Model formulation, *J. Clim.*, **9**, 676–705.
- Tian, H., J. M. Melillo, D. W. Kicklighter, A. D. McGuire, J. V. K. Helfrich, B. Moore, and C. J. Vörösmarty (1998), Effect of interannual climate variability on carbon storage in Amazonian ecosystems, *Nature*, **396**, 664–667.
- Townshend, J. R. G. (1994), Global data sets for land applications from the Advanced Very High Resolution Radiometer: An introduction, *Int. J. Remote Sens.*, **15**, 3319–3332.
- Tucker, C. J. (1979), Red and photographic infrared linear combination for monitoring vegetation, *Remote Sens. Environ.*, **8**, 127–150.
- Tucker, C. J., J. R. G. Townshend, and T. E. Goff (1985), African land-cover classification using satellite data, *Science*, **227**, 369–375.
- Tucker, C. J., Y. Fung, C. D. Keeling, and R. H. Gammon (1986), Relationship between atmospheric CO<sub>2</sub> variations and a satellite-derived vegetation index, *Nature*, **319**, 195–199.
- Tucker, C. J., W. W. Newcomb, and H. E. Dregne (1994), AVHRR datasets for the determination of desert spatial extent, *Int. J. Remote Sens.*, **15**, 3547–3566.
- Tucker, C. J., J. E. Pinzon, M. E. Brown, D. A. Slayback, E. W. Pak, R. Mahoney, E. F. Vermote, and N. El Saleous (2005), An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data, *Int. J. Remote Sens.*, **26**, 4485–4498.
- vander Werf, G. R., J. T. Randerson, G. J. Collatz, L. Giglio, P. S. Kasibhatla, A. F. Arellano, S. C. Olsen, and E. S. Kasichke (2004), Continental-scale partitioning of fire emissions during the 1997 to 2001 El Niño/La Niña period, *Science*, **303**, 73–76.
- 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**, 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.
- Wanner, W., X. Li, and A. H. Strahler (1995), On the derivation of kernels for kernel-driven models of bidirectional reflectance, *J. Geophys. Res.*, **100**(D10), 21,077–21,089.
- Zeng, N., J. D. Neelin, K. M. Lau, and C. J. Tucker (1999), Enhancement of interdecadal climate variability in the Sahel by vegetation interaction, *Science*, **286**, 1537–1540.
- Zeng, N., A. Mariotti, and P. Wetzel (2005), Terrestrial mechanisms of interannual CO<sub>2</sub> variability, *Global Biogeochem. Cycles*, **19**, GB1016, doi:10.1029/2004GB002273.
- 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.
- Zhou, L., R. K. Kaufmann, Y. Tian, R. B. Myneni, and C. J. Tucker (2003), Relation between interannual variations in satellite measures of vegetation greenness and climate between 1982 and 1999, *J. Geophys. Res.*, **108**(D1), 4004, doi:10.1029/2002JD002510.

G. J. Collatz, F. Hall, and J. G. Masek, NASA Goddard Space Flight Center, Code 614.4, Greenbelt, MD 20771, USA. (fghall@ltpmail.gsfc.nasa.gov)

## Sensitivity of surface climate to land surface parameters: A case study using the simple biosphere model SiB2

Lahouari Bounoua,<sup>1</sup> Jeffrey Masek,<sup>1</sup> and Yves M. Tourre<sup>2,3</sup>

Received 16 March 2006; revised 30 May 2006; accepted 24 August 2006; published 15 November 2006.

[1] Significant uncertainties may result from numerical models if fed with inappropriate input data. Biosphere-atmosphere transfer models are sensitive to input vegetation parameters, and the degree to which parameter estimates rely on the definition of the land cover type varies with models. In this study we use the simple biosphere model (SiB2) to evaluate uncertainties associated with misclassification of the land cover type and how they propagate to surface climate variables. We estimate that in regions with heterogeneous landscapes, the aggregation of land cover types from  $1 \times 1$  km to  $100 \times 100$  km horizontal resolution overestimates the area of the dominant type by up to 70%. The largest uncertainties associated with land cover misclassification are found in leaf area index and roughness length both of which have significant impact on the fluxes of carbon, water and energy at the earth surface. Other important uncertainties occur when the misclassification confuses plants with different carbon pathways. An assessment of the uncertainties is obtained comparing outcomes resulting from a choice of a dominant type in a  $100 \times 100$  km area to those obtained using a mosaic of land cover composition weighted by its fractional cover. The difference shows the choice of the dominant type to be cooler by  $0.6^\circ\text{C}$  than the average of the mosaic at local noon, while at night it is warmer by  $1.7^\circ\text{C}$ . Our results indicate that the diurnal temperature range (dtr) varies from  $13^\circ\text{C}$  for the dominant type to  $15^\circ\text{C}$  for the weighted average. The difference in the dtr is due to higher minimum temperature simulated with the dominant type. The choice of a dominant type also results in a daily carbon assimilation loss of 28,000 gC compared to the average.

**Citation:** Bounoua, L., J. Masek, and Y. M. Tourre (2006), Sensitivity of surface climate to land surface parameters: A case study using the simple biosphere model SiB2, *J. Geophys. Res.*, **111**, D22S09, doi:10.1029/2006JD007309.

### 1. Introduction

[2] Land surface models such as the Biosphere-Atmosphere Transfer Scheme-BATS [Dickinson, 1984] or the simple biosphere model SiB2 [Sellers *et al.*, 1996a] simulate the exchange of energy, water and momentum between the soil, vegetation and the atmosphere. These land surface models are coupled to atmospheric General Circulation Models (GCMs) and often operate at the same spatial resolution as the host GCM. The grid cell used in current GCMs is rather coarse (e.g.,  $2^\circ \times 2^\circ$ ) compared to the resolution of most satellite-derived land cover products. On the other hand the land cover within the grid is represented either by one biome in the case of “dominant” type or several biomes when “mosaics” are used [Koster and Suarez, 1996]. The biomes are used to prescribe many land

surface parameters to describe optical, morphological and physiological characteristics.

[3] There are obvious problems associated with prescribing the dominant land cover type within a coarse resolution grid cell. In areas of heterogeneous landscapes, the most common (modal) land cover type may represent a relatively small proportion of the total cell area. For example, comparing the 1 km MODIS land cover classification for North America with a  $1^\circ$  resolution version, the  $1^\circ$  version overestimates the proportion of the dominant type within each cell, sometimes by as much as 70% (Figure 1). As a result, outputs from models using coarse resolution may carry significant uncertainties at regional and local scales. In addition, these errors do not necessarily “cancel out” when considered across the continent as a whole (Table 1). Some subsidiary classes (e.g., urban land cover) disappear completely from the landscape when using the dominant type aggregation.

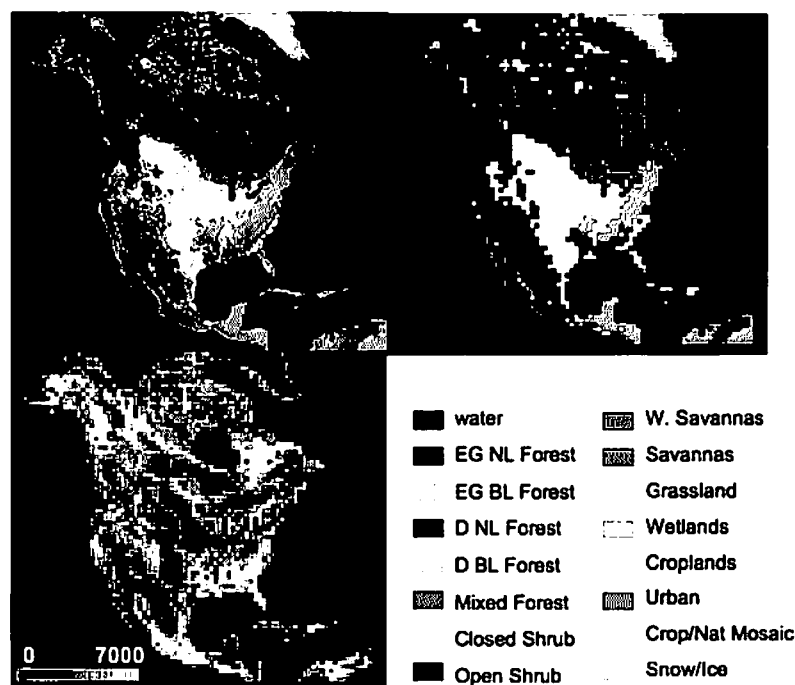
[4] The implication of these errors for modeling depends in part on the type of model under consideration. Climate models typically compute grid-size fluxes of water and energy at the land surface. The dominant biome concept used by most climate models is not representative of the mix of ecosystems coexisting within the same grid. However, it

<sup>1</sup>Biospheric Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.

<sup>2</sup>MEDIAS-France, Centre National d'Études Spatiales, Toulouse, France.

<sup>3</sup>Also at Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York, USA.





**Figure 1.** Differences between 1 km and 100 km ( $\sim 1^\circ$ ) land cover representations for North America. (top left) MODIS 1 km IGBP land cover classification, (top right) MODIS land cover map aggregated to  $1^\circ$  resolution using dominant type, and (bottom left) the difference between the area of the dominant type in the  $1^\circ$  cell (e.g., 10,000 km<sup>2</sup>) and the actual area of that type within the cell from the 1 km representation. In heterogeneous landscapes, the area of the dominant type is overestimated by up to 7000 km<sup>2</sup>, or 70% of the cell area. Color appears in back of the print issue.

is a reasonable compromise between high computational overhead and simple description of the physics. For land surface models that incorporate a carbon cycle such as SiB2, the dominant biome classification is even more problematic, especially in the tropics where it may misclassify plants species with different carbon pathways. These “third generations” land surface models [Sellers *et al.*, 1997] have a coupled photosynthesis-stomatal conductance module [Collatz *et al.*, 1991, 1992; Farquhar *et al.*, 1980] that allows the simultaneous computation of carbon assimilation while regulating the water loss through the stomates to the atmosphere [Cowan, 1977]. C4 plants, dominant in tropical and subtropical herbaceous vegetation [Collatz *et al.*, 1998] are much more water use efficient than their C3 counterparts occurring in all woody and temperate herbaceous plants and for given environmental conditions, they are capable of higher photosynthetic rates [Collatz *et al.*, 1992].

[s] Plant Functional types advocated by several land surface and climate modelers [e.g., Bonan *et al.*, 2002, and references therein] appear more appropriate for ecological models while landscapes represented by mosaics of ecosystems types seem to better capture the water and energy exchanges than a single dominant biome type. However, even when high-resolution satellite data provide a better description of landscape composition and land cover types, lack of data describing the morphological, optical and physiological properties of different types is still a problem. Vegetation affects surfaces fluxes through different mechanisms. The morphological aspect alters the

turbulent transfer of energy through roughness elements. The Leaf reflectance and absorbance of light in different wavelengths are important optical properties that determine the net amount of radiation absorbed by the canopy while the physiological activity controls the partitioning of the incoming energy into sensible and latent heat fluxes and determines the amount of carbon uptake.

**Table 1.** Area of Land Cover Types Across North American Domain (Figure 1) Using the  $1^\circ$  Dominant Type Classification and the 1 km Original Resolution<sup>a</sup>

IGBP Type	$1^\circ$ Area	1 km Area	Difference	Percent Difference
1, evergreen needleleaf forest	3.941	3.316	-0.625	-18.838
2, evergreen broadleaf forest	1.047	1.095	0.048	4.361
3, deciduous needleleaf forest	0.000	0.025	0.025	100.000
4, deciduous broadleaf forest	0.889	0.880	-0.009	-0.968
5, mixed forests	1.927	2.235	0.308	13.798
6, closed shrublands	0.060	0.190	0.130	68.486
7, open shrublands	5.881	5.129	-0.752	-14.655
8, woody savannas	1.138	1.428	0.290	20.298
9, savannas	0.250	0.491	0.241	49.097
10, grasslands	2.040	2.142	0.102	4.780
11, permanent wetlands	0.060	0.127	0.067	52.980
12, croplands	2.244	2.192	-0.052	-2.393
13, urban and built-up	0.000	0.081	0.081	100.000
14, cropland/natural vegetation mosaic	0.619	0.862	0.243	28.237
15, snow and ice	1.095	1.129	0.034	3.012
16, barren or sparsely vegetated	0.547	1.082	0.535	49.445

<sup>a</sup>Area is in km<sup>2</sup>.

Table 2. Example of Time Invariant Biome Type–Dependent Land Surface Characteristics in SiB2<sup>a</sup>

Parameter	Units	Biome Types			
		4	12	3	7
Morphological properties					
Canopy height	m	17.0	1.0	20.0	0.5
Canopy base height	m	8.5	0.1	10.0	0.1
Inflection height for LAI	m	10.0	0.55	15.0	0.3
Leaf angle distribution factor	...	0.01	−0.3	0.01	0.01
Leaf width	m	0.001	0.01	0.04	0.003
Leaf length	m	0.055	0.3	0.1	0.03
Soil depth	m	2.0	1.5	2.0	1.5
Root depth	m	1.5	1.0	1.5	1.0
Optical properties					
Leaf reflectance, visible, live	...	0.07	0.105	0.07	0.1
Leaf reflectance, visible, dead	...	0.16	0.36	0.16	0.16
Leaf reflectance, NIR, live	...	0.35	0.58	0.4	0.47
Leaf reflectance, NIR, dead	...	0.39	0.58	0.39	0.39
Leaf transmittance, visible, live	...	0.05	0.07	0.05	0.05
Leaf transmittance, visible, dead	...	0.001	0.22	0.001	0.001
Leaf transmittance, NIR, live	...	0.10	0.25	0.15	0.25
Leaf transmittance, NIR, dead	...	0.001	0.38	0.001	0.001
Physiological properties					
Maximum Rubisco capacity Vmax	$\mu\text{mol m}^{-2} \text{s}^{-1}$	60	100	80	60
High temperature stress factor	...	303	308	307	313
Low temperature stress factor	...	278	281	281	288
Half inhibition water potential	...	−200	−200	−200	−300

<sup>a</sup>For more detailed list see *Sellers et al.* [1996b].

[6] Remote sensing data has been used to map global distributions of land cover and vegetation parameters for two decades. Early efforts focused on fixed land cover classification from the NOAA AVHRR instrument [Townshend *et al.*, 1991; DeFries and Townshend, 1994; Loveland *et al.*, 2000]. Since 2000, the Terra MODIS instrument has been used to produce land cover products at 1 km resolution [Friedl *et al.*, 2002]. Implications of land cover misclassification for parameter estimates for global land surface models was examined by DeFries and Los [1999]. They reported a realistic assessment of the implications of misclassifications on leaf area index and surface roughness, as estimated for use in the simple biosphere model (SiB2). However, they did not evaluate the propagation of the impact to surface climate.

[7] In this paper, we use data from the International Satellite Land Surface Climatology Project initiative II [Hall *et al.*, 2005] and the simple biosphere model SiB2 [Sellers *et al.*, 1996a] to evaluate and illustrate uncertainties associated with aggregating land cover information to GCM resolution, and the impact of these uncertainties on surface climate variables.

## 2. Model and Data

### 2.1. Model

[8] The simple biosphere model (SiB2) of Sellers *et al.* [1996a] is used to assess the impact of land cover and its associated parameters on surface energy and water components. SiB2 is a biophysical land surface model that computes the exchanges of energy, water, momentum and carbon between the biosphere and atmosphere accounting explicitly for 12 different biome types. The version used for this assessment uses the vegetation dominant type as representative of the grid cell ecosystems. Time-varying parameters describing the phenology such as the fraction of

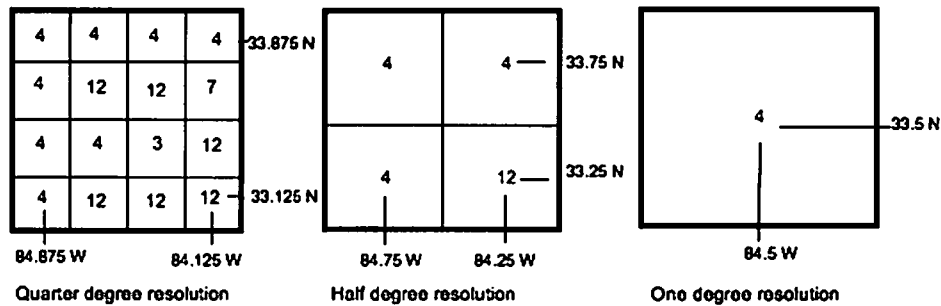
the photosynthetically active radiation (FPAR), the leaf area index (LAI), the green fraction of the canopy, the roughness length and the zero plan displacement are all derived from satellite data. However, many other parameters describing optical, morphological and physiological properties are prescribed from lookup tables and are vegetation type–dependent (Table 2).

[9] SiB2 also includes a photosynthesis–conductance module to simulate the simultaneous exchange of carbon and water vapor in and out of the leaf. Carbon assimilation is calculated from FPAR constrained by light limitation, environmental conditions, and a prescribed maximum photosynthetic capacity ( $V_{\max}$ ) which depends on the biome type. From the canopy net assimilation rate ( $A_c$ ) canopy stomatal conductance ( $g_c$ ) is calculated as

$$g_c = m \frac{A_c}{C_s} h_s p + b L_T; \quad (1)$$

where  $p$  is the atmospheric pressure and  $L_T$  is total leaf area index.  $m$  is the stomatal slope factor and  $b$  the minimum stomatal conductance, and both are biome type–dependent parameters that differentiate between C3 and C4 vegetation. While the control of stomatal conductance by atmospheric humidity is explicit through  $h_s$ , effects of temperature and soil moisture stresses are implicit in  $A_c$  [Bounoua *et al.*, 2004].  $C_s$  is the intercellular CO<sub>2</sub> concentration.

[10] The capacity of vegetation to convert soil moisture into latent heat flux is determined by FPAR and stomatal conductance. The former is prescribed and derived from global NDVI data sets while the latter depends on atmospheric conditions and the amount of water available in the model's root-zone layer; thus establishing a strong coupling between the atmospheric forcing and the soil hydrology. The satellite-derived FPAR is converted to LAI, which is used in a radiative transfer model [Sellers *et al.*, 1996b] to



**Figure 2.** Land cover classification at quarter, half and  $1^\circ$  resolutions showing the different vegetation classes within the  $1^\circ$  grid, where 3 indicates mixed broadleaf and needleleaf trees (6.25%), 4 indicates needleleaf-evergreen trees (50%), 12 indicates agriculture/C3-grassland (37.5%), and 7 indicates C4-grasslands (6.25%). Values in parentheses represent fractional coverage for each biome type from the quarter degree to the  $1^\circ$  grid.

calculate albedo and radiation partition between the canopy and ground. The surface roughness length also depends on the LAI and the biome's canopy height.

[11] The two-stream approximation is used to compute surface albedo from soil and canopy reflectance. Leaf reflectance and transmittance for near infrared and visible wavelengths for both green and dead leaves is prescribed for each of the 12 biomes defined in SiB2.

[12] Fluxes of energy and water are calculated from the potential differences of temperature and specific humidity between the vegetated surface and the canopy air space. Therefore the definition of the biome type and the canopy leaf density in this model virtually affect all components of the energy, water and momentum budgets.

[13] SiB2 has been extensively used in previous studies in offline mode [e.g., Zhang *et al.*, 1996; Bounoua *et al.*, 2004; Pongratz *et al.*, 2006] and coupled with a GCM [e.g., Randall *et al.*, 1996; Bounoua *et al.*, 1999, 2000, 2002; DeFries *et al.*, 2002].

## 2.2. Data

[14] We use the land cover map developed by the University of Maryland and the 1987 monthly FASIR-NDVI both provided by International Satellite Land Surface Climatology Project Initiative II (ISLSCP II) [Hall *et al.*, 2005] to compute all biophysical fields necessary for SiB2 [Sellers *et al.*, 1996a]. These include LAI, FPAR, the canopy greenness fraction, the roughness length and some other aerodynamic parameters. The climate data are provided by the ISLSCP initiative I [Meeson *et al.*, 1995] and correspond to 1987. The climate data consist of surface short-wave and long-wave diffuse and direct radiation, wind speed, surface air temperature and large and convective scales precipitation. The soil map is constructed using source soil data at 4 km horizontal resolution from Food and Drug Organization. Soil texture was generated using the 9 classes defined by Zohler [1986] and soil properties were derived from Clapp and Hornberger [1978] for each of the soil texture class following Sellers *et al.* [1996b]. All data are provided at  $1^\circ \times 1^\circ$  horizontal resolution and the climate drivers have a temporal resolution of 1 hour. In addition to the  $1^\circ \times 1^\circ$  land cover map, we use a  $0.25^\circ \times 0.25^\circ$  land cover map [Hall *et al.*, 2005] from which the  $1^\circ \times 1^\circ$  has

been derived by simple aggregation retaining the dominant type as is used in most land surface models.

## 3. Model Runs

[15] We selected a small domain over the Eastern part of the United States that presents a spatial heterogeneity at a scale of  $1^\circ \times 1^\circ$ . We overlaid that domain over a quarter degree land cover map and selected all biomes existing over the  $1^\circ \times 1^\circ$  grid (Figure 2). We run the model in a single column mode for the month of July 1987 at  $1^\circ \times 1^\circ$  resolution using the different biomes forced by the same climate and soil data. For each individual run we assume that only one biome type covers a fraction of the grid cell and is described by the same NDVI. The only land characteristic that changed is the cover type and its associated parameters. This way we are able to isolate changes in model outputs that are exclusively due to land cover misclassification and the impact of their associated parameters.

## 4. Results and Discussions

[16] In the following sections, we summarize the impact of biome misclassification on the biophysical fields as well as the propagation of this impact to surface climate variables through the SiB2 land surface model.

### 4.1. Biophysical Fields

[17] Table 3 shows biophysical fields for the month of July computed with the same value of NDVI for the different biomes coexisting in the quarter degree grid (Figure 2). The biophysical parameters are computed using algorithms described by Sellers *et al.* [1996b] and are extensively reviewed by Los *et al.* [2000]. Note that some of these biophysical parameters are common to more than one submodel in SiB2. For example, the leaf area index (LAI) is used in the radiative transfer, the momentum and the stomatal resistance calculations, whereas the canopy greenness fraction and the roughness length affect only the radiative and momentum transfers, respectively. In Table 3, we also provide a weighted average, which gives an "accurate" parameter value for the entire  $1^\circ$  cell, weighted by the fractional coverage of each type.

**Table 3.** Comparison of Biophysical Parameters for Selected Biome Types With the Same Value of the NDVI for July<sup>a</sup>

Biome	3 (Mixed Forest)	4 (Evergreen Needleleaf Forest)	12 (Cropland/C3-Grassland)	7 (C4-Grassland)	Weighted Value
LAI	2.29	3.93	1.51	1.26	2.75
FPAR	0.63	0.60	0.64	0.57	0.62
Greenness	0.88	0.90	0.85	0.85	0.88
Roughness, m	1.02	1.20	0.13	0.13	0.72

<sup>a</sup>Weighted average value represents average weighted by the fractional coverage. LAI, FPAR and greenness are dimensionless and the roughness length is in meters.

[18] There is not much difference in the greenness fraction between the different biomes. The canopy greenness fraction describes the change in LAI between two consecutive time periods and is mostly representative of the amount of dead leaves between the two periods [Sellers *et al.*, 1996a]. The largest difference between biome 4 (needleleaf evergreen) and 7 (C4-grassland) is about 0.04. Similarly, FPAR does not show large differences between biomes. FPAR strongly depends on leaf optical properties and the time-mean radiation weighted extinction coefficient. While FPAR varies exponentially for small leaf area index it reaches saturation at higher LAI values.

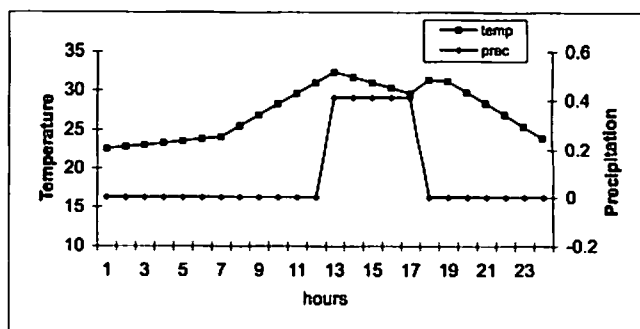
[19] The Leaf area index (LAI) presents the largest difference. The definition of the biome type is extremely important for the computation of LAI, an exponential extinction coefficient is usually used for broadleaves (types 7 and 12) and a linear model for clumped canopies (type 4) while a mixed model is used for the mixed forest (type 3) which includes broadleaf and needleleaf trees. Biomes types have prescribed values of maximum LAI ( $LAI_{max}$ ) which combined with the extinction coefficient have a large influence on LAI [Sellers *et al.*, 1996b]. For example for forests (biomes 4 and 3), the  $LAI_{max}$  is 8 and 7.5, respectively while it is only 5 for short vegetation, biomes 7 and 12. Therefore even if assigned the same satellite observed NDVI value, a grid cell in such land surface models can produce differences in leaf area index large enough to cause significant differences in the surface fluxes. In this case the largest discrepancy of 2.67 is obtained between biomes 4 and 7 and its absolute value is larger than the LAI for biome 3. The weighted average value, obtained as an average weighted by the fractional cover of each biome type, is closer to that of the dominant biome which represents half of the grid cell. The aerodynamic transfer parameter (roughness length) depends on the shape of the canopy and is therefore strongly affected by the choice of the biome type. It is used to determine the aerodynamic resistance which in turn enters the computation of the shear stress [Sellers *et al.*, 1996a, 1996b]. The difference in roughness between tall vegetation (forests) and short vegetation cropland/grassland is about 1 m (Table 3).

#### 4.2. Model Response

[20] In this section we document the model response to the different types of biomes forced by the same climate. We present a comprehensive comparison between the four different biomes' responses and provide a plausible scenario of the error associated with the choice of the dominant type rather than using a mosaic approach where each biome within the grid is represented in proportion to its fractional coverage. We select for discussion a day during the month of July characterized by a strong morning convective

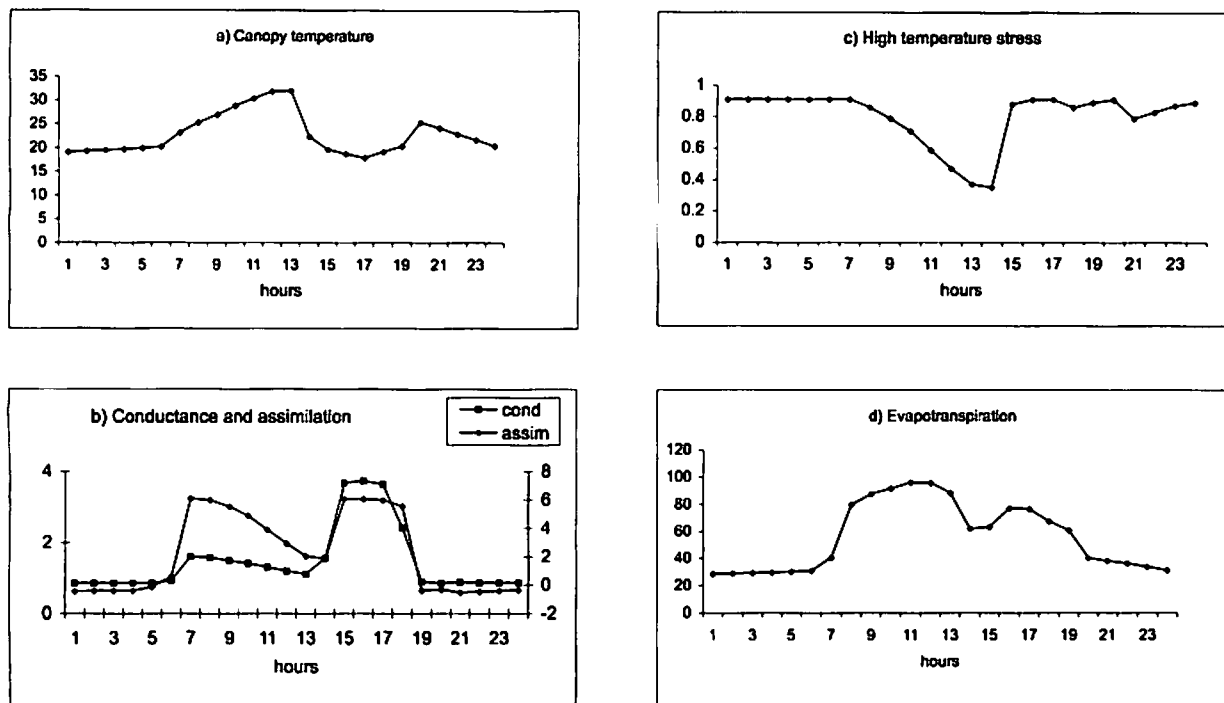
warming followed by a midday shower (4 July 1987). Figure 3 shows the diurnal variation of the surface air temperature and the total precipitation used to drive the land surface model for the selected day. The selection of such day with strong warming gives a good example as to the responses of different biomes to high temperature stress and allows clear evaluation of extreme uncertainties in model outcomes associated with land cover aggregations and scaling. The meteorological conditions of the selected day are typical of the month of July for the region.

[21] Figure 4 shows the response, for evergreen needleleaf forest (biome 4), of canopy temperature, high temperature stress factor, canopy conductance and canopy carbon assimilation rates as well as the canopy evapotranspiration, i.e., the sum of canopy interception and transpiration rates, as both are affected by the biome type through the leaf area. In SiB2, needleleaf evergreen forest is assigned a high temperature stress factor of 30°C which allows a normal physiological operating range below this critical temperature. When canopy temperature approaches this critical value, the vegetation is high-temperature stressed resulting in closure of stomates, reduction in stomatal conductance, and reduced carbon uptake. For this particular time of the year (July), the vegetation was stressed during most of the day when ambient air temperature (Figure 3) was close to 30°C. The high temperature stress (Figure 4c) increases in the morning as the ambient temperature increases following the solar heating. Around 02 PM local time the temperature stress reached its maximum, however the carbon assimilation reduction started earlier during the day around 8 AM driving down the stomatal conductance. By 11 AM the canopy temperature reached 30.36°C which corresponds approximately to half inhibition of the photosynthetic activity rate. The warming intensified until 01 PM which resulted in a "crash" of the conductance and assimilation functions. The canopy temperature reached a maximum of



**Figure 3.** Boundary layer temperature in Celsius (left axis) and total precipitation ( $\text{mm h}^{-1}$ ) for 4 July 1987 over the selected region.





**Figure 4.** Diurnal response for key surface variables for biome 4: (a) canopy temperature ( $^{\circ}\text{C}$ ); (b) canopy conductance ( $\text{mm s}^{-1}$ ) and carbon assimilation ( $\mu\text{mol m}^{-2} \text{s}^{-1}$ ), where the left axis represents conductance; (c) high temperature stress (dimensionless, 1 is no stress, and 0 is high stress); and (d) evaporation ( $\text{W m}^{-2}$ ).

$32.03^{\circ}\text{C}$ . The evolution of the canopy transpiration (not shown), controlled mainly by the stomatal conductance, dropped from a value of about  $89 \text{ W m}^{-2}$  at 01 PM to  $9 \text{ W m}^{-2}$  at 02 PM illustrating the strong sensitivity of the physiological activity to surface characteristics. The evapotranspiration curve (Figure 4d) shows the compensating effect of canopy interception loss on canopy evapotranspiration. Because of relatively high LAI value, the canopy intercepts more rain which started at 1 PM and ended at 5 PM. The canopy interception loss rate was about  $50 \text{ W m}^{-2}$  between 2 PM and 5 PM and only  $10 \text{ W m}^{-2}$  between 6 and 7 PM. As the forcing conditions changed during early afternoon and rainfall started, the ambient air temperature dropped and so did the high temperature stress (Figure 4c). Consequently, carbon assimilation and canopy conductance increased which led to a cooling. This situation is typical of a summer day over the U.S. northeastern region.

[22] When the exact same forcing conditions are imposed upon a pixel labeled as mixed forest (biome 3), results were similar in form to those of biome 4 but slightly different in terms of intensity. Biome 3 represents a mix of broadleaf and needleleaf trees and uses a linear mixing model for LAI. The high temperature stress assigned to this biome is  $34^{\circ}\text{C}$  making it more heat tolerant than biome 4. In this case, the high temperature stress started 2 hours later and increased to reach a maximum value of 0.53 at 01 PM compared to 0.37 for biome 4 at the same time. Consequently the assimilation rate was larger reaching a maximum of  $8 \mu\text{mol m}^{-2} \text{s}^{-1}$  versus 6 for biome 4 and the decrease was not as sharp as observed in the case of biome 4. Although the carbon assimilation rate is larger for biome 3, the canopy conductance remains smaller than that of biome 4 and this is

because of the smaller LAI value (see Table 3 and equation (1)). Consequently the canopy temperature reached a higher maximum of  $33.51^{\circ}\text{C}$  during the midday hours (Figure 5).

[23] The response of a C3-grassland (biome 12) to warm conditions was not as severe as seen for tall trees. The stress function (not shown) showed similar but less intense diurnal variation. This is partially related to a smaller LAI value and a lesser water demand compared to tall trees. However, short vegetation access water from a root zone layer shallower than that of tall trees. C3-grasses are more heat tolerant than some trees and have a higher photosynthetic capacity, about 60% more than a tree (Table 2). This resulted in much higher carbon assimilation rate and higher conductance to transpiration. Because of the small leaf area index, the canopy interception was much smaller than in the case of the forest. From the morphological perspective however, grasslands are much shorter than trees; the canopy height for biomes 3 and 4 are 17 and 20 m, respectively while it is only of 1 m for C3-grassland. This is an important biome-dependent parameter that plays a central role in the turbulent transfers. The aerodynamic resistance regulating these transfers is inversely related to the roughness length, which going from a tall tree to a grassland has been decreased by about ninefold. Since the climate forcing was not allowed to vary in our simulations, the increase in aerodynamic resistance resulted in warming of the surface. This type of misclassification of land cover from tall forest to grassland is a classical case of morphological versus physiological interactions. The physiological activity tends to cool the canopy and canopy air space but the morphological impact alters the aerodynamic transfer and as such counterbalances the physiological effect for an overall

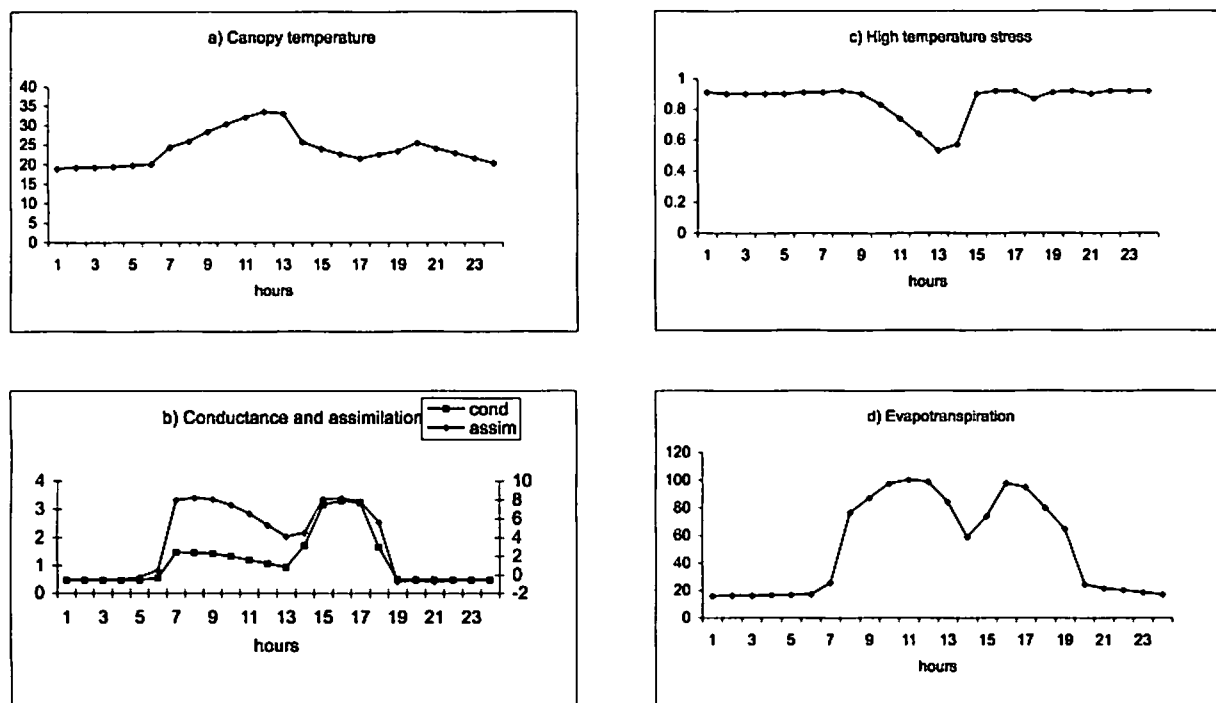


Figure 5. (a–d) Same as Figure 4 except for biome 3.

daytime warming [Bounoua *et al.*, 2002, 2004; DeFries *et al.*, 2002].

[24] An interesting case is observed when the confusion in the selection of the cover type occurs between short vegetation with similar morphological and optical characteristics but different physiological characteristics. This is the case when an agriculture/C3-grassland (biome 12) grid cell is misclassified as C4-grassland (biome 7). Biome 7 has a

high temperature stress factor of 40°C and is therefore much more resistant to heat stress compared to any of the biomes described above. C4-plants are water efficient and assimilate about twice as much carbon than C3-plants for the same amount of water use. In the model, this characteristic is described by a difference in the stomatal conductance slope factor (1). The diurnal variation of the response of biome 7 is shown in Figure 6. During daylight hours when photo-

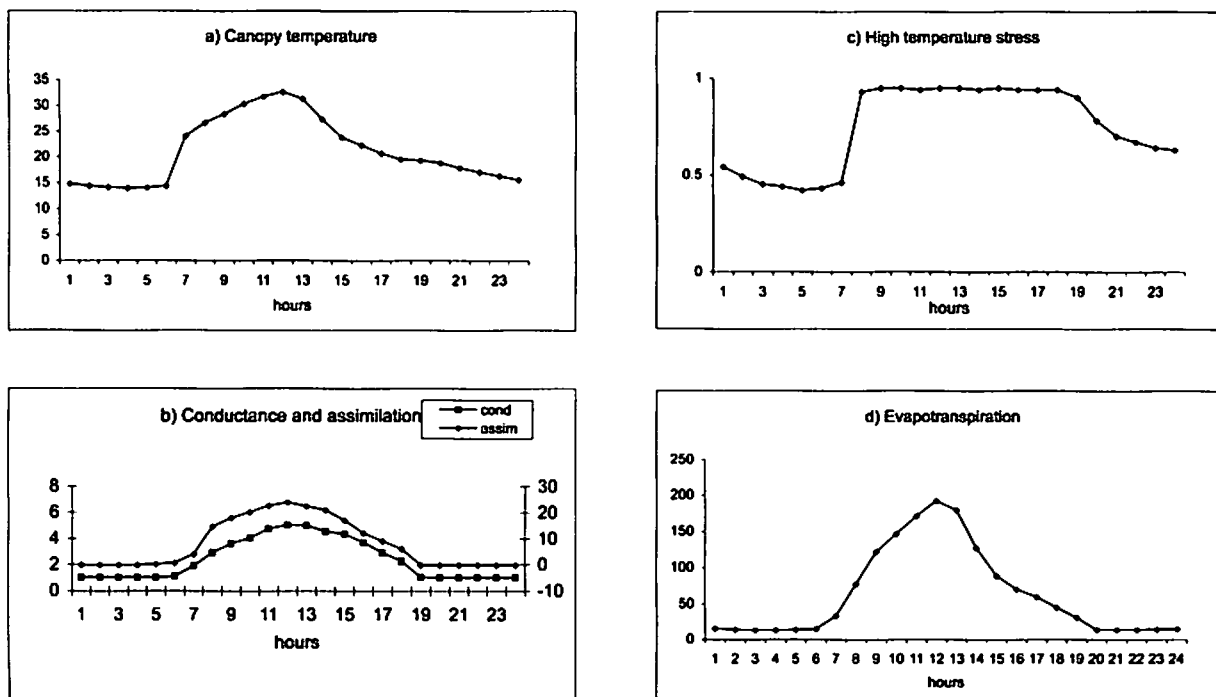
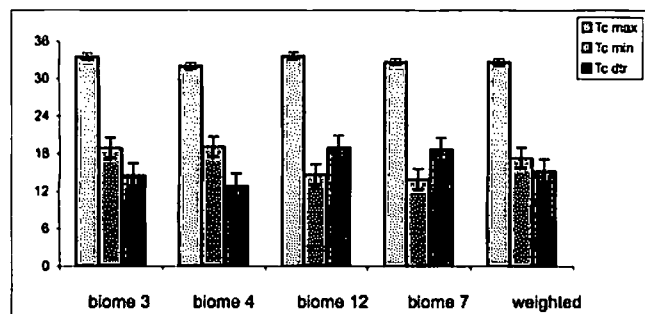


Figure 6. (a–d) Same as Figure 4 except for biome 7.



**Figure 7.** Canopy maximum temperature (left bar), minimum temperature (middle bar) and diurnal temperature range (right bar) for all four biomes and the weighted average. Bars represent standard deviations from the weighted average.

synthesis is operating, the high temperature stress was negligible (Figure 6c), the small level of stress seen during nighttime is due to the fact that biome 7 is more sensitive to low temperature; but nighttime physiological stress is irrelevant as photosynthetic activity is “shut down.” The photosynthetic activity of this biome was not altered by warm ambient daytime temperature, which reached about 33°C, and the net carbon assimilation rate reached a maximum of about 24  $\mu\text{moles m}^{-2} \text{s}^{-1}$ . The canopy conductance shows a normal behavior following the assimilation rate and reaches a maximum of 5.04  $\text{mms}^{-1}$  around local noon. This led to a higher evapotranspiration rate than that obtained with the C3-grassland, mixed forest and needleleaf forest, respectively (192 versus 155, 100 and 96  $\text{W m}^{-2}$ ) and resulted in a maximum temperature of 32.69°C or 1°C cooler than that obtained using grassland (Figure 6).

[25] Results obtained from the different biome configurations are compared to those obtained using all four biomes weighted by their respective fractional coverage as described in Figure 2. In this particular case, the largest difference is due to tall versus short vegetation rather than the fractional coverage. This is in part due to the fact that a large part of the grid is covered by biome 4 and 12. During daytime the weighted average canopy temperature appears representative of the spatial heterogeneity of the landscape, however nighttime temperatures are much warmer than the weighted average for trees and cooler for short biomes. Rainfall occurrence strongly alters the canopy temperature diurnal pattern for tall biomes because of their capacity to intercept rainwater which reevaporates at potential rate, but has almost no effect on short biomes. The temperature response for the four different biomes showed a strong asymmetry in the diurnal temperature range (dtr) mainly due to difference in minimum temperatures (Figure 7). Clearly minimum temperatures for short biomes are lower than those simulated by the model for tall biomes leading to different diurnal temperature amplitude. This indicates that through its vertical structure vegetation plays an important role in the determination of local climates and its accurate classification in land surface models is central for improved climate prediction. Our results show that the diurnal temperature range could vary between about 13°C for tall

vegetation (biome 4) to about 19°C for short vegetation (biome 12) with a standard deviation between the four biomes and the weighted average being about 1.8°C.

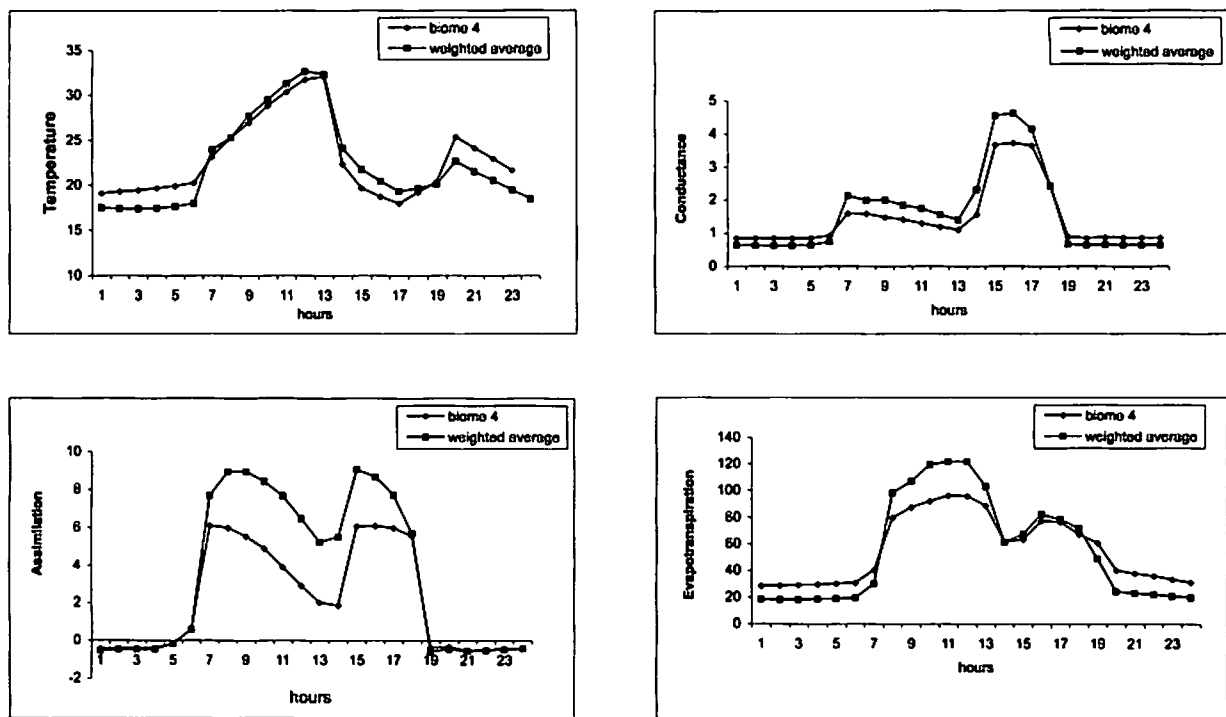
[26] It is important to note that results presented here are highly dependent on the land surface model parameters and its physical parameterizations and should not be considered as predictions. Two of the most important parameters that affected results in this assessment are the canopy height and the biome tolerance to high temperature. Other parameters with potentially important impact are leaf reflectance and transmittance which enter the computation of snow free albedo and affect the surface radiation balance. However, albedo alters the net energy absorbed by the surface but has no direct effect on the partitioning of this energy among different surface fluxes.

[27] For this example, a reasonable assessment of the uncertainties is obtained between outcomes resulting from a choice of an evergreen needleleaf forest, the dominant type, and the weighted average representing a “statistical biome.” This difference (Figure 8) shows the needleleaf forest (biome 4) 0.6°C cooler than the weighted average during local noon, while at nighttime it is warmer by 1.7°C. Most importantly the choice of biome 4 as a dominant type results in a daily area integrated carbon assimilation loss of 28,000 gC compared to the weighted average. Composite diurnal cycles for the month of July have been constructed using hourly outputs over 744 realizations. The monthly mean diurnal response is similar to that of the selected day for all variables analyzed (not shown). However, as expected, the magnitude of the change is different. In general the monthly average resulted in slightly cooler canopy temperature response. For example, the monthly composite diurnal temperature difference shows the dominant type (biome 4) to be cooler by about 0.5°C than the weighted average during daytime, while it is warmer by up to 1.4°C during the night. This leads to a monthly composite dtr of 10.7°C and 12.8°C for the dominant type and the weighted average, respectively. The area integrated monthly composite carbon loss associated with the choice of biome 4 instead of the weighted average is about 23,000 gC.

## 5. Concluding Remarks

[28] The results from this study illustrate the uncertainties encountered by a numerical model if fed with inappropriate input data. Results are also affected to some extent by the quality of the climate input data. Development of data sets such as those produced by the International Satellite Land Surface Climatology Project Initiative II (ISLSCP II) is of central importance for improving model inputs or to validate their outputs. Most current land surface models are complex and require a large set of internally consistent input data that is not readily available from a single uniform source. Therefore sources of uncertainties are large.

[29] The problem of scaling land cover type to models resolution is typical and all modelers face it. This study shows that scaling data up from fine to coarse resolution for modeling use can produce significant uncertainties in model's simulations. The degree to which parameter estimate rely on the definition of the land cover type varies with models. The simple biosphere model assessed in this study uses satellite data in addition to land cover type to determine



**Figure 8.** Comparison of key surface climate variables as obtained from the dominant type (biome 4) and the weighted average, for canopy temperature ( $^{\circ}\text{C}$ ), canopy conductance ( $\text{mm s}^{-1}$ ), assimilation ( $\mu\text{moles m}^{-2} \text{s}^{-1}$ ) and evapotranspiration ( $\text{W m}^{-2}$ ).

some of its biophysical parameters and therefore has a reduced reliance on land cover type because satellite data carry much of the spatial heterogeneity of the surface characteristics. Other land surface models that do not use satellite data may be more sensitive to misclassification errors and their implications in the computation of surface fluxes. Our study also shows that modeling uncertainties associated with land cover misclassification are more important in heterogeneous regions.

[30] There are many other parameters within SiB2 and other land surface models that rely on land cover type. Our study is an example illustrating land cover misclassification errors and how they propagate down to the surface climate variables. It raises the need to develop global data sets to characterize the vegetation morphological, optical and physiological parameters necessary for current surface vegetation transfer schemes common to most climate models. Some of these properties such as canopy top and base height as well as canopy shape could probably be developed from radar or lidar data, but other algorithms depicting vegetation health, temperature and water stress levels from remotely sensed data are highly desirable. This problem will become more acute in the near future as high-resolution satellite vegetation data such as MODIS data at 250 m horizontal resolution will be more accessible. However, if this fine resolution data set is not accompanied by vegetation morphological, optical and physiological characteristics at an appropriate spatial scale, the modeling of surface water and energy transfers at the land surface-atmosphere interface will suffer and will hamper our ability to improve climate predictability. Not only will surface characteristics ensure that major model's components are appropriately

parameterized, but they will also and most importantly help solve some problems associated with scaling of vegetation data and the loss of information contained in the spatial heterogeneity. Projects such as ISLSCP II should be encouraged and should be oriented toward collection, formatting and distribution of specific sets of data that are most needed.

[31] **Acknowledgments.** This work was funded by NASA Energy and Water Cycle Science NRA NN-H-04-Z-YS-005-N LCLUC part (PI: L. Bounoua). Y. Tourre's contribution was supported by LDEO 6943.

## References

- Bonan, G. B., S. Levis, L. Kergoat, and K. W. Oleson (2002), Landscapes as patches of plant functional types: An integrating concept for climate and ecosystem models, *Global Biogeochem. Cycles*, **16**(2), 1021, doi:10.1029/2000GB001360.
- Bounoua, L., G. J. Collatz, P. J. Sellers, D. A. Randall, D. A. Dazlich, S. O. Los, J. A. Berry, I. Fung, C. J. Tucker, and C. B. Field (1999), Interactions between vegetation and climate: Radiative and physiological effects of doubled atmospheric  $\text{CO}_2$ , *J. Clim.*, **12**, 309–323.
- Bounoua, L., G. J. Collatz, S. O. Los, P. J. Sellers, D. A. Dazlich, C. J. Tucker, and D. A. Randall (2000), Sensitivity of climate to changes in NDVI, *J. Clim.*, **13**, 2277–2292.
- Bounoua, L., R. S. DeFries, G. J. Collatz, P. J. Sellers, and H. Khan (2002), Effects of land cover conversion on surface climate, *Clim. Change*, **52**, 29–64.
- Bounoua, L., R. S. DeFries, M. L. Imhoff, and M. K. Steininger (2004), Land use and local climate: A case study near Santa Cruz, Bolivia, *Meteorol. Atmos. Phys.*, **86**, 73–85.
- Clapp, R. B., and G. M. Hornberger (1978), Empirical equations for some soil hydraulic properties, *Water Resour. Res.*, **14**(4), 601–604.
- Collatz, G. J., J. T. Ball, C. Grivet, and J. A. Berry (1991), Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: A model that includes a laminar boundary layer, *Agric. For. Meteorol.*, **54**, 107–136.
- Collatz, G. J., M. Ribas-Carbo, and J. A. Berry (1992), Coupled photosynthesis-stomatal conductance model for leaves of  $\text{C}_4$  plants, *Aust. J. Plant Physiol.*, **19**(1), 519–538.



- Collatz, G. J., J. A. Berry, and J. S. Clark (1998), Effects of climate and atmospheric CO<sub>2</sub> partial pressure on the global distribution of C<sub>4</sub> grasses: Present, past, and future, *Oecologia*, **114**, 441–454.
- Cowan, I. R. (1977), Stomatal behavior and environment, *Adv. Bot. Res.*, **4**, 117–228.
- DeFries, R. S., and S. O. Los (1999), Implications of land cover misclassification for parameter estimates in global land surface models: An example from the simple biosphere model (SiB2), *Photogramm. Eng. Remote Sens.*, **65**(9), 1083–1088.
- DeFries, R. S., and J. R. G. Townshend (1994), NDVI-derived land cover classifications at a global scale, *Int. J. Remote Sens.*, **15**, 3567–3586.
- DeFries, R. S., L. Bounoua, and G. J. Collatz (2002), Human modification of the landscape and surface climate in the next fifty years, *Global Change Biol.*, **8**, 438–458.
- Dickinson, R. (1984), Modeling evapotranspiration for three-dimensional global climate models, in *Climate Processes and Climate Sensitivity*, *Geophys. Monogr. Ser.*, vol. 29, pp. 58–72, AGU, Washington, D. C.
- Farquhar, G. D., S. von Caemmerer, and J. A. Berry (1980), A biochemical model of photosynthetic CO<sub>2</sub> assimilation in leaves of C<sub>3</sub> species, *Planta*, **149**, 78–90.
- Friedl, M. A., et al. (2002), Global land cover mapping from MODIS: Algorithms and early results, *Remote Sens. Environ.*, **83**(1–2), 287–302.
- Hall, F. G., G. Collatz, S. Los, E. B. de Colstoun, and D. Landis (Eds.) (2005), ISLSCP Initiative II [DVD/CD-ROM], NASA, Washington, D. C.
- Koster, R. D., and M. J. Suarez (1996), Energy and water balance calculations in the mosaic LSM, *NASA Tech. Memo. 104606*, vol. 9.
- Los, S. O., G. J. Collatz, P. J. Sellers, C. M. Malstrom, N. H. Polack, R. S. Defries, L. Bounoua, M. T. Parris, C. J. Tucker, and D. A. Dazlich (2000), Global 9-year biophysical land-surface data set from NOAA AVHRR data, *J. Hydrometeorol.*, **1**, 183–199.
- Loveland, T. R., B. C. Reed, J. F. Brown, D. O. Ohlen, Z. Zhu, L. Yang, and J. W. Merchant (2000), Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data, *Int. J. Remote Sens.*, **21**, 1303–1330.
- Meeson, B. W., F. E. Corpew, J. M. P. McManus, D. M. Myers, J. W. Closs, K.-J. Sun, D. J. Sunday, and P. J. Sellers (1995), ISLSCP Initiative I—Global data sets for land-atmosphere models, 1987–1988 [CD-ROM], NASA Goddard Space Flight Cent., Greenbelt, Md.
- Pongratz, J., L. Bounoua, R. S. DeFries, D. C. Morton, L. O. Anderson, W. Mauser, and C. A. Klink (2006), The impact of land cover change on surface energy and water balance in Mato Grosso, Brazil, *Earth Interactions*, in press.
- Randall, D. A., et al. (1996), A revised land surface parameterization (SiB2) for GCMs. Part III: The greening of the Colorado State University general circulation model, *J. Clim.*, **9**, 738–763.
- Sellers, P. J., D. A. Randall, G. J. Collatz, J. A. Berry, C. B. Field, D. A. Dazlich, C. Zhang, and L. Bounoua (1996a), A revised land surface parameterization (SiB2) for atmospheric GCMs. Part I: 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 (1996b), A revised land surface parameterization (SiB2) for atmospheric GCMs. Part II: The generation of global fields of terrestrial biophysical parameters from satellite data, *J. Clim.*, **9**, 706–737.
- Sellers, P. J., et al. (1997), Modeling the exchanges of energy, water, and carbon between continents and the atmosphere, *Science*, **275**(5299), 502–509.
- Townshend, J., C. Justice, W. Li, C. Gurney, and J. McManus (1991), Global land cover classification by remote sensing: Present capabilities and future possibilities, *Remote Sens. Environ.*, **35**(2–3), 243–255.
- Zhang, C., D. A. Randall, P. J. Sellers, and A. S. Denning (1996), Calculation of the global land surface energy, water and C fluxes with an off-line version of SiB2, *J. Geophys. Res.*, **101**, 19,061–19,075.
- Zobler, L. (1986), A world soil file for global climate modeling, *NASA Tech. Memo.*, 87802.

L. Bounoua and J. Masek, Biospheric Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA. (bounoua@ltpmail.gsfc.nasa.gov; jmasek@ltpmail.gsfc.nasa.gov)

Y. M. Tourre, MEDIAS-France, Centre National d'Études Spatiales, Bpi 2102, 18 avenue Edouard Belin, F-31401 Toulouse Cedex 9, France. (yves.tourre@medias.cnes.fr)

## Toward a new generation of satellite surface products?

F. Aires<sup>1</sup> and C. Prigent<sup>2</sup>

Received 3 April 2006; revised 18 September 2006; accepted 16 October 2006; published 11 November 2006.

[1] Despite the abundance and variety of remote sensing measurements, land surface characterization from satellite observations is still very challenging. The links between the three sources of surface information, namely the satellite observations, the in situ measurements, and the land surface model outputs, are complex. Innovative techniques have to be developed to merge these information sources and optimize the use of satellite measurements for better surface products and more predictability. Concepts such as multi-instrument/multiparameter retrieval algorithms are discussed, as well as the synergetic use of satellite observations, model outputs, and in situ data. The need for careful satellite calibration is stressed, and the scaling problem is emphasized. Recent results are reviewed to indicate what the land surface remote sensing problems are and how they might be attacked. Two concrete applications are presented: an “all weather” retrieval of surface skin temperature from combined microwave and infrared observations and a soil moisture analysis from the merging of multisatellite observations and land surface model outputs. This paper is intended to stimulate debates and collaborations between the land surface modelers and the satellite remote sensing community for the design of the next generation of land surface products.

**Citation:** Aires, F., and C. Prigent (2006), Toward a new generation of satellite surface products?, *J. Geophys. Res.*, *111*, D22S10, doi:10.1029/2006JD007362.

### 1. Introduction

[2] In addition to their key role in many applications (e.g., hydrology or agriculture), land surface satellite-derived products are very important in the framework of Land Surface Models (LSM): first as inputs, to initialize the models, to serve as boundary conditions, or for assimilation; second, to validate the model outputs in order to complement the in situ measurements that are spatially and temporally limited. Depending on their use in the LSM and depending on the model application (e.g., energy-water exchanges, biochemistry, ecosystem function), the set of necessary land parameters differs as well as their spatial and temporal requirements.

[3] The International Satellite Land Surface Climatology Project (ISLSCP), started in 1983, provides a large number of global gridded land surface data sets, related to land cover, hydrometeorology, radiation, and soils, over 10 years (more for some variables) with spatial resolution from 0.25° to 1° (see Sellers *et al.* [1995], Hall *et al.* [1995], and the other papers in this special ISLSCP issue). The available surface products are not all estimated from satellite observations, some being derived from meteorological reanalysis

(e.g., ECMWF or NCEP reanalysis) or from upscaling of in situ measurements (e.g., the Climate Research Unit near surface information or the GlobalView CO<sub>2</sub> and CH<sub>4</sub> products). Most satellite products directly derived from the ISLSCP activity focus on the vegetation and land cover characterization, using essentially visible and near infrared observations.

[4] A wealth of Earth satellite observations is now available, over long time series, covering the entire globe and providing a large diversity of information, from the visible to the microwave. However, land surface characterization from satellite observations is still very challenging:

[5] 1. The signal received by the satellite is generally the combination of contributions from different surface characteristics (vegetation, soil, soil moisture, snow, roughness, among others) and disentangling these various effects to quantify one variable is often very difficult. In addition, depending on the wavelength, the atmospheric contamination might need to be subtracted.

[6] 2. Limitations also come from the fact that no Radiative Transfer Model (RTM) for soil/vegetation/snow is satisfactory for global applications and for each wavelength range. Empirical relationships are often fitted locally for a given frequency range and their extension is questionable. In addition, even if such RTM existed, it would need a large variety of ancillary information that are not available at a global scale.

[7] 3. The spatial resolution of the satellite observations is not always compatible with the model scales, especially when local processes are involved. Within a given satellite field-of-view, the land surface parameters can exhibit large spatial variability (e.g., soil moisture or vegetation), making

<sup>1</sup>Laboratoire de Météorologie Dynamique, Institut Pierre-Simon Laplace/Centre National de la Recherche Scientifique, Université de Paris VI/Jussieu, Paris, France.

<sup>2</sup>Laboratoire d'Études du Rayonnement et de la Matière en Astrophysique, Centre National de la Recherche Scientifique, Observatoire de Paris, Paris, France.

it difficult to compare the satellite estimates with in situ point measurements or with model outputs. Aggregation and disaggregation techniques need to be developed in order to change scales.

[8] 4. Up to recently, there was no satellite optimized for the observations of key surface parameters like the soil moisture. This means that the designs of the satellite payloads are not optimal for the retrieval of these parameters in terms of frequency selection or spatial/temporal resolution and as a consequence, estimating these parameters is more difficult. The GRACE mission launched in 2002 [Tapley *et al.*, 2004] is the first mission specifically designed to quantify the terrestrial hydrological cycle, including the aquifer, the soil moisture, and the snow pack, through the gravimetric measurements of the vertically integrated water mass changes. The Advanced Microwave Scanning Radiometer–EOS (AMSR-E) is a passive microwave radiometer on board the Aqua satellite, launched in 2002. It observes atmospheric, land, oceanic, and cryospheric parameters, including soil moisture [Njoku *et al.*, 2003]. The European SMOS [Kerr *et al.*, 2001] is a soil moisture dedicated mission with measurements at low microwave frequencies (L-band) that will be launched in 2007.

[9] Figure 1 evidences the complex relationships between the three sources of land surface information: the satellite observations, the LSM, and the in situ measurements. The objective of this paper is to analyze these links and suggest methodologies to derive optimum land surface information from the combination of these three sources.

[10] Since the start of the ISLSCP, the needs of the land surface modelers in terms of satellite-derived products have grown dramatically. Entekhabi *et al.* [1999] proposed an agenda to answer the priority science questions in hydrology, and it included optimizing the use of remote sensing products. As an example, calibration of a model parameterization was traditionally performed over a given region and for a given period of time, using comparisons between one model output and the corresponding in situ measurements. There is today a demand to account for the full variability of the model dynamics, temporally and spatially, and simultaneously for different outputs of the model [McCabe *et al.*, 2005]. In addition, several models are now often intercompared (e.g., the Global Soil Wetness Project (GSWP) [Entin *et al.*, 1999]) and multimodel approaches tend to evaluate the model uncertainties by using outputs from different models [Murphy *et al.*, 2004]. In these cases, the role of the satellite observations could be to diagnose problems in one model (or to invalidate one model), not to validate a model: this difference in perspective can induce significant changes in the way the satellite/model relationship is considered.

[11] The three sources of land surface information have their own limitations. One model using different sources of inputs for the same variable can produce differences in outputs that are difficult to explain [e.g., Schaake *et al.*, 2004]. Even separating the effects of erroneous forcing from a lack of realism in the model is not obvious [Robock *et al.*, 1998]. Model intercomparisons have shown that even when fed with the same inputs, the model outputs can be significantly different (e.g., the Project for Intercomparison of Land Surface Parameterization Schemes (PILPS) [Henderson-Sellers *et al.*, 1995]). Although often consid-

ered as the “truth,” in situ measurements are by nature very local, are often labor intensive, and for the measurements of some variables the experimental protocol can vary from a campaign to the other with resulting biases. Satellite remote sensing is clearly the solution for large-scale and long-term calibration of the model as it encompasses all the variability of the land surface system. However, as already stressed, developments of the satellite retrieval is often difficult and if several surface parameters have to be examined simultaneously, consistency is necessary between the retrieved parameters.

[12] How to reconcile and merge the three sources of information for a better final product and for more predictability? How to optimize the use of satellite observations for that goal?

[13] 1. Satellite retrieval is very often an ill-posed problem. Retrieval schemes have to optimize the use of complementary observations and ancillary data. Merging of different satellite observations is an attractive option.

[14] 2. For an efficient use of the satellite information in the models, consistency has to be reached between the three sources of land surface information, as well as between the different satellite-derived parameters if several of them have to be used. The simultaneous use of the three spatially different information requires special attention to the scaling problem.

[15] 3. Each information source having its own errors, it is important to design techniques that can handle these uncertainties.

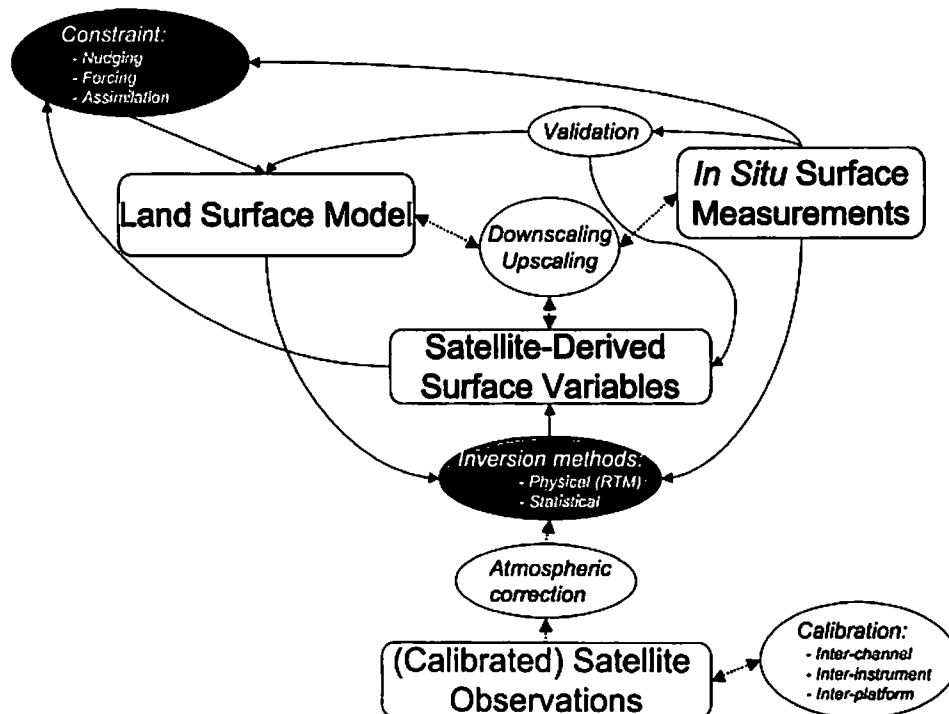
[16] 4. The satellite derived parameters have to have consistent characteristics globally and over long time series, especially for climate monitoring objectives, thus imposing strong constraint on the satellite intercalibration.

[17] These new challenges make it necessary to develop new strategies. In this paper, we will focus on the methodologies to put in place to optimize the combined use of satellite, model, and in situ data, and in particular, to optimize the use of satellite data in the framework of land surface modeling. We do not intend to examine each issue and to suggest solution for each one, but rather to suggest a few elements of reflection, describe few general strategies, and list original efforts to make progress.

[18] We will first discuss the different possibilities and describe some promising methodologies. The discussion will be limited to schemes that can be realistically implemented at a global scale. Second, examples of derived parameters using these types of techniques will be briefly presented in order to illustrate the concepts. Although these examples concern the hydrometeorology, similar methodologies can be applied to other parameters as well. The conclusion tends to describe, pragmatically, how to proceed, insisting on the necessity of a dialog between modelers and satellite product experts to optimize the use of satellite observations in the land surface modeling framework.

## 2. Retrieval Methodologies

[19] The radiation that impinges on the satellite is often the result of several contributions from the land surface. Even for a homogeneous field-of-view, the satellite will receive radiation from the soil, the vegetation, and potentially the snow. The problem is even worse for a heteroge-



**Figure 1.** Schematic representation of the complex links between the three sources of land surface information.

neous pixel. Depending on the wavelength, contamination from the atmosphere (gases, clouds, or rain) might also interfere with the signal and needs to be accounted for by using ancillary atmospheric data sets from satellites, models, or mixed outputs (e.g., reanalysis). Retrieval of land surface parameters from the satellite measurement is thus often an ill-posed problem that can require multiple and independent measurements as well as quality a priori information in order to disentangle the mixing effects in the observations.

[20] There are different solutions to retrieve one or multiple land surface parameters, using one or several instruments from one or different satellites at the same time. The various solutions are now discussed, from the simplest to the most complex ones.

## 2.1. Satellite-Only Methods

### 2.1.1. One Instrument/Multi-Instrument

[21] Methodologies are developed that use one type of wavelengths, measured on board one satellite to derive a single land surface parameter. In order to suppress ambiguities related to the contribution from other surface parameters, these algorithms exploit the complementarity of close frequency bands, of various incidence angles or the temporal information available from consecutive measurements. The Normalized Difference Vegetation Index (NDVI) for instance combines visible and near infrared observations to isolate the effect of the vegetation photosynthetic activity on the absorption in the visible and limit the contribution from the soil [e.g., *Tucker, 1979; Tucker et al., 1985*]. The soil moisture index developed by *Wagner et al. [2003]* from the ERS scatterometer observations at 5.25 GHz capitalizes on

the multiangle observations and the temporal evolution of the signal to subtract the vegetation effect.

[22] These methods have the advantage to provide parameters that are independent from other sources of information, for potential comparison with other estimates. However, the apparent simplicity of the algorithm should not mask specific difficulties, like the treatment of the atmospheric contamination in the case of the NDVI [*Gutman, 1999*]. In addition, these methodologies can suffer from their limited spectral range. This can translate into saturation effect: for instance, although the NDVI offers a good sensitivity over crops and grasslands, it tends to saturate over denser vegetation types. The retrieval can also experience interference with other surface parameters. As an example, the estimation of snow depth from the differences of two passive microwave channels is hampered by contamination from vegetation as well as by the snow metamorphism during the winter season [*Kelly et al., 2003*]. Since the variations of the contaminating variables are treated implicitly as random noise, they affect the quality of the retrieved quantity.

[23] A strategy to avoid such deficiencies consists in gathering satellite observations in different wavelength ranges that have different sensitivities to the various surface parameters. The synergetic use of satellite observations helps separate the contribution of the various parameters. For example, in clear sky condition, passive microwave signal in window channels depends on both the surface temperature and the surface emissivity: another source of information like thermal infrared measurements can help untangle the contribution of these two parameters [*Aires et al., 2001*].



[24] The sensitivity of the different satellite observations for a given parameter has to be investigated, to analyze their complementarity, and to assess the ability of combinations of these satellite measurements for the estimation of the parameter. A detailed information content analysis has to be performed, on a global basis, to select the optimum satellite information to be merged.

[25] The estimation of one parameter using several wavelength ranges can involve one satellite only if the various instruments are on the same platform. This could be the case for precipitation retrievals that would use the synergy between the radar and the radiometer on board the Tropical Rain Measuring Mission (TRMM). In the case of multiplatforms, the appropriate time window has to be selected to allow meaningful merging of the information. In both cases, the satellite measurements from different instruments have to be collocated in space, with potential problems related to the various spatial resolutions.

[26] An additional benefit of satellite data fusion is that the retrievals are more robust to noise since they use more information. They can also be less sensitive to missing data in one sensor. Furthermore, it can help fill up temporal and spatial gaps. This is the case when infrared observations from geostationary satellites are used to complement the rain estimates from microwave instrument on board polar satellites like in GPCP [Adler *et al.*, 2003]. It can also help “calibrate” one retrieval with the other: this is the case when the passive microwave rain retrieval is used to “calibrate” the rain estimates from infrared measurements.

[27] A multi-instrument retrieval scheme in this paper refers to an algorithm that uses simultaneously or hierarchically the observations of two or more instruments in order to benefit from the instrument synergy. The a posteriori combination of retrievals derived independently from each instrument does not constitute a multi-instrument retrieval since there is no benefit from the synergy of instruments in this case.

### 2.1.2. One Parameter/Multiparameter

[28] For the retrieval of a surface parameter (e.g., the soil moisture), it is often necessary to use an auxiliary information (such as surface temperature). This auxiliary information can be derived from another satellite information, and no control is possible on the coherency between these two quantities. Independent and inconsistent calibration or assumptions can be made for the two retrievals. These various parameters put together can lack coherency, or contrarily be too much dependent, because they are derived from a very limited number of observations. Even if the independence of the retrieved parameters was satisfactory and would allow for the intercomparison of the retrieved products, it is clear that the next step would consist in using together multiple satellite observations and benefit from their synergy in the retrieval of multiple surface parameters. One major advantage of merging satellite observations for multiple parameter retrieval is that the various retrieved parameters are coherent. This means that the same set of assumptions is adopted for the two retrieved parameters, that these assumptions are controlled and known. This also constrains part of the incoherencies among parameters (incompatible surface temperature and soil moisture for instance).

[29] Once the satellite observations are merged, in spatiotemporal collocation, two basic strategies can be used: the retrieval of the different parameters can be hierarchical or simultaneous. In a hierarchical scheme, a major surface parameter (e.g., the surface temperature) is first retrieved by using all available satellite observations. Then, this retrieved parameter is used together with the satellite measurements for the subsequent retrieval of another variable (e.g., soil moisture). This approach uses the fact that some surface parameters are dependent on others and that some specific retrieval algorithms need to follow this dependency structure. As an example, within the International Satellite Cloud Climatology Project (ISCCP) [Rossow and Schiffer, 1999], using IR observations from both geostationary and polar satellites, the cloud presence is first analyzed and then, for clear scenes, the surface skin temperature is estimated. An inconvenience of this scheme is the cascade of errors in the suite of retrieval algorithms. The other approach consists in performing the multiparameter retrieval at the same time in the algorithm. One advantage of this strategy is that the uncertainty characterization is easier. The solution is determined simultaneously for each parameter (called “global solution”). This is preferable from an optimization point of view than a solution built piece by piece. The optimization process might be more difficult, but it is easier to obtain a compromise choosing a solution that satisfies all the satellite observations.

[30] A multiparameter retrieval scheme refers to the simultaneous or hierarchical retrieval of two or more parameters using a single algorithm in order to benefit from the coherency among the retrieved variables. Again, the a posteriori association of various parameters retrieved by independent algorithms does not constitute a multiparameter retrieval because the dependence among retrieved parameters is not exploited.

### 2.1.3. Illustration With a Synthetic Example

[31] To illustrate the two concepts of multi-instrument and multiparameter retrievals, a highly idealized example is presented here that uses information content ideas [Tarantola, 1987]. The general principles are developed in the classical references [Twomey, 1977; Rodgers, 1976, 2000]. We use a bivariate/two-measurement linear case (or a linearized nonlinear model):

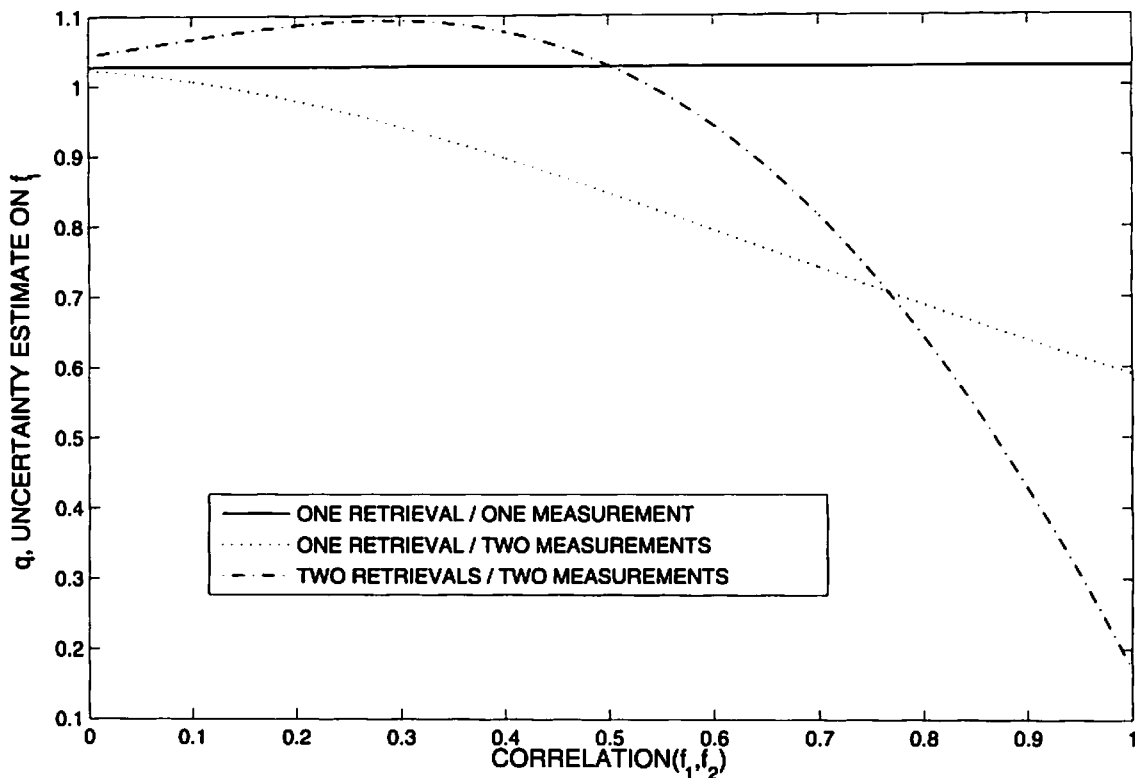
$$F = A \cdot f + e \quad (1)$$

where  $f = (f_1, f_2)$  is the state vector composed by two geophysical variables (with covariance matrix  $S_f$ ) that we intend to retrieve from the measurements  $F = (F_1, F_2)$ ,  $e = (e_1, e_2)$  is the measurement noise (with covariance matrix  $S_e$ ) and  $A$  is the forward model (linear in this case) that needs to be inverted.

[32] A Bayesian solution to the inverse problem is, under Gaussian hypothesis:

$$f = f_g + \left[ A' \cdot S_e^{-1} \cdot A + S_f^{-1} \right]^{-1} \cdot A' \cdot S_e^{-1} \cdot (F - F_g) \quad (2)$$

where  $f_g$  is the first guess for  $f$  to which is associated the measurement  $F_g$ ,  $A'$  is the transpose of matrix  $A$ ,  $S_e$  is the covariance matrix of the measurement errors  $e$ , and  $S_f$  is



**Figure 2.** Sensitivity of the error estimate to  $cor(f_1, f_2)$ , the correlation between the two geophysical variables to retrieve: for one retrieval and one measurement (solid line), for one retrieval and two measurements (dotted line), and for two retrievals and two measurements (dot-dashed line).

the covariance matrix of the first guess errors [Chédin *et al.*, 1985]. The uncertainty matrix on this retrieval is:

$$Q = [A' \cdot S_e^{-1} \cdot A + S_f^{-1}]^{-1} \quad (3)$$

For illustrative purpose, we use the following numerical values:

$$S_f = \begin{pmatrix} 3 & cor(f_1, f_2) \cdot \sqrt{3}\sqrt{4} \\ cor(f_1, f_2) \cdot \sqrt{3}\sqrt{4} & 4 \end{pmatrix}$$

$$S_e = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

and

$$A = \begin{pmatrix} 0.8 & coef(f_1, F_2) \\ 0.1 & 0.9 \end{pmatrix}$$

The values of  $S_f$  and  $S_e$  impose that  $coef(f_1, F_2) = [cor(f_1, f_2) - 0.08]/0.9$ . Only  $cor(f_1, f_2)$ , the correlation between the two geophysical variables  $f_1$  and  $f_2$ , is now varying.

[33] This formulation allows to study the sensitivity of  $q = Q(1,1)$  (i.e., the uncertainty estimate on  $f_1$ ) to the parameter  $cor(f_1, f_2)$ . In Figure 2, three curves for three different retrieval configurations are represented:

[34] 1. In case 1, only one observation  $F_1$  and one retrieval  $f_1$  are considered in the top curve (continuous line). Since the measurement  $F_2$  is not used here for the retrieval

of  $f_1$ ,  $coef(f_1, F_2)$  has no impact on  $q$  and the curve is flat with an uncertainty level at:

$$q = (0.8 \cdot 1^{-1} \cdot 0.8 + 1/3)^{-1} = 1.0309$$

[35] 2. In case 2, a unique variable  $f_1$  is retrieved but the two observations  $F = (F_1, F_2)$  are used in the dotted curve. As expected, the higher  $cor(f_1, f_2)$  is, the higher  $coef(f_1, F_2)$  is, and the lower the uncertainty  $q$  becomes.

[36] 3. In case 3, the dotted/dashed curve is for the full problem that considers the two measurements  $F = (F_1, F_2)$  and the two variables to be retrieved  $f = (f_1, f_2)$ . In this configuration, again, the higher  $coef(f_1, F_2)$  is, the lower the uncertainty becomes. Furthermore, it can be noticed that when  $cor(f_1, f_2)$  is lower than about 0.6, retrieving only  $f_1$  gives better results than retrieving both geophysical variables  $f_1$  and  $f_2$ , meaning that the retrieval of  $f_1$  is hampered, part of the information  $F = (F_1, F_2)$  being “diverted” to retrieve  $f_2$  which reduces the quality of the  $f_1$  retrieval. When  $cor(f_1, f_2)$  is higher than 0.6, the uncertainty on  $f_1$ , using the same measurements  $F = (F_1, F_2)$ , can decrease dramatically and make a significant difference with case 2. This improvement is only due to the simultaneous retrieval of the second geophysical variable  $f_2$ . The explanation is that  $f_1$  being correlated to  $f_2$  and the second measurement  $F_2$  being a  $f_2$  information-carrier, the retrieval of  $f_1$  benefits from this indirect relationship.

[37] This example illustrates well how information about the dependencies among the geophysical variables can be extremely beneficial for the retrieval.

[38] It has been shown that multi-instruments cannot bring improvements in some situations (no synergy) but that it will not degrade results. On the contrary, the use of multiparameter techniques can be detrimental (when there is independence of the retrieved parameters) but is extremely beneficial in some cases (strong correlations among geophysical variables). No dogma in the use of such or such methodology should be imposed: each application has its specificities and the needs have to be carefully analyzed to design the optimum retrieval scheme. The good news is that with the kind of simple tools used in this section, we can quantify the possible synergies and therefore the type of approach to be used.

[39] The example provided here is a linear model, or the linearization of a nonlinear model. It should be pointed out that the linear methods could be insufficient because the relationships can be highly nonlinear among the surface variables. This is particularly true with surface problems that undergo high discontinuities, with threshold effects and very different interactions depending upon the situations. This means that the synergy can be different in some situations and the use of nonlinear techniques is a plus.

#### 2.1.4. Technical Implications for the Multivariate Case

[40] In this section, we comment on some technical implications arising when multivariate strategies are used. The major requirement for this kind of approaches is the exploitation of the dependencies between the measurements, that need to be coherent, and between the geophysical variables to be retrieved.

[41] We first consider multichannel, multi-instrument or multiplatform algorithms. Such retrieval strategies need to use coherent measurements. Is it possible to rely on RTMs to elaborate physically based retrieval algorithms? The radiation emerging from the surface is sensitive to a large number of surface parameters, through intricate mechanisms for each surface type. For instance, even for a bare soil in the microwave, the response will depend upon the surface roughness (at small and large topographic scales), upon the dielectric properties of the medium (related to the soil composition, texture, and humidity), and potentially upon volume scattering below the surface. In addition, it is also related to the observation characteristics, its wavelength, its incidence angle, and its polarization among other parameters. The interactions between the surface and the radiation are complex to model, being dependent on a large number of highly variable and difficultly accessible parameters. Efforts have been made to better understand the mechanisms responsible for the interaction between the surface and the radiation, from both theoretical analysis and field experiments, at all the wavelength ranges of interest. Ground or airborne based measurements have also been performed to help develop the RTM. The resulting RTMs, even the most elaborated ones, still have difficulties to perform well for all environments at the global scale, although they can be efficient for the specific conditions for which they have been tuned. Even assuming that a perfect land surface RTM exists for a given wavelength range, would the inputs it will require (e.g., soil composition, texture, surface roughness, vegetation density, geometry, water content) be available on a global basis with a resolution compatible with the satellite one and with the

required accuracy? The problem is even more complex when using multiple instruments that cover different wavelength ranges: the RTM should be coherent and perform identically well over the full frequency domain. Even for the same wavelength range, it can be difficult to simulate the radiative transfer with the same accuracy when combining active and passive modes. For instance, there are very few RTMs developed to simulate both the active and passive responses of the surface in the microwave for a joint analysis of microwave radiometer and radar observations.

[42] Most classical inversion techniques are able, and were actually designed to deal with multiparameter retrievals. In equation (2), an inversion formula was given, describing the Bayesian estimator with Gaussian hypotheses. Classical variational assimilation or iterative methods are also based on the same estimator. Neural Network (NN) schemes are technically different but are based on the minimization of similar quality criteria, so the essence of the estimation is very similar, and the same type of a priori information is required (e.g., measurement characteristics, statistics on the variable dependences). Each technique needs to use an information describing the dependencies among the physical variables to be retrieved. (First guess error dependencies are also very important, for example in variational assimilation [Rabier *et al.*, 1998].) When a physical relationship linking two or more of the variables exists, it is extremely interesting to use it in the retrieval process. For example, during the retrieval of atmospheric water vapor, if one instrument provides information on the total vertical content and another instrument gives a profile description, then the total content can constrain the profile retrieval. This type of physical constraints can be used in each step of the iterative inversion methods or it can be added as an additional strong constraint in the quality criterion used by statistical techniques. Most of the time, however, the dependencies are described statistically. In the estimator of equation (2) (Bayesian, iterative, or variational assimilation), it is given by the covariance matrix  $S_f$ . For neural networks, the learning data set describes implicitly the various complex relationships and the NN disentangles them during the learning stage to reproduce them adequately. This approach could appear easier to use because no hypothesis is formerly required a priori. In reality, extreme care needs to be exercised during the development of the learning data set, with sometimes the use of complex a priori information. For example, in the work by Aires *et al.* [2001] the distributions of liquid water cloud top temperature are carefully analyzed and then used in the development of the learning data set.

[43] Finally, as already mentioned in the previous section, the relationships between the geophysical variables can be dependent on the situation. This is referred to as the nonlinearity of the a priori relationship and the retrieval scheme has to be flexible enough to account for it. A NN inversion is by nature designed to deal with this difficulty. For classical approaches, a sensitivity analysis of the a priori information must be conducted. For example, the first guess errors on specific humidity are dependent on the atmospheric situation and for atmo-

spheric temperature, a different covariance matrix is used for each  $10^\circ$  latitude bands [Rabier *et al.*, 1998].

## 2.2. In Situ Measurements of Land Surface Model Outputs to Help Satellite Retrievals

### 2.2.1. In Situ Measurements

[44] The complexity of surface parameter retrieval together with the inadequacy of current surface RTMs make the use of in situ measurements an attractive source of additional information.

[45] First, in situ measurements are essential to understand at a local and detailed scale how the parameters interact with the radiation. This is generally done during well-documented measurement campaigns that combine in situ measurements of the land parameters with coincident observations from ground-based or airborne instruments (e.g., radiometers, radars, lidars). However, this analysis can also be performed with coincident in situ and satellite measurements: it enables comparisons over longer time series and more diverse environments, provided that a large data bank of consistent in situ measurements exists for the variable of interest. This exercise has been recently conducted for the surface soil moisture using the Global Soil Moisture Data Bank [Robock *et al.*, 2000] and a suite of satellite observations [Prigent *et al.*, 2005a], as well as for the snow depth using a large collection of in situ measurements in the Northern Hemisphere in coincidence with satellite measurements [Cordisio *et al.*, 2006].

[46] Second, in situ measurements can also be exploited for the parameterization of the retrieval algorithms. Many surface retrieval algorithms are actually based on the parameterization (also called calibration) of regression models between the satellite observation and the surface variable of interest. When performed on a reduced number of locations during a restricted period of time, the validity of the parameterization for other conditions is questionable. For instance, when developing a soil moisture algorithm, the large-scale variability of the vegetation can be misrepresented by a parameterization performed on a local basis. For a retrieval algorithm to be valid globally, its parameterization needs to use as diverse as possible in situ measurements to sample local and large-scale variabilities. The parameterization data set should account for the spatial and temporal variability of the parameters in the satellite observations. Note that depending on the variable, the relevant temporal scale can span from the hour to several years.

[47] Finally, in situ measurements are necessary for the evaluation of the retrieval algorithms. This imposes that independent measurements are available for the parameterization of the algorithm and for its evaluation. A portion of the in situ measurement data set is used to parameterize the retrieval algorithm (i.e., the learning data set) and the remaining part can be exploited to validate it (i.e., the testing data set).

[48] One essential difficulty in using in situ measurements is related to the scaling problems. Models and satellite measurements have different spatial resolutions, which makes difficult the direct comparison with point in situ measurements. The link has to be found between the two scales (i.e., local measurements and large pixels from models or satellite observations) (see section 3.3).

[49] In addition, except for well-organized measurement campaigns [e.g., Leese *et al.*, 2003], the large majority of in situ measurements performed regularly by multiple institutions is very difficult to access, and can show very different accuracy. An effort has to be done to create homogeneous and quality-controlled collections of the key surface variables that are not part of the weather station routine measurements. Entekhabi *et al.* [1999] already insist on this necessity for hydrological purposes. The Global Soil Moisture Data Bank [Robock *et al.*, 2000] is such a database for the soil moisture and similar efforts should be encouraged for other variables.

### 2.2.2. Land Surface Models

[50] More and more studies couple LSM outputs and satellite observations to help solve the retrieval of land surface parameters by using the relationships between surface variables and satellite observations. The inversion problem being ill-posed, the LSM can provide auxiliary information that helps constrain the retrieval.

[51] Outputs from a LSM can be directly adopted as auxiliary input information to the retrieval algorithm. For instance, a snow depth retrieval scheme could require a LSM derived surface temperature information as input [Boone *et al.*, 2006].

[52] Differently, LSM outputs can be used to generate databases from which a retrieval algorithm is derived. For example, Lakshmi *et al.* [1997] adapt a LSM and couple it to a RTM to simulate soil moisture and the corresponding microwave responses that could be the basis for a retrieval algorithm. In the work by Aires *et al.* [2005], the link between the modeled surface variables and the satellite observations is statistical: a database of coincident multi-satellite observations and soil moisture model outputs is created and an updated estimation of the soil moisture is deduced. In these cases, the retrieval is clearly linked to a specific LSM.

[53] The LSM can also be part of an iterative least square inversion schemes. For instance, Ramillien *et al.* [2005] develop such a method based on the LaD model [Milly and Shmakin, 2002] to separate the contribution of the various water reservoirs (surface water, soil moisture, ground-water, and snow pack) that are vertically integrated within GRACE gravimetric measurements.

[54] The important message is that the choices should be clearly stated so that the user of the retrieved products knows about the different assumptions and about the potential links between the variables, other instruments, and other variables. Otherwise, satellite retrieved parameters could be illegitimately used to validate LSM from which they are not independent.

## 2.3. Constraining Land Surface Models With Satellite Data

[55] Methodologies have to be developed to combine optimally satellite observations and land surface schemes in order to produce better final products and more predictability. As already mentioned, it is very difficult to simulate the satellite observations directly from the surface parameters (e.g., soil moisture, vegetation, snow) using a RTM. As a consequence, traditional direct assimilation of raw satellite observations is difficult. Satellite remote sensing is a solution to constrain the model for



one or several variables over large-scale and long time periods. The retrieval parameter can be assimilated as a state variable in the model, or as a geophysical variable linked physically to the state variables of the model (i.e., wind, temperature, humidity, surface fluxes for soil moisture retrievals [Mitchell *et al.*, 2004]).

[56] The recent interest for techniques that evaluate simultaneously different outputs predicted by the model [Gupta *et al.*, 1999; Franks *et al.*, 1999; Beven and Freer, 2001; McCabe *et al.*, 2005] is driven by the fact that the calibration of one variable only can bias the model toward that variable while the others are not well reproduced. Multiobjective calibration tends to evaluate the set of model parameterizations that best reproduces multiple outputs. However, several parameter combinations can reproduce similarly different sets of one output (the “equifinality” as defined by Beven and Freer [2001]). Likelihood methods such as the Generalized Likelihood Uncertainty Estimation (GLUE) are developed within the LSMs to select the parameterization that fits the observed parameters better [McCabe *et al.*, 2005].

[57] Variational assimilation is a particular technique to perform the assimilation of observations (satellite observations, weather station measurements, radiosondes) into a numerical model [Ide *et al.*, 1997]. Kalman filtering is another technique that emphasizes the sequential assimilation in forecasting problems. In the work by Reichle *et al.* [2002a], soil moisture is assimilated using the ensemble Kalman filter, an extension that uses ensembles of runs to estimate the error covariances instead of propagating them theoretically which is computationally intensive. An experiment in similar conditions is conducted by Reichle *et al.* [2002b] to test the extended Kalman filter designed for nonlinear problems. Nudging is also a possible way to constrain models with observations [Mitchell *et al.*, 2004].

[58] A simpler approach consists also in using the satellite remote sensing soil moisture to initialize climate model simulations [Walker and Houser, 2001]. Other more specialized methods have been developed. For example, in the work by Cordisio *et al.* [2006] a surface model is calibrated by using in situ snow depth measurements from a network of stations.

[59] For use in LSM, the absolute requirements on the satellite retrievals in terms of accuracy or spatial and temporal resolutions are often unclear (see the analysis for soil moisture in the work by Walker and Houser [2004]). However, there is a clear demand for consistency between the satellite retrievals and the model variables, prior to the use of the satellite retrievals within the model. Berg *et al.* [2003] insist on the impact of potential biases in data sets to force models and recommend the implementation of bias reduction scheme to reduce the associated errors. Reichle *et al.* [2004] compare soil moisture estimates from Scanning Multichannel Microwave Radiometer, modeled soil moistures, and in situ measurements, for nine years all over the globe. The time averaged fields from the model and the satellite agree well but the magnitude and variability of the soil moisture estimates are very different. Local bias correction or rescaling have to be performed before assimilation of the satellite data into LSMs: tuning the local statistics of the satellite retrievals to the model ones can be a solution.

[60] Classical variational assimilation in surface models suffers from limitations. First, surface parameters can have a very strong spatial inhomogeneities with strong discontinuities. Second, relation between surface parameters and observations can be highly nonlinear. Lastly, no RTM dedicated to surface is fully satisfactory. For all these reasons, we believe that special techniques need to be developed when using observations into a surface model. Instead of the traditional approach that assimilates directly the satellite observations, we suggest to use the inverse model derived from a remote sensing algorithm. This has various benefits: (1) It avoids the estimation of the Jacobians of the RTM, (2) it does not add up uncertainties from the forward model to the uncertainties from the Jacobians, and (3) it makes it possible to work directly with the surface state variables that are more directly related to the numerical model.

### 3. Ancillary Data Processing

[61] Several processing steps are often necessary, especially when homogeneous long time periods of accurate global products are expected, derived from multisatellite approaches. Although sometimes neglected and often strenuous to perform, these processings are crucial if quality land surface parameters are required. It is particularly important to quantify the errors associated to these treatments (or their absence) and to document them well.

#### 3.1. Satellite Calibration

[62] A reliable instrument calibration is a prerequisite for any remote sensing algorithm. The calibration has to be stable over long time period and free from any biases. For instance, Colton and Poe [1999] performed significant intercalibration work on the series of SSM/I microwave instruments. Even if only one type of instrument is involved, satellite drift within the life span of one given satellite and satellite intercalibration between successive satellites of the same family can be very difficult to achieve. The analysis of the AVHRR NDVI over long time series suffers from these difficulties [Gutman, 1999]. In addition, when trying to cover the whole globe and the full diurnal cycle, the simultaneous use of several satellite types is necessary and stringent constraints are then imposed on the satellite intercalibration. For instance, the ISCCP [Rossow and Schiffer, 1999] that supplies global cloud information every three hours, combines all the visible and infrared observations from the NOAA polar orbiters and the geostationary satellites to provide both the spatial and temporal coverages. A huge effort has been dedicated to the accurate radiance calibration and intercalibration of all the instruments over the long time series to archive the ISCCP results [Brest *et al.*, 1997].

[63] In addition to a detailed and systematic analysis of each sensor calibration from an engineering point of view, various methodologies have to be developed to intercalibrate the observations. In order to perform multiparameter retrievals from multiple satellite observations, the calibration of the instruments needs to be consistent at several levels. First, the calibration has to be consistent among the various channels of a given instrument. From an information theory point of view, it is not optimal to calibrate

independently the various channels (essentially by monitoring the mean and standard deviation over the time, or by indirect comparison with in situ measurements). The covariance among channels can be exploited to perform the calibration and this potential should be examined. At least, monitoring the interdependence among the channels can help check the calibration quality. However, caution has to be exercised in this procedure as too stringent constraints could mask some real extreme behavior in one channel. Second, the intersatellite calibration has to be performed and consistency has to be ensured across platforms. Observations from coincident overpasses can be compared, provided that the measurements are performed exactly in the same conditions (frequency, incidence angle, polarization); otherwise, products retrieved from the two coincident satellite measurements can be compared. Rigorous statistical comparison of RTM calculations and satellite observations can also be performed, with the same assumptions for the two satellites, in order to diagnose the radiance biases between instruments. The retrieved products across satellites can also be compared.

[64] A generic calibration method that includes many of these above aspects is under development at NOAA/NESDIS to intercalibrate radiometer in the visible, infrared, and microwave [Weng *et al.*, 2005]. It has already been applied to MSU. In the Global Precipitation Mission framework, efforts are also conducted to intercalibrate the passive microwave imagers on a common standard to ensure consistency among precipitation products [Kummerow *et al.*, 2001; see also <http://mrain.atmos.colostate.edu/LEVELIC/index.html>].

[65] Having together all the satellite observations, and using the same methods and data to calibrate them would inevitably benefit the retrieval. For instance, it is worrying to realize that there is not one uniformly calibrated set of SSMR and SSM/I observations over the full period of satellite operation that is easily available to the community. A strong effort, with dedicated funding, should be supported by the agencies to systematically apply these new calibration approaches. The resulting multiplatform calibrated data sets should be easily accessible to the community and would stimulate the developments of the next generation of retrieval algorithms.

### 3.2. Subtracting the Atmospheric Contribution

[66] Depending on the wavelength, the satellite observations can be contaminated by the atmospheric contribution. At low microwave frequencies, the atmospheric effect is negligible, making this wavelength range particularly suitable for land surface analysis (e.g., the ERS scatterometer data at 5.25 GHz [Prigent *et al.*, 2001, 2005b]). For other wavelength ranges, even if this effect is limited, it can modulate the received signals both in time and space and be mistakenly interpreted in terms of land surface variations. Suppressing the atmospheric signal is particularly important when analyzing the interactions between the land surface and the atmosphere. Different techniques are adopted to eliminate this effect. For instance, the traditional NDVI processing relies on the selection of the maximum value of the NDVI for a location for a given period of time to minimize the atmospheric contamination [Holben, 1986]. Physical methods based on radiative transfer calculations

have also been developed. In the ISCCP analysis the surface skin temperature is retrieved from clear IR radiances using the TOVS products to specify the atmospheric temperature and humidity profiles and a RTM to calculate the atmosphere radiative contribution. The passive microwave observations up to 37 GHz are adopted in a large number of studies to provide information on soil moisture [Vinnikov *et al.*, 1999], snow [Kelly *et al.*, 2003], or floods [Sippel *et al.*, 1998]: usually, the brightness temperatures are directly used in the algorithms, assuming that the atmospheric effect is negligible. However, as noted by several authors [e.g., Kerr and Njoku, 1993], atmospheric effects, especially cloud cover, may be responsible for a large part of the 37 GHz signal, casting doubt on the interpretation of simple brightness temperature combinations solely in terms of surface properties. In addition, although very sensitive to some surface parameters like the snow [Cordisco *et al.*, 2006], the 85 GHz channel that is present on the SSM/I or TMI is often not used because it is deemed too contaminated by the water vapor absorption. In order to avoid such limitations, some work has been conducted to extract from the passive microwave observations the information that is directly related to the land surface, its emissivity, by removing the contributions from the atmosphere, clouds, and rain using ancillary satellite data, atmospheric profiles from meteorological reanalysis, and a RTM [Prigent *et al.*, 1997, 2006].

[67] This preprocessing step is often strenuous, as it can involve significant amount of coincident ancillary information as well as complex RTM to estimate the atmospheric contribution. However, it is necessary if unambiguous surface information is required, with good accuracy.

### 3.3. Scaling

[68] Spatial and temporal scales of observed or measured surface variables cover a wide range. General principles of upscaling (i.e., aggregation (when the scale change is performed in the same variable, which is not always the case) or downscaling (i.e., disaggregation) have been developed to link one scale to another but the definition of such general terms is sometimes confusing. In this paper, we simply refer to the “downscaling” (respectively “upscale”) as a technique that increases (respectively decreases) the spatial resolution of the original data.

[69] “Regionalization” [Von Storch *et al.*, 1993] is one application of the downscaling: it makes it possible to describe the specific behavior of a region by combining large-scale outputs from a climate model with small-scale information from the surface. It is not directly of interest in the study. Another application of the down/up-scaling is the simple scale change of a particular field with the same variable in both scales: the terms “aggregation” for “upscale” and “disaggregation” for “downscaling” could be used for clarity purpose. This (dis)aggregation is performed for various reasons. For example, the downscaling from a satellite estimate or a climate output to a local measurement allows for the comparisons of the sources of information and can be used for validation or the calibration of models and retrieval algorithms. Techniques for (dis)aggregation require some knowledge of the scale spectrum of variation magnitude.

[70] In dynamical/physical approach, a physical model describes the dynamics of the system at the regional

resolution of interest. The physical model is then run under some constraints coming from the other information scale. This can be done through simple forcing, nudging, or (variational) assimilation. The inconveniences of this approach are twofold: such physical model is not always available and it can be computationally prohibitive.

[71] Statistical methods try to infer cross-scale relationships using a random or a deterministic function. The relationship is deduced from a data set of pairs of cross-scale samples, derived from observations (empirical scale change) or from model outputs (model-based scale change). A statistical technique (e.g., singular value decomposition, canonical correlation analysis, regular regression, classification, or neural networks) is “trained” to reproduce the link between each pair. Different approaches can be used: (1) The weather generator is a stochastic model defined to describe the evolution of the (dis)aggregated field. For example, a Markov chain can be used, allowing for temporal coherency in the field. This approach is essentially a complex random generator that depends upon cross-scale information. The sample data set is used to calibrate the stochastic model. (2) In the weather typing approach, weather regimes are defined (by using the data set of samples) and a classification algorithm links one scale-field of the weather regimes to the other scale in a deterministic or stochastic way. (3) In the transfer function method, a deterministic or stochastic model is defined to perform the cross-scale transfer, using a linear regression technique (e.g., singular value decomposition, canonical correlation analysis, regular regression) or a nonlinear one (e.g., artificial neural networks [Cavazos, 1999], kriging [Biau *et al.*, 1999]). Weather typing could also be included in this general approach when considering the categorization of the transfer function response. Other techniques are interpolation schemes (e.g., cubic spline, geostatistics).

[72] Statistical methods are computationally inexpensive and very flexible so that they can be adapted to specific situations. However, their extension to unobserved conditions is questionable (for global warming experiments or for regions not yet observed).

[73] The data set used to parameterize the statistical models has to describe correctly all the relationships among the variables. Spurious results such as artificial correlations can be misleading. Furthermore, small-scale heterogeneities relevant to the (dis)aggregation must be represented. For example, a few ground stations inside a large-scale cell can help estimate this heterogeneity. When this data set is available, the (dis)aggregation process uses explicitly or not this heterogeneity information. Some methods require a statistic-only information (geostatistics use a variogram (covariance of a same variable but at two varying distant locations) [Wackernagel, 2003]) but this information might not represent well enough the spatial nature of the information, in relation with surface properties (vegetation, elevation, geology of the soil, etc.). In this case, very fine geographic information must be extracted from the spatial patterns in the observed or modeled output data sets. The limitation of this approach is that the fine spatial information is region-dependent and cannot be extrapolated to other regions. Depending on the particular problem to solve and the quality of the available data set, a compromise will be

found between using very fine-scale information specific to a region and more widely applicable statistic analysis.

#### 4. Examples of Retrieved Surface Parameters Using Some of These Methods

##### 4.1. Surface Skin Temperature: An Example of Multisatellite/Multiparameter Retrieval

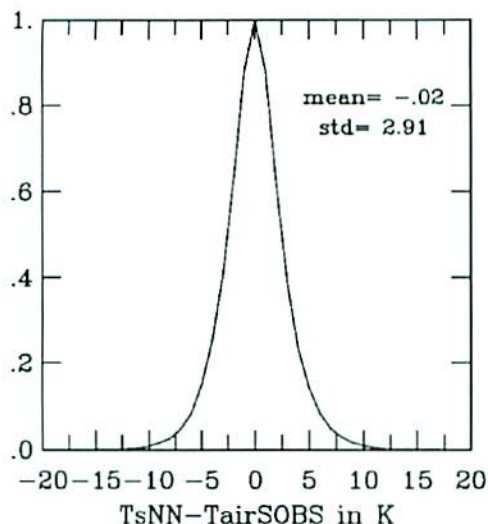
[74] The surface skin temperature and the soil moisture largely control energy and water exchanges at the land-atmosphere interface. The skin temperature represents the soil skin temperature for bare soil, the canopy surface temperature in densely vegetated region, and an average of the above for sparse vegetation. Measurements of the skin temperatures are required to study the energy and water exchange processes at the land-atmosphere interface, with time resolution high enough to resolve the diurnal cycle under all synoptic conditions, and covering a long enough period to examine how different seasonal and interannual conditions affect them.

[75] In situ surface skin temperature can be calculated from observations with an infrared radiometer, if the land surface emissivity is known: this measurement is not performed at weather stations and is not part of the conventionally measured data. Skin temperatures have been estimated from satellite infrared observations, with the limitation that cloud free observations are required (i.e., infrareds are blocked by clouds). Microwave land surface skin temperature retrieval is a very promising complement to infrared estimates, with the significant advantage that it can be effective even under cloudy conditions. However, because of the larger surface emissivity variations in the microwave than in the infrared, a combined analysis is required to isolate the temperature variation accurately.

[76] The analysis of microwave observations over land to determine atmospheric and surface parameters is still limited because of the complexity of the inverse problem. A Neural Network (NN) having the particularity of using first-guess information has been developed in [Aires *et al.*, 2001] to retrieve simultaneously the surface skin temperature, the integrated water vapor content, the cloud liquid water path, and the microwave surface emissivities between 19 and 85 GHz over land from SSM/I observations. The simultaneous retrieval of all these quantities improves the results for consistency reasons. A database to train the NN has been calculated with a RTM and a global collection of coincident surface and atmospheric parameters extracted from NCEP reanalysis, from ISCCP data, and from microwave emissivity atlases previously calculated. The results of the NN inversion are satisfactory. The theoretical r.m.s. error (based on radiative transfer simulations) of the surface temperature ( $T_s$ ) retrieval over the globe is 1.3 K in clear sky conditions and 1.6 K in cloudy scenes. Similar methodology has also been applied with success over snow and ice regions [Prigent *et al.*, 2003b].

[77] The thorough evaluation of the retrieved product is difficult. In the absence of routine in situ surface skin measurements, retrieved  $T_s$  values have been evaluated by comparison with the surface air temperature  $T_{air}$  measured by the meteorological station network [Prigent *et al.*, 2003a]. The  $T_s - T_{air}$  difference shows all the expected variations with solar flux, soil characteristics, and cloudi-





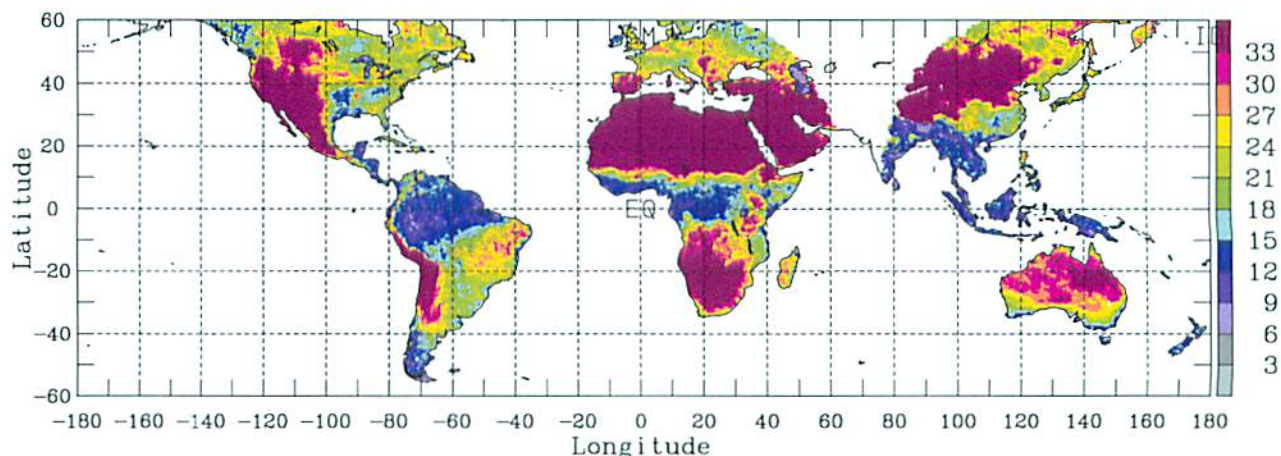
**Figure 3.** Normalized histogram of the difference between retrieved surface skin temperature from combined microwave and infrared observations and the surface air temperature at meteorological stations, for all coincident sites, accumulated over July and December 1992. The diurnal cycle has been removed from the two data sets separately.

ness. During daytime the  $T_s - T_{air}$  difference is driven by the solar insulation, with positive differences that increase with increasing solar flux. With decreasing soil and vegetation moisture, the evaporation rate decreases, increasing the sensible heat flux, thus requiring larger  $T_s - T_{air}$  differences. Nighttime  $T_s - T_{air}$  differences are governed by the longwave radiation balance, with  $T_s$  usually closer or lower than  $T_{air}$ . The presence of clouds dampens all the difference. After suppression of the variability associated to the diurnal solar flux variations, the  $T_s$  and  $T_{air}$  data sets show very good agreement in their synoptic variations, even for cloudy cases, with no bias and a global r.m.s. difference of  $\sim 2.9$  K (Figure 3). This value is an upper limit of the retrieval r.m.s. because it includes errors in the in situ data as well as errors

related to imperfect time and space collocations between the satellite and in situ measurements.

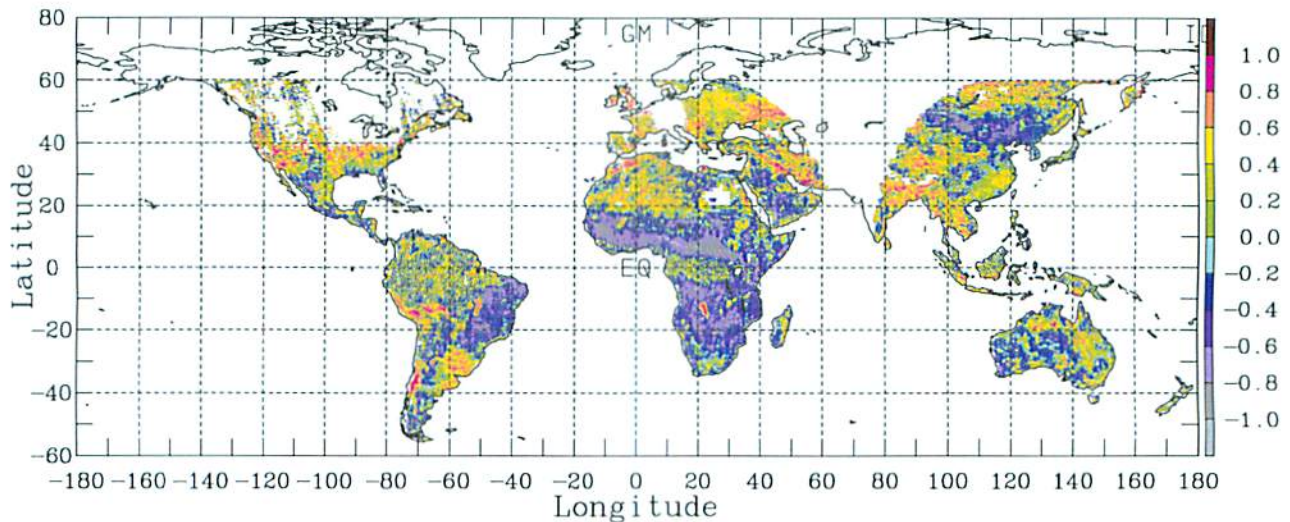
[78] The diurnal cycle of surface skin temperature has also been analyzed almost globally ( $60^\circ\text{N}$ – $60^\circ\text{S}$  over snow-free areas), using a Principal Component Analysis (PCA) [Aires *et al.*, 2004]. The first three components are identified as the amplitude, the phase, and the width (i.e., daytime duration) of the diurnal cycle and represent 97% of the variability. The PCA is used to regularize estimates of the diurnal cycle at a higher time resolution. A new temporal interpolation algorithm, designed to work when only a few measurements of surface temperature, has been developed on the basis of the PCA representation and an iterative optimization algorithm using good quality monthly climatological first guess information. The method is very flexible: only temperature measurements are used (no ancillary data), no surface model constraints are used (which is interesting for the comparison to model outputs, or for the assimilation of satellite estimates on surface models), and the time and number of measurements are not fixed. The performance of this interpolation algorithm has been tested for various diurnal sampling configurations. In particular, the potential to use the satellite microwave observations to provide a full diurnal surface temperature cycle in cloudy conditions has been investigated [Aires *et al.*, 2004]. Figure 4 represents the resulting averaged diurnal cycle amplitude of the surface temperature in June 1993. No geostationary satellites were available at that time over large parts of Eurasia explaining the impossibility to estimate accurate surface temperature diurnal cycles. The next step is to use AMSU (from NOAA platforms) or AMSR (from Aqua or Adeos II missions) radiometers in addition to the SSM/I instruments. The time sampling of these various instruments should provide a better characterization of the full diurnal cycle of  $T_s$ , even under cloudy skies.

[79] An all-weather time record of land surface skin temperatures has been produced from the merging of 10 years of microwave SSM/I satellite measurements and ISCCP products. The resulting global satellite observations of land surface skin temperature is now merged with surface weather observations of near-surface air temperature, humidity, and winds to study the diurnal, synoptic, and



**Figure 4.** Amplitude of the reconstructed surface skin temperature diurnal cycle for June 1993 (in K).





**Figure 5.** Linear correlation between the SSM/I emissivity polarization difference at 19 GHz (V-H) and the ECMWF surface soil moisture estimates, for the 1993–1994 period.

seasonal variations of land-atmosphere energy and water exchanges.

#### 4.2. Soil Moisture: An Example of Synergetic Use of Multisatellite Observations and LSM Outputs

[80] Soil moisture is a key land surface variable that partly controls the surface energy and water exchanges at the atmosphere interface. It is also very important for agriculture, water management, or flood monitoring. The SMOS dedicated mission that will provide soil moisture estimates from measurement in L-band (1.4 GHz) will not be launched before several years. Land surface modelers are now producing soil moisture estimates (e.g., GSWP-2) and there is an urgent need for consistent global data sets to evaluate model outputs [Entin *et al.*, 1999]. What can be done now with the available observations?

[81] A number of existing satellite observations have shown sensitivity to the soil moisture: it includes passive and active microwave measurements, as well as thermal infrared observations. Most studies are geographically limited but there are at least two global attempts. Owe *et al.* [2001] derive a global soil moisture index over 9 years from SMMR observations. The 6.63 GHz polarization difference makes it possible to take vegetation into account whereas the 37 GHz band gives access to the surface temperature. Wagner *et al.* [2003] analyze the temporal variation of the ERS scatterometer observations at 5.25 GHz to retrieve a global soil water index over 10 years from ERS.

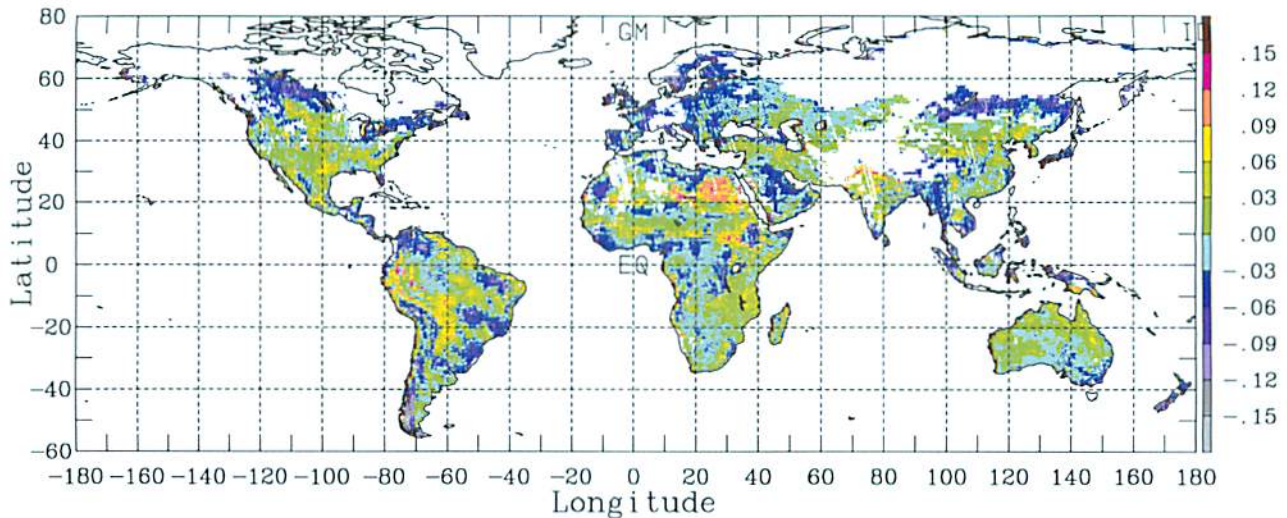
[82] We examined systematically and objectively the sensitivity of the available satellite observations on a global basis, in order to analyze their complementarity, and to assess the ability of combinations of these satellite measurements for soil moisture estimation [Prigent *et al.*, 2005a; Aires *et al.*, 2005]. For each type of observations, the optimum products are selected. The thermal infrared observations come from both the NOAA polar orbiters and the geostationary meteorological satellites, as processed by the ISCCP to obtain direct determination of the diurnal cycle of land surface skin temperature [Aires *et al.*, 2004]. Passive microwave information is provided by the SSM/I from

which the microwave land surface emissivities have been calculated; this analysis separates the atmosphere,  $T_s$ , and emissivity contributions to the observed signal, unlike previous studies [Prigent *et al.*, 1997, 2006]. The active microwave observations are extracted from the ERS scatterometer. In addition, the AVHRR NDVI is used to quantify and eventually separate the vegetation contribution from the other factors.

[83] First, for a two year period (1993–1994), the selected satellite observations are compared to the Global Soil Moisture Data Bank [Robock *et al.*, 2000] that provides in situ soil moisture measurements in five separate regions. This analysis makes it possible to objectively compare the sensitivity of each measurement type to the soil moisture, and to estimate the relative contribution of the vegetation for each one. Most studies are limited to one instrument, usually assuming it is the best for the given purpose, and it is very difficult to assess the relative sensitivity of the observations to the studied surface characteristic. The linear correlation coefficients between the in situ soil moisture measurements and the satellite variables are low when considered over all regions (−0.15 for the microwave emissivity polarization difference at 19 GHz from SSM/I and 0.41 for the active microwave ERS scatterometer measurements at small angles). Each satellite observation is differently sensitive to a large number of surface characteristics such as soil moisture, vegetation, soil texture, or roughness. Some of these parameters vary strongly from region to region but not strongly at each location and locally, the time variability of some of these parameters being limited, the correlation between the satellite information and the soil moisture is much stronger [Prigent *et al.*, 2005a].

[84] Second, the in situ soil moisture data set being unable to represent the full range of variability, outputs from Numerical Weather Prediction (NWP) reanalysis from ECMWF and NCEP along with the satellite variables have been analyzed over the globe, for two years. Figure 5 shows the linear correlation between the SSM/I emissivity polarization difference at 19 GHz (V-H) [Prigent *et al.*, 1997,





**Figure 6.** Volumetric soil moisture difference between the ECMWF estimate and the NN retrieved values for October 1993.

2006] and the ECMWF surface soil moisture estimates, for the 1993–1994 period. A positive correlation is expected between these two variables: with increasing soil moisture, the emissivity polarization difference should increase. In midlatitude regions, this is confirmed. However, strong negative correlations dominate in semiarid regions, the passive microwaves reacting primarily to the vegetation density. When the vegetation density and the soil moisture are negatively correlated, the passive microwaves vary as expected with soil moisture only because of the opposing effects of soil moisture and vegetation on the signal. This confirms what was observed with in situ measurements. On the basis of the statistical analysis of the comparison between the NWP soil moisture estimates and the satellite variables in addition to the understanding gained from the study with coincident in situ measurements, a method is derived to establish a statistical relationship between the soil moisture and satellite observations. A NN model is developed to describe the link between the satellite observations and the NWP soil moisture. No RTM today can accurately replicate this link on a global basis, for the wavelength range covered by the observations used. The NN model can reproduce the NWP soil moisture outputs (Figure 6) with a r.m.s. error of 5% volumetric soil moisture, close to what is expected from the future SMOS mission retrieval (4% volumetric soil moisture). More details are given by Aires *et al.* [2005].

[85] Although the NN model cannot be strictly considered as a retrieval algorithm because of its tight relationship with the NWP soil moisture, the fact that the independent satellite observations can be related to model outputs with this level of accuracy is a positive sign for relating these observations to the real-world soil moisture. The fact that the NN model is able to work on a global scale comes from the synergetic use of observations from various wavelengths.

[86] Comparisons between the NN model outputs and the NWP soil moisture reanalysis revealed some particular problems with the NCEP land surface models that have been confirmed by modelers. This statistical link can be

used to check the consistency between modeled soil moisture and satellite measurements and diagnose specific model problems (i.e., to invalidate them). We suggest applying the same analysis to the GSWP-2 results.

[87] The soil moisture estimates from our NN model can be assimilated in the LSM model. The necessary consistency between the satellite estimate and the LSM [Reichle *et al.*, 2004; McCabe *et al.*, 2005] is intrinsically verified.

## 5. Conclusion and Perspectives

[88] A great diversity of satellite measurements exists today and they are sensitive to various land surface parameters. Although these satellite observations might not be optimal for surface estimates, they have already shown some potential and optimized methodologies can be developed to fully exploit these satellite data for continental surface characterization. In this paper, innovative retrieval schemes are suggested that benefit from the synergy between the satellite observations, in addition to in situ measurements or land surface model outputs. Two concrete examples are presented for illustration: an “all weather” retrieval of surface skin temperature from combined microwave and infrared observations and a soil moisture analysis from the merging of multisatellite observations and LSM outputs.

[89] This study focuses on the application of the land surface remote sensing for land surface modeling. However, similar methodologies can also benefit other applications such as hydrology or agriculture as the primary objective of these methods is to derive better land surface products.

[90] How to proceed in a practical way to extend the use of the new methodologies and provide the community with improved satellite land surface products?

[91] First, the multi-instrument approaches have to be adopted, although it is more complex to put in place. Different methods can be applied and the selection of a specific one depends upon the application and upon the variable of interest. The retrieval methodology has to be clearly documented for the users and a detailed analysis of

the uncertainties associated with the satellite estimates has to be performed. The underlying assumptions and the use of any ancillary information have to be described. Modelers are often reluctant to use remote sensing data: the retrieval algorithms are unclear and there are doubts on their accuracy. As a consequence, they often prefer the in situ measurements that they know well and that represent the "truth." With better estimates of the errors, the remote sensing data could be used more quantitatively in the models [McCabe *et al.*, 2005].

[92] Second, there is a strong need for consistent and accurate calibration of satellite data, across instruments and across platforms, over the life spans of the missions. Within the ISCCP framework, IR and visible radiances are carefully intercalibrated, over more than 20 years and for a large number of platforms. Similar work has to be performed for other sensors. Some actions have recently been undertaken but this tedious long-term effort has to be strongly supported.

[93] Third, the in situ observations required to evaluate the satellite retrieval and the models have to be easily accessible, over long periods of time and covering a large range of environments. Variables that are not part of the routinely measured meteorological observations are generally difficult to access, although frequent measurements might be performed all over the world. Initiatives such as the Global Soil Wetness Database must be encouraged, to foster the development of databases of unified, well formatted, and quality controlled long-term in situ measurements. In addition, there is a need for the development of measurement sites for some key variables such as turbulent fluxes for which only a limited number of observations exists.

[94] These efforts could be conducted within the framework of a future ISLSCP initiative, that would aim at providing a synthetic data record in which inconsistencies among LSMs, satellite data, and in situ measurements are reconciled to the degree possible. In order to reach these ambitious goals, an improved communication between the different actors of the land science community is necessary. The link between modelers and the field experimenters has been naturally and historically tight. There is now a necessity for the satellite community to work in close collaboration with the land surface modelers to understand their needs and define common strategies. This collaboration will benefit both communities.

[95] The next challenge of the satellite community is the estimation of the turbulent fluxes over land, at a global scale, which are key variables of the land surface models. It will inevitably require a synergetic use of multiple sources of information, including a large range of satellite observations and model outputs. This can be a very good opportunity for the modelers and the satellite experts to join forces toward a common goal.

[96] **Acknowledgment.** We would like to thank Bill Rossow (NASA/GISS), Hervé Douville (Météo-France), Alan Robock (Rutgers University), and John Bates (NOAA) for their comments and suggestions on the manuscript.

## References

Adler, R. F., et al. (2003), The Version 2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979-present), *J. Hydrometeorol.*, 4, 1147–1167.

- Aires, F., C. Prigent, W. B. Rossow, and M. Rothstein (2001), A new neural network approach including first-guess for retrieval of atmospheric water vapor, cloud liquid water path, surface temperature and emissivities over land from satellite microwave observations, *J. Geophys. Res.*, 106, 14,887–14,907.
- Aires, F., C. Prigent, and W. B. Rossow (2004), Temporal interpolation of global surface skin temperature diurnal cycle over land under clear and cloudy conditions, *J. Geophys. Res.*, 109, D04313, doi:10.1029/2003JD003527.
- Aires, F., C. Prigent, and W. B. Rossow (2005), Sensitivity of satellite microwave and infrared observations to soil moisture at a global scale: 2. Global statistical relationships, *J. Geophys. Res.*, 110, D11103, doi:10.1029/2004JD005094.
- Berg, A. A., J. S. Famiglietti, J. P. Walker, and P. R. Houser (2003), Impact of bias correction to reanalysis products on simulations of North American soil moisture and hydrological fluxes, *J. Geophys. Res.*, 108(D16), 4490, doi:10.1029/2002JD003334.
- Beven, K. J., and J. Freer (2001), Equifinality, data assimilation, and uncertainty assessment in mechanistic modeling of complex environmental systems using the GLUE methodology, *J. Hydrol.*, 249, 11–29.
- Biau, G., E. Zorita, H. von Storch, and H. Wackemagel (1999), Estimation of precipitation by kriging in EOF space, *J. Clim.*, 12, 1070–1085.
- Boone, A., N. Mognard, B. Decharme, H. Douville, M. Grippa, and K. Kerrigan (2006), The impact of simulated soil temperatures on the estimation of snow depth over Siberia from SSM/I, *Remote Sens. Environ.*, 101, 482–494.
- Brest, C. L., W. B. Rossow, and M. Roiter (1997), Update of radiance calibrations for ISCCP, *J. Atmos. Oceanic Technol.*, 14, 1091–1109.
- Cavazos, T. (1999), Large-scale circulation anomalies conducive to extreme precipitation events and derivation of daily rainfall in northeastern Mexico and southeastern Texas, *J. Clim.*, 12, 1506–1523.
- Chédin, A., N. A. Scott, C. Wahiche, and P. Moulinier (1985), The improved initialization inversion method: A high resolution physical method for temperature retrievals from satellites of the TIROS-N series, *J. Appl. Meteorol.*, 24, 128–143.
- Colton, M. C., and G. A. Poe (1999), Intersensor calibration of DMSR SSM/I's: F-8 to F-14, 1987–1997, *IEEE Trans. Geosci. Remote Sens.*, 37, 418–439.
- Cordisco, E., C. Prigent, and F. Aires (2006), Snow characterization at a global scale with passive microwave satellite observations, *J. Geophys. Res.*, 111, D19102, doi:10.1029/2005JD006773.
- Entekhabi, D., et al. (1999), An agenda for land surface hydrology research and a call for the Second International Hydrological Decade, *Bull. Am. Meteorol. Soc.*, 80, 2043–2058.
- Entin, J. K., A. Robock, K. Y. Vinnikov, V. Zabelin, S. Liu, A. Namkhani, and T. Adyasuren (1999), Evaluation of global soil wetness project soil moisture simulations, *J. Meteorol. Soc. Jpn.*, 77, 183–198.
- Franks, S. W., K. J. Beven, and J. H. C. Gash (1999), Multi-objective conditioning of a simple SVAT model, *Hydrol. Earth Syst. Sci.*, 3, 477–489.
- Gupta, H. V., L. A. Bastidas, S. Sorooshian, W. J. Shuttleworth, and Z. L. Yang (1999), Parameter estimation of a land surface scheme using multi-criteria methods, *J. Geophys. Res.*, 104, 19,491–19,503.
- Gutman, G. (1999), On the use of long-term global data of land reflectances and vegetation indices from the advanced very high resolution radiometer, *J. Geophys. Res.*, 104, 6241–6255.
- Hall, F. G., J. R. Townshend, and E. T. Engman (1995), Status of remote sensing algorithms for estimation of land surface parameters, *Remote Sens. Environ.*, 51, 138–156.
- Henderson-Sellers, A., A. J. Pitman, P. K. Love, P. Irannejad, and T. H. Chen (1995), The Project for Intercomparison of Land Surface Parameterization Schemes (PILPS): Phases 2 and 3, *Bull. Am. Meteorol. Soc.*, 76, 489–503.
- Holben, B. N. (1986), Characteristics of maximum-value composite images from temporal AVHRR data, *Int. J. Remote Sens.*, 7, 1417–1434.
- Ide, K., P. Courtier, M. Ghil, and A. C. Lorenc (1997), Unified notation for data assimilation: Operational, sequential and variational, *J. Appl. Meteorol.*, 36, 181–189.
- Kelly, R. E., A. T. C. Chang, L. Tsang, and J. Foster (2003), A prototype AMSR-E global snow area and snow depth algorithm, *IEEE Trans. Geosci. Remote Sens.*, 41, 230–242.
- Kerr, Y. H., and E. G. Njoku (1993), On the use of passive microwave at 37GHz in remote sensing of vegetation, *Int. J. Remote Sens.*, 14, 1931–1943.
- Kerr, Y. H., P. Waldteufel, J.-P. Wigneron, J.-M. Martinuzzi, J. Font, and M. Berger (2001), Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission, *IEEE Trans. Geosci. Remote Sens.*, 39, 1729–1735.
- Kummerow, C., et al. (2001), The evolution of the Goddard Profiling Algorithm (GPROF) for rainfall estimation from passive microwave sensors, *J. Appl. Meteorol.*, 40, 1801–1820.



- Lakshmi, V., E. F. Wood, and B. J. Choudhury (1997), A soil-canopy-atmosphere model for use in satellite microwave remote sensing, *J. Geophys. Res.*, **102**, 6911–6927.
- Leese, J., S. Williams, R. Jenne, and A. Ritchie (2003), Data collection and management for Global Energy and Water Cycle Experiment (GEWEX) Continental-Scale International Project (GCIP), *J. Geophys. Res.*, **108**(D16), 8620, doi:10.1029/2002JD003196.
- McCabe, M. F., S. W. Franks, and J. D. Kalma (2005), Calibration of a land surface model using multiple data sets, *J. Hydrol.*, **302**, 209–222.
- Milly, P. C. D., and A. B. Shmakin (2002), Global modeling of land water and energy balances: I. The Land Dynamics (LaD) model, *J. Hydrometeorol.*, **3**, 283–299.
- Mitchell, K. E., et al. (2004), The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system, *J. Geophys. Res.*, **109**, D07S90, doi:10.1029/2003JD003823.
- Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth (2004), Quantification of modeling uncertainties in a large ensemble of climate change simulations, *Nature*, **430**, 768–772.
- Njoku, E. G., T. J. Jackson, V. Lakshmi, T. K. Chan, and S. V. Nghiem (2003), Soil moisture retrieval from AMSR-E, *IEEE Trans. Geosci. Remote Sens.*, **41**, 215–229.
- Owe, M., R. de Jeu, and J. Walker (2001), A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization index, *IEEE Trans. Geosci. Remote Sens.*, **39**, 1643–1654.
- Prigent, C., W. B. Rossow, and E. Matthews (1997), Microwave land surface emissivities estimated from SSM/I observations, *J. Geophys. Res.*, **102**, 21,867–21,890.
- Prigent, C., F. Aires, W. B. Rossow, and E. Matthews (2001), Joint characterization of vegetation by satellite observations from visible to microwave wavelength: A sensitivity analysis, *J. Geophys. Res.*, **106**, 20,665–20,685.
- Prigent, C., F. Aires, and W. B. Rossow (2003a), Land surface skin temperatures from a combined analysis of microwave and infrared satellite observations for an all-weather evaluation of the differences between air and skin temperatures, *J. Geophys. Res.*, **108**(D10), 4310, doi:10.1029/2002JD002301.
- Prigent, C., F. Aires, and W. B. Rossow (2003b), Retrieval of surface and atmospheric geophysical variables over snow from microwave satellite observations, *J. Appl. Meteorol.*, **42**, 368–380.
- Prigent, C., F. Aires, and W. B. Rossow (2005a), Sensitivity of satellite microwave and infrared observations to soil moisture at a global scale: Relationship of satellite observations to in situ soil moisture measurements, *J. Geophys. Res.*, **110**, D07110, doi:10.1029/2004JD005087.
- Prigent, C., I. Tegen, F. Aires, B. Marticorena, and M. Zribi (2005b), Estimation of the aerodynamic roughness length in arid and semi-arid regions over the globe with the ERS scatterometer, *J. Geophys. Res.*, **110**, D09205, doi:10.1029/2004JD005370.
- Prigent, C., F. Aires, and W. B. Rossow (2006), Land surface microwave emissivities over the globe for a decade, *Bull. Am. Meteorol. Soc.*, in press.
- Rabier, F., A. McNally, E. Andersson, P. Courtier, P. Undén, J. Eyre, A. Hollingsworth, and F. Boutier (1998), The ECMWF implementation of three-dimensional variational assimilation (3D-Var). II Structure functions, *Q. J. R. Meteorol. Soc.*, **124**, 1809–1829.
- Ramillien, G., F. Frappart, A. Cazenave, and A. Guntner (2005), Time variations of land water storage from an inversion of 2 years of GRACE geoids, *Earth Planet. Sci. Lett.*, **235**, 283–301.
- Reichle, R. H., D. McLaughlin, and D. Entekhabi (2002a), Hydrology data assimilation with the ensemble Kalman filter, *Mon. Weather Rev.*, **130**, 103–114.
- Reichle, R. H., J. P. Walker, R. D. Koster, and P. R. Houser (2002b), Extended versus ensemble Kalman filtering for land data assimilation, *J. Hydrometeorol.*, **3**, 728–740.
- Reichle, R. H., R. D. Koster, J. Dong, and A. A. Berg (2004), Global soil moisture from satellite observations, land surface models, and ground data: Implication for data assimilation, *J. Hydrometeorol.*, **5**, 430–442.
- Robock, A., C. A. Schlosser, K. Y. Vinnikov, N. A. Speranskaya, J. Entin, and S. Qiu (1998), Evaluation of AMIP soil moisture simulations, *Global Planet. Change*, **19**, 181–208.
- Robock, A., K. Y. Vinnikov, G. Srinivasan, J. K. Entin, S. E. Hollinger, N. A. Speranskaya, S. Liu, and A. Namkhai (2000), The Global Soil Moisture Data Bank, *Bull. Am. Meteorol. Soc.*, **81**, 1281–1299.
- Rodgers, C. D. (1976), Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation, *Rev. Geophys.*, **14**, 609–624.
- Rodgers, C. D. (2000), *Inverse Methods for Atmospheric Sounding: Theory and Practice*, World Sci., Hackensack, N. J.
- Rossow, W. B., and R. A. Schiffer (1999), Advances in understanding clouds from ISCCP, *Bull. Am. Meteorol. Soc.*, **80**, 2261–2287.
- Schaafe, J. C., et al. (2004), An intercomparison of soil moisture fields in the North American Land Data Assimilation System (NLDAS), *J. Geophys. Res.*, **109**, D01S90, doi:10.1029/2002JD003309.
- Sellers, P. J., et al. (1995), Remote sensing of the land surface for studies of global change: Models, algorithms, experiments, *Remote Sens. Environ.*, **51**, 3–26.
- Sippel, S. J., S. K. Hamilton, J. M. Melack, and E. M. Novo (1998), Passive microwave observations of inundation area and area/stage relation in the Amazon river floodplain, *Int. J. Remote Sens.*, **19**, 3055–3074.
- Tapley, B. D., S. Bettadpur, J. C. Ries, P. F. Thompson, and W. Watkins (2004), GRACE measurements of mass variability in the Earth system, *Science*, **305**, 503–505.
- Tarantola, A. (1987), *Inverse Problem Theory: Models for Data Fitting and Model Parameter Estimation*, Elsevier, New York.
- Tucker, C. J. (1979), Red and photographic infrared linear combinations for monitoring vegetation, *Remote Sens. Environ.*, **8**, 127–150.
- Tucker, C. J., J. R. G. Townshend, and T. E. Goff (1985), African land-cover classification using satellite data, *Science*, **227**, 369–375.
- Twomey, S. (1977), *Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements*, Elsevier, New York.
- Vinnikov, K. Y., A. Robock, S. Qiu, J. K. Entin, M. Owe, B. J. Choudhury, S. E. Hollinger, and E. G. Njoku (1999), Satellite remote sensing of soil moisture in Illinois, United States, *J. Geophys. Res.*, **104**, 4145–4168.
- Von Storch, H., E. Zorita, and U. Cubasch (1993), Downscaling of global climate change estimates to regional scales: An application to Iberian rainfall in wintertime, *J. Clim.*, **6**, 1161–1171.
- Wackernagel, H. (2003), *Multivariate Geostatistics: An Introduction With Applications*, 3rd ed., Springer, New York.
- Wagner, W., K. Seipal, C. Pathe, D. Gerten, W. Lucht, and B. Rudolf (2003), Evaluation of the agreement between the first global remotely sensed soil moisture data with model and precipitation data, *J. Geophys. Res.*, **108**(D19), 4611, doi:10.1029/2003JD003663.
- Walker, J. P., and P. R. Houser (2001), A methodology for initializing soil moisture in a global climate model: Assimilation of near surface soil moisture observations, *J. Geophys. Res.*, **106**, 11,761–11,774.
- Walker, J. P., and P. R. Houser (2004), Requirements of a global near-surface soil moisture satellite mission: Accuracy, repeat time, and spatial resolution, *Adv. Water Resour.*, **27**, 785–801.
- Weng, F., B. Yan, H. Xu, and N. Sun (2005), Assessment of DMSP series of SSM/I data for NOAA weather and climate applications paper presented at SSMIS Working Group Meeting, Monterey, Calif., 26–28 Oct.

F. Aires, Laboratoire de Météorologie Dynamique, Institut Pierre-Simon Laplace/Centre National de la Recherche Scientifique, Université Pierre et Marie Curie, case 99 4, place Jussieu, F-75252 Paris Cédex 05, France. (filipe.aires@lmd.jussieu.fr)

C. Prigent, Laboratoire d'Études du Rayonnement et de la Matière en Astrophysique, Centre National de la Recherche Scientifique, Observatoire de Paris, 61, av. de l'Observatoire, F-75014 Paris, France. (catherine.prigent@obspm.fr)



## Exploring global patterns of net primary production carbon supply and demand using satellite observations and statistical data

Marc L. Imhoff<sup>1</sup> and Lahouari Bounoua<sup>1</sup>

Received 7 April 2006; revised 14 August 2006; accepted 9 October 2006; published 22 November 2006.

[1] A unique combination of satellite and socioeconomic data were used to explore the relationship between human consumption and the carbon cycle. The amount of Earth's net primary production (NPP) required to support human activities is a powerful measure of the aggregate impact on the biosphere and indicator of societal vulnerability to climate change. Biophysical models were applied to consumption data to estimate the annual amount of Earth's terrestrial net primary production humans require for food, fiber (including fabrication) and fuel using the same modeling architecture as satellite-supported NPP measurements. The amount of NPP required was calculated on a per capita basis and projected onto a global map of population to create a spatially explicit map of NPP-carbon "demand" in units of elemental carbon. NPP demand was compared to a map of Earth's average annual net primary production or "supply" created using 17 years (1982–1998) of AVHRR vegetation index to produce a geographically accurate balance sheet of NPP-carbon "supply" and "demand" for the globe. Globally, humans consume 20% of Earth's total net primary production on land. Regionally, the NPP-carbon balance percentage varies from 6% to over 70% and locally from near 0% to over 30,000% in major urban areas. Scenarios modeling the impact of per capita consumption, population growth, and technology suggest that NPP demand is likely to increase substantially in the next 40 years despite better harvesting and processing efficiencies.

**Citation:** Imhoff, M. L., and L. Bounoua (2006), Exploring global patterns of net primary production carbon supply and demand using satellite observations and statistical data, *J. Geophys. Res.*, *111*, D22S12, doi:10.1029/2006JD007377.

### 1. Introduction

[2] An important but relatively little studied part of the global carbon cycle is the fraction of the planet's net primary production (NPP) appropriated by human beings [Vitousek *et al.*, 1997; Postel *et al.*, 1996]. Human consumption of NPP in the form of food, fiber (including fabrication), and wood-based fuel products has significant implications both in terms of its proportion relative to total planetary NPP (up to 55% by some estimates) and its impact on a wide range of ecological and biophysical processes [Wackernagel *et al.*, 2002; Vitousek *et al.*, 1986; Rojstaczer *et al.*, 2001]. Human cooption of the products of photosynthesis alters the composition of the atmosphere [Schimel *et al.*, 2000], modulates the flow of important ecosystem services [Daily *et al.*, 1997], affects levels of biodiversity [Pimm and Gittleman, 1992; Sala *et al.*, 2000; Haberl, 1997] and diverts energy flows within food webs [Field, 2001; Cardoch *et al.*, 2002].

[3] The portion of Earth's NPP supporting human activity occupies a pivotal position in the carbon cycle through its dependence and feedback on socioeconomic conditions,

ecosystem function, and climate. How it functions has immediate as well as long-term implications to human welfare and has been identified as an important focus area for scientific research and policy formulation [Rosegrant and Cline, 2003; Hasselmann *et al.*, 2003; Smith, 2003]. Of particular importance is how increasing human demands on Earth's ecosystems for producing food and fiber will affect the functioning of the biosphere and a sustainable future for the human enterprise within the context of global change.

### 2. NPP, the Biological Engine, and the Human Requirement

[4] From a biological perspective, NPP represents the primary energy source for Earth's ecosystems and complex food webs by supplying food energy to the planet's heterotrophic organisms (organisms that require preformed organic compounds for food, including human beings). Humans appear to exert a remarkable demand on this part of the carbon cycle for a species that represents roughly 0.5% of Earth's total heterotroph biomass [Smil, 1983]. An influential study by Vitousek *et al.* [1986], for example, estimated that humans appropriate 31% of global NPP (intermediate calculation) with "high" (39%) and "low" (3%) estimates, based on more or less inclusive definitions of human appropriation. Rojstaczer *et al.* [2001] in an approach similar to Vitousek *et al.* [1986] used improved data and robust statistical methods to estimate that humans use

<sup>1</sup>Biospheric Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.

roughly 32% of global NPP, but they reported high uncertainty in this result (10% to 55% appropriation).

[5] Because these previous studies based their calculations on a mix of aggregated biome-wide averages and consumption information, they were unable to fully account for spatially heterogeneous processes (e.g., human caloric intake, agricultural productivity, NPP spatial distribution). As a result the spatial patterns of human NPP appropriation remained hidden and the methodologies did not lend themselves well to spatial comparisons with spatially explicit satellite-derived indices of biological productivity and global change [Field, 2001; Haberl *et al.*, 2002].

[6] In a previous work, we described an approach for estimating the fraction of NPP required to support human activities using biogeochemical relationships that match those used in satellite-based methods [Imhoff *et al.*, 2004]. This approach allowed a comparison of the rate of NPP required to support human consumption (NPP demand) with the rate of terrestrial production (NPP supply). Described here are new results from this approach showing the spatial characteristics of an NPP supply and demand relationship driven by population distribution and per capita consumption. Also included is an exploration of how changing population and socioeconomic conditions are reflected as potential forcings in NPP carbon demand under different consumption scenarios. In order to avoid confusion with the various published definitions of HANPP [Haberl *et al.*, 2002], quantities reported here for Human Appropriated NPP or HANPP represent the amount of total NPP required (as elemental carbon) to produce consumed products including; food, fiber, wood, and wood-based fuels (same as Imhoff *et al.* [2004]). NPP required, NPP demand, and HANPP are synonymous terms in this paper. Although humans also consume the products of primary production from aquatic and marine systems, this analysis is limited to terrestrial sources.

### 3. Methods

#### 3.1. Estimate of NPP Supply Using ISLSCP Data

[7] Terrestrial NPP supply (here after designated as NPP) in the form of elemental carbon was estimated by applying the Carnegie-Ames-Stanford Approach (CASA) terrestrial carbon model [Potter *et al.*, 1993] to global fields of normalized difference vegetation index (NDVI) from the Advanced Very High Resolution Radiometer (AVHRR) and surface climatology data from ISLSCP II (International Satellite Land Surface Climatology Project *initiative II*) [Hall *et al.*, 2005] and the Global Inventory Monitoring and Modeling System (GIMMS). The data were composed of the maximum observed monthly NDVI spanning a 17 year time period from 1982 to 1998. The data processing for this time series included improved navigation, calibration of the four different sensors, corrections for sensor degradation, and atmospheric correction including Rayleigh absorption and scattering, and El Chichon and Pinatubo aerosols [Los *et al.*, 2000; Tucker *et al.*, 2001]. The correction of the satellite artifacts in this data set and the comparatively long period of coverage make it attractive for investigations of long-term trends in biological productivity [Hicke *et al.*, 2002].

[8] From the 17-year data series, we compiled a single set of monthly NDVI averages representing a composite annual

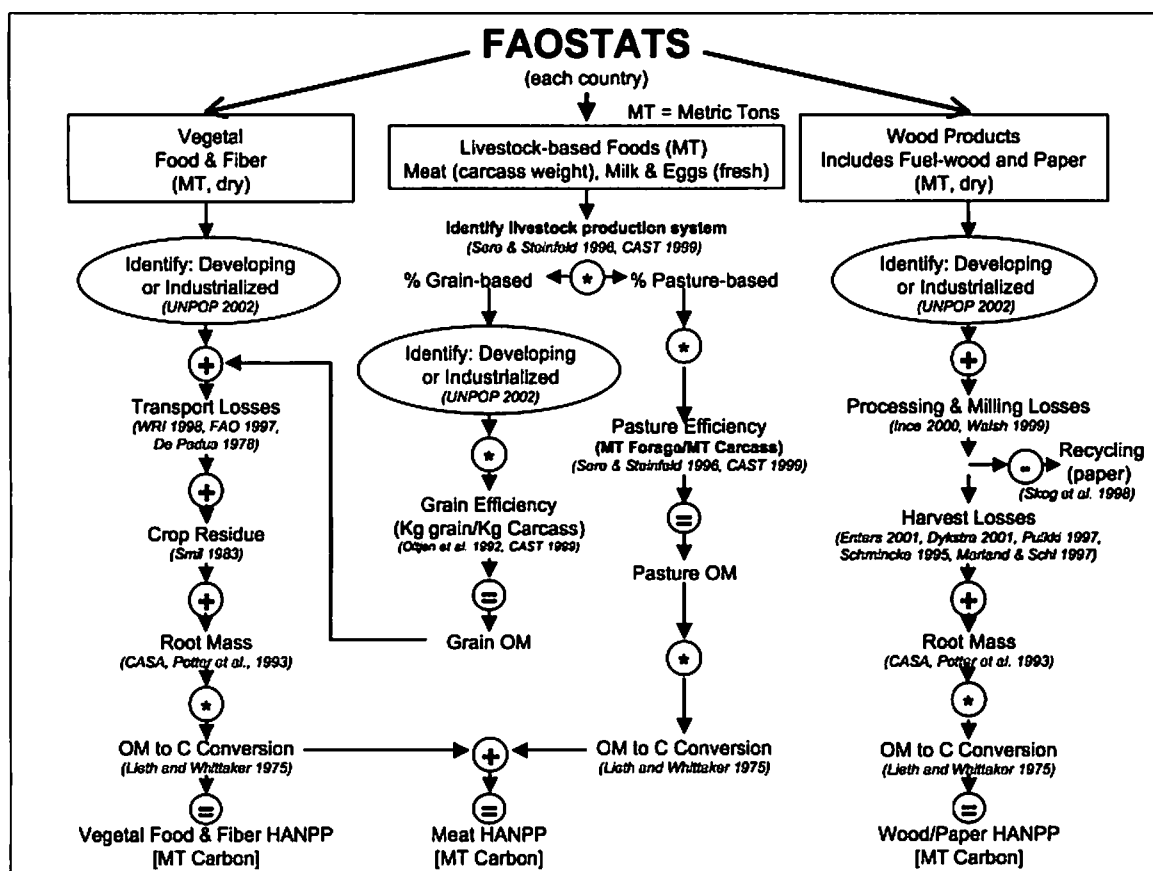
cycle (the composite NDVI for January, for example, is the average of the observed monthly NDVI for all of the Januaries from 1982 through 1998).

[9] NPP was estimated by applying the CASA terrestrial carbon model to the satellite data and surface climatology. The CASA model characterized the fixation and release of carbon on the basis of a spatially and temporally resolved prediction of NPP in a steady state [Potter *et al.*, 1993]. NPP was estimated on a monthly timescale as the amount of intercepted photosynthetically active radiation (IPAR) modulated by a light use efficiency factor. IPAR was determined by the product of the total incident solar radiation and the fraction of the incoming PAR intercepted by the green fraction of the vegetation (FPAR) derived from the AVHRR data [Sellers, 1985; Sellers *et al.*, 1996a, 1996b]. The light efficiency factor was controlled by environmental stresses for temperature and water [Monteith, 1977; Kumar and Monteith, 1981]. The allocation of carbon to woods, leaves, and roots as well as the turnover times was determined by vegetation type from the vegetation classification map defining 12 classes of vegetation cover [Hansen *et al.*, 2000]. In addition to vegetation classification and its associated monthly biophysical fields derived from NDVI data, CASA also required monthly fields of temperature and precipitation [Shea, 1986], solar radiation [Bishop and Rossow, 1991] and soil texture [Zobler, 1986]. The climate drivers, temperature, precipitation and solar radiation were resampled from global  $1^\circ \times 1^\circ$  resolution to  $0.25^\circ \times 0.25^\circ$  resolution using a bilinear interpolation algorithm and averages generated from the historical data matching the satellite data. In a model intercomparison study including seventeen global models of terrestrial biogeochemistry, the annual NPP from CASA was close to the annual average value from the seventeen participating models including some that did not use satellite data [Cramer *et al.*, 1999]. The NPP calculation also compares well to other more recent satellite-supported estimates using the same AVHRR series [Nemani *et al.*, 2003] and MODIS [Zhao *et al.*, 2005]. In this analysis, only the vegetation existing on land was considered. Aquatic or marine systems were not included.

#### 3.2. Estimating NPP Carbon Demand

[10] NPP carbon demand is defined as the annual amount of terrestrial NPP required to derive the food and fiber products (construction material and fuel wood) consumed by humans as reported in the United Nations FAOSTATS database and the biomass lost in harvest and processing (e.g., crop residues).

[11] Starting with input data on food and fiber products consumed on an annual basis, models were developed estimating the amount of NPP required in the field (i.e., at the landscape level) to generate the various end products so that it can be compared on the same biophysical basis to satellite-supported estimates of NPP supply. The NPP-based products derived on land for 230 countries were compiled into seven categories: vegetal foods, meat, milk, eggs, wood (building and fuel), paper, and fiber. Harvest, processing, and efficiency multipliers and estimates of below-ground production were applied to successively add mass thereby reconstructing the total amount of NPP required, at-the-source, to derive the final products. Separate efficiency multipliers for industrialized and developing countries were



**Figure 1.** Logic chain (simplified) for estimating the amount of NPP required for food and fiber products. To constrain the calculation within country boundaries the FAOSTAT sums input to this process (top of graph) are the domestic supply (i.e., production + imports – exports).

derived from the literature based the UNFAO designated development status of the source countries (Figure 1).

[12] The country level FAO data for 1995 were scanned to correct for missing data and reporting errors. Over reporting due to multiple entries for the same country was eliminated and national entities or territories reporting under another administrative country were identified (e.g., Guam reports under the United States). For calculations of per capita consumption at the national level, the population of these entities was added to that of the administrative country and a national per capita consumption was obtained by dividing the administrative country's consumption by the total population.

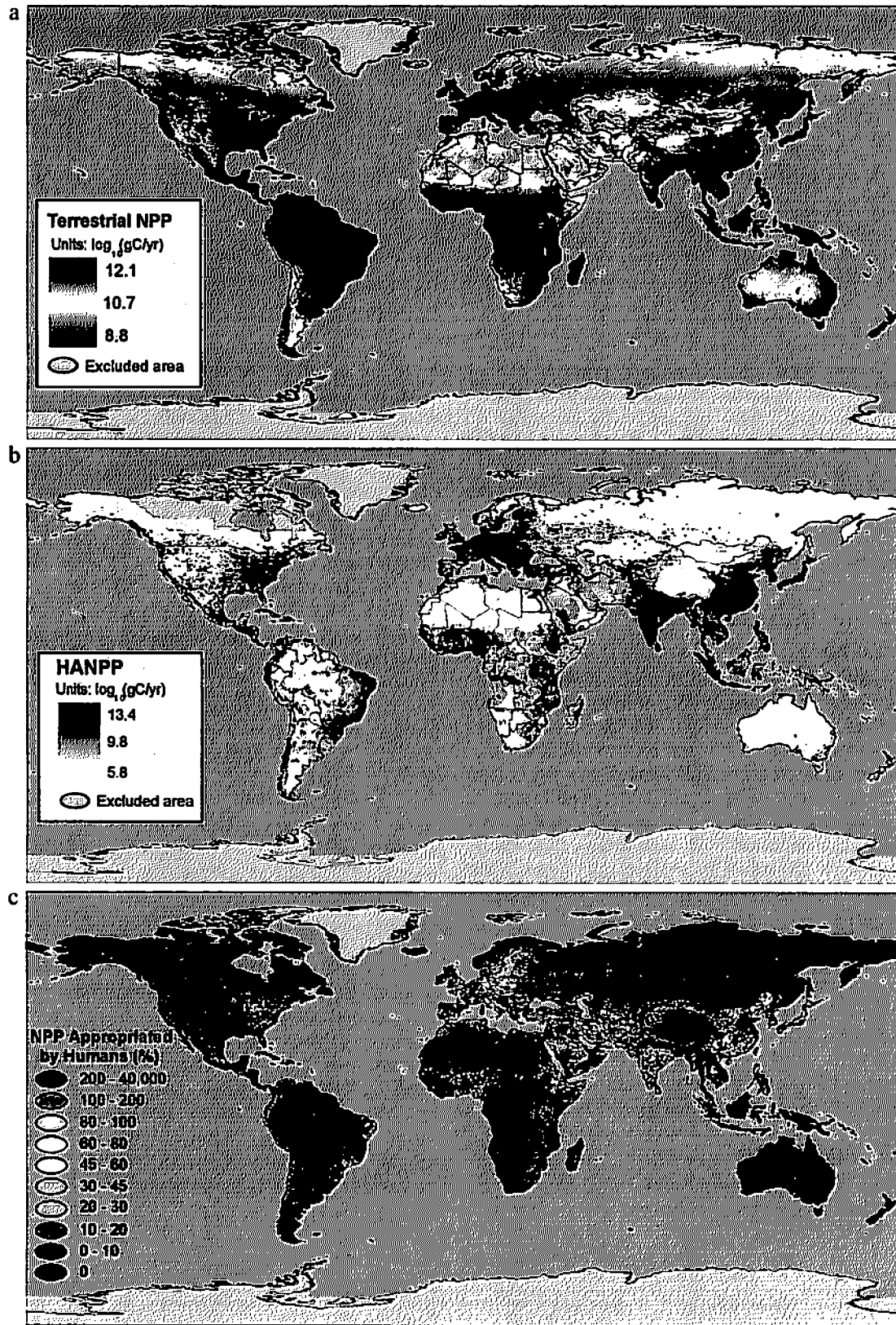
[13] Product sums for all plant products were rendered in terms of dry mass. For vegetal foods and fiber, mass was successively added to the reported value to account for post harvest processing, transport losses [World Resources Institutes, 1998; United Nations Food and Agriculture Organization, 1997; De Padua, 1978] and crop residue left in the fields [Smil, 1983]. For the intermediate case the weighted mean for major world crops was used while high and low estimates were  $\pm 1$  standard deviation [see Imhoff et al., 2004]. The large variation in residue among crop types yielded a high standard deviation.

[14] For wood and paper products, organic matter (OM) was added to account for milling [Ince, 2000; Walsh et al., 1999] and harvest losses [Enters, 2001; Dykstra, 2001; Pulkki,

1997; Schmincke, 1995; Marland and Schlamadinger, 1997]. For paper, recycling was accounted for by subtracting the quantity recycled annually reported by Skog et al. [1998].

[15] In cases where the individual plant is killed in the process (all cases except pasture grasses), the mass of the root system was also included. Root organic matter was estimated using the same values employed in the NPP supply calculations (see below) with multipliers for short vegetation (2.0) applied to vegetal foods, fiber, and grain used in meat production, and woody vegetation (1.5) for wood and paper [Potter et al., 1993]. No root OM was added to pasture grasses used for livestock.

[16] Meat production was reported as carcass weight (wet) including all meat types. The NPP required for meat was estimated by summing the NPP required for grain and pasture-based (forage) feed using a global average of 62% grain and 38% forage [Sere and Steinfeld, 1996]. The amount of feed as organic matter (OM) was estimated using feed use efficiency values (kg feed OM/kg carcass) for grain (2.3:1 average for all meat types) and pasture (21.46:1, ruminant [Council for Agricultural Science and Technology, 1999; Oltjen et al., 1992]). The NPP required for total feed-grain was calculated in the same way as for vegetal foods, adding residue and loss factors appropriate to each country's development status and final conversion from organic matter to carbon. Since pasture grazing is in situ, no loss



**Figure 2.** (a) Annual terrestrial NPP “supply” (56.8 PgC) estimated using a 17 year average of maximum monthly NDVI from AVHRR, the CASA model, and climate drivers. (b) Map showing the amount of NPP required (in  $\log_{10}$  grams of carbon) to support the population in each grid cell on an annual basis. Statistics can be aggregated globally, regionally, subregionally, or by country. (c) Map comparing NPP supply versus demand at 0.25° resolution (NPP required as a % of NPP supply). See color version of this figure in the HTML.



**Table 1.** Annual Estimates of NPP Required (PgC; 1Pg = 10<sup>15</sup>g) for a Population of 5.69 Billion in 1995

Consumed Products	Low Estimate	Intermediate Estimate	High Estimate
Vegetal food	0.89	1.73	2.96
Meat	1.69	1.92	2.21
Milk	0.15	0.27	0.43
Eggs	0.09	0.17	0.27
Human food (subtotal)	2.83	4.09	5.85
Paper	0.20	0.28	0.38
Fiber	0.32	0.37	0.42
Wood products (construction and fuel)	4.64	6.81	8.15
Human commodities (subtotal)	5.17	7.45	8.95
Total NPP required	8.00	11.54	14.80
Total as % of NPP supply (56.8 Pg)	14.10	20.32	26.07

or residue factors were applied to pasturage. Efficiency factors for milk and eggs are for grain component only. Final representation of NPP required was converted to elemental carbon using Carbon/OM ratios following Lieth [1975].

[17] For spatial representation, we calculated RNPP at the country level using domestic supply (i.e., production + imports – exports) to constrain the country totals to products consumed in situ.

## 4. Results and Discussion

### 4.1. NPP Supply

[18] We produced a global map of average annual terrestrial NPP for the 17 year time interval (Figure 2a). Summed across the land surface, we estimated the average total global NPP supply to be 56.8 Pg of elemental carbon (Pg, 10<sup>15</sup> grams), a value within the range of other estimates using various models [Cramer *et al.*, 1999]. The averaging approach makes a good baseline estimate of NPP supply because using data collected over such a long time period reduces short-term variations in surface conditions while still incorporating decadal-scale effects of human influence on the land surface.

### 4.2. NPP Demand

[19] Summing the amount of NPP required for all products yielded an intermediate global NPP demand estimate of 11.5 Pg of elemental carbon, equivalent to 24.3 Pg of dry organic matter. To address uncertainty, we bracketed our intermediate calculation with low and high estimates using the range of reported efficiency, loss, and residue multipliers. Differences in these multipliers correspond to a country's technical proficiency linked to its development status (e.g., timber harvest and milling losses are typically

lower in industrialized nations). For the intermediate estimate, we applied harvest and processing efficiencies to each country on the basis of its UNFAO-designated development status [United Nations Food and Agriculture Organization, 2001]. For high and low estimates, we applied to all countries the multipliers that produce maximum and minimum estimates, respectively. Our low and high calculations yield NPP demand estimates of 8.0 and 14.8 Pg C, respectively (Table 1).

[20] To produce a global map of NPP demand (Figure 2b), we applied the NPP required per capita (calculated at the country level) to a global database of human population produced by the Center for International Earth Science Information Network [Center for International Earth Science Information Network, 2000]. These data were gridded at 0.25° which equates to a spatial resolution of about 28 × 28 km at the equator.

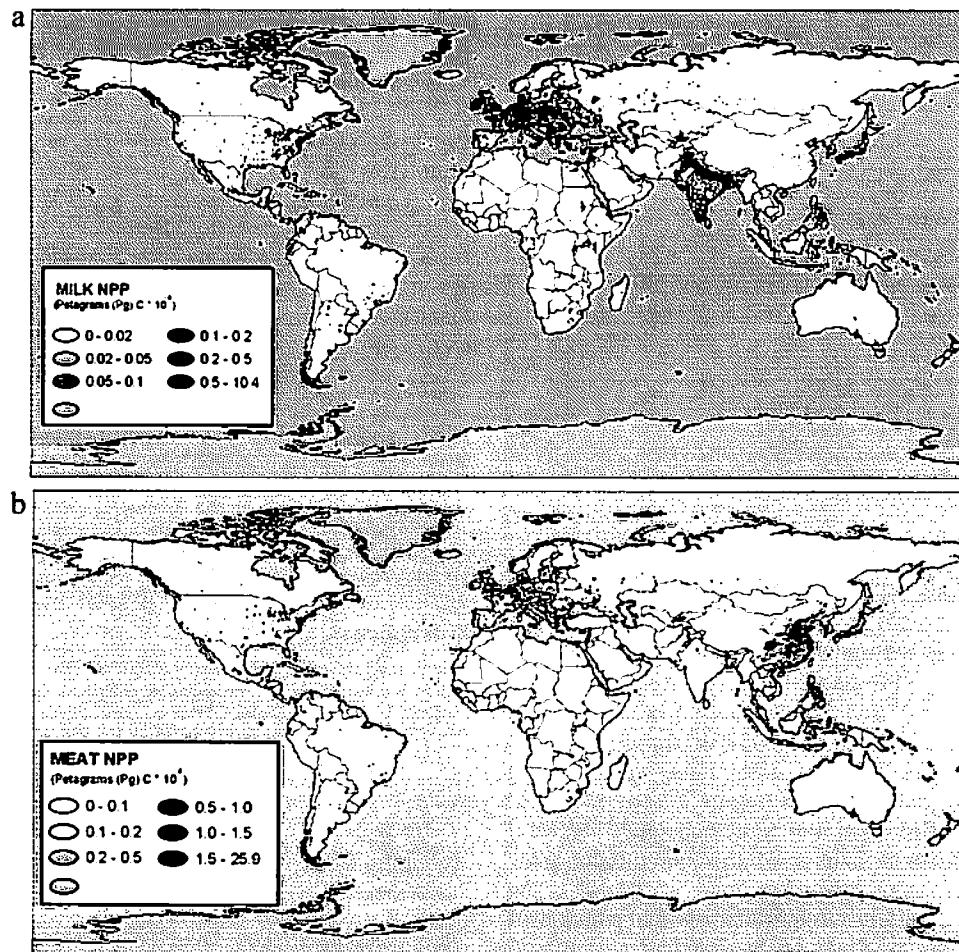
### 4.3. Comparing Rates of NPP Supply and Demand

[21] Comparing our global total values for NPP supply and demand, we find that humans appropriate approximately 20% of terrestrial NPP globally, with low and high estimates of 14% and 26%, respectively (Table 1). We were also able to show spatial patterns revealing the regional balance between NPP supply and demand (Figure 2c). Since the NPP demand estimates are ultimately tied to population, the spatial comparisons show the direction toward which NPP carbon must flow. Specific sources for the NPP are not explicitly delineated in this representation. However, socio-economically meaningful comparisons can be made constraining the analysis using regional or national boundaries. For example, some regions, such as western Europe and south central Asia, consume more than 70% of their regional NPP supply. Conversely, NPP demand in other regions is less than 15% of supply, with the lowest value of

**Table 2.** NPP Required for Selected Regions (Intermediate Estimate)

Region <sup>a</sup>	Population, millions	Per Capita NPP Required, Mt	NPP Supply, Pg	Total NPP Required "Demand," Pg	NPP Required As % of Supply
Africa	742	2.08	12.50	1.55	12.40
East Asia	1400	1.37	3.02	1.91	63.25
South-central Asia	1360	1.21	2.04	1.64	80.39
Western Europe	181	2.86	0.72	0.52	72.22
North America	293	5.40	6.67	1.58	23.69
South America	316	3.11	16.10	0.98	6.09

<sup>a</sup>Regions [United Nations Population Division, 2002].



**Figure 3.** Maps of NPP required for (a) milk and (b) meat products. See color version of this figure in the HTML.

about 6% in South America (Table 2). At more local scales, these spatial differences in NPP balance are even more striking, varying from nearly 0% of local NPP in sparsely populated areas to over 30,000% in large urban centers.

[22] In addition to the basic drivers of population and consumption level, cultural differences in the amount of NPP required are also visible through product preferences as expressed within our categorization of the NPP-based products. This is clearly seen in the graphic representations of NPP required for meat versus milk products (most especially between south Asia and east Asia; see Figures 3a and 3b).

#### 4.4. Drivers of Demand

[23] The population-consumption based approach for estimating NPP demand allows for some basic socioeconomic factors to be introduced as forcings to the NPP-carbon cycle. *Holdren and Ehrlich [1974]* introduced a simple relationship ( $I = PAT$ ) describing the overall ecological impact ( $I$ ) of human activities as product of population size ( $P$ ), affluence ( $A$ ) and technology ( $T$ ). Our model reflects the influence of these three factors in the form of population numbers ( $P$ ), per capita consumption level ( $A$ , affluence), and the different harvesting and milling efficiencies used for developing and industrialized countries ( $T$ ). To explore this relationship in the context of NPP-carbon, we

performed quantitative assessment of the impact of each of these factors under different combinations of  $P$ ,  $A$ , and  $T$  using current conditions (1995) and estimates of future population.

[24] The role of population is obvious despite vast differences in consumption among nations. For example, Asia (east and south central Asia), with almost half the world's population (Table 2), appropriates 72% of its regional NPP supply despite having the lowest per capita consumption of any region ( $1.29 \text{ Mt} \cdot \text{yr}^{-1}$ ). A model simulation assuming no appreciable change in global NPP and allowing the population to grow to 8.92 billion by 2050 shows that RNPP could rise to 17.4 PgC or nearly 31% of global NPP (combination 1, Table 3).

[25] Affluence also plays a significant role. From our intermediate calculation, we find the average annual per capita NPP required for industrialized countries (3.2 Mt, Metric tons) is almost double that of the developing nations (1.8 Mt) which host 83% of the global population. If the per capita NPP required of developing nations is increased to match that of industrialized countries without increase in technological efficiency, NPP required increases to 20.2 PgC or 35% of current global NPP (combination 2). If we increase Technology and Affluence together (combination 3), the effect of rising consumption is considerably moderated by; better processing and harvest efficiencies and a reduction

**Table 3.** Changes in NPP Required With Shifts in Population, Affluence, and Technology<sup>a</sup>

Combination	P <sup>b</sup>	A <sup>c</sup>	T <sup>d</sup>	NPP Required, PgC
1	↑	—	—	17.42
2	—	↑	—	20.19
3	—	↑	↑	16.26 <sup>e</sup>
4	↑	↑	—	31.59
5	↑	↑	↑	25.5 <sup>e</sup>

<sup>a</sup>Dashes denote no change from baseline (i.e., 1995 intermediate estimate); up arrows denote an increase relative to baseline.

<sup>b</sup>Population increase from 5.69 billion (global population in 1995) to 8.92 billion (estimated global population in 2050 [United Nations Population Division, 2002]).

<sup>c</sup>Affluence increase applies average per capita consumption of industrialized countries (in 1995) for all countries.

<sup>d</sup>Technology increase applies technological efficiencies of industrialized countries (in 1995) to all countries.

<sup>e</sup>Per capita fuel wood use in developing countries reduced to average for industrialized countries in 1995.

in wood fuel use to levels comparable with industrialized countries. In this case, NPP required is 16.3 PgC. Even with better technology, however, increased consumption in the developing countries will have significant regional impacts compared to current NPP. In south central Asia, for example, regional NPP required would grow from 80% to a regionally unsustainable rate of 224% of supply. Without major changes to the productivity of the landscape, it would take more than 2 years for this region's ecosystems to produce the amount of food and fiber consumed by local populations in 1 year. A change of this magnitude, aside from signaling increasing ecological impoverishment in the region itself, would certainly require substantial imports of NPP, creating greater pressure on natural and agricultural systems worldwide.

[26] The positive influence of technology is best seen in the following example. If both population and affluence are increased without change in technology (combination 4), global NPP required increased to 31.6 PgC or nearly 56% of current supply. However, if technology is improved as well, the global NPP required increases to only 25.5 PgC (combination 5).

[27] These combinations help highlight potential changes in NPP demand that may result from different development trajectories. As the human population and per capita consumption increase, pressure will be exerted on global ecosystems to increase NPP supply to meet the growing demand.

## 5. Conclusions

[28] Our results show one dimension of the human interaction with the NPP carbon cycle by comparing the rate of human NPP demand for products generated on land with the average rate of supply for the mid 1990s. The use of consumption data provides an independent estimate of NPP demand (eliminating circularity issues when comparing to satellite-based estimates of supply) and allows shifts in socioeconomic conditions to be readily incorporated. Because the FAO data reflect the influence of population, consumption level, and style (product preferences), and our model included the effect of technology through harvest and processing efficiencies on NPP demand, we were able to model current conditions as well as potential future trajec-

tories. This approach does not explicitly portray the spatial aspect of the sources of NPP required by various populations of consumers. It shows an endpoint-oriented gradient of NPP carbon flow spatially oriented around population distribution. When constrained by physical or political boundaries, this viewpoint is useful for elucidating NPP supply and demand rate balance issues around conservation, policy, and food security. In order to fully account for impacts to particular ecosystems and land surface climatology, this approach needs to be augmented by identifying the specific source areas for NPP required as well as an accounting of the fate of the carbon with respect to relocation or transport.

[29] **Acknowledgments.** We would like acknowledge our colleagues Taylor Ricketts and Colby Loucks at the World Wildlife Fund, Robert Harriss of the Houston Advanced Research Center and the National Center for Atmospheric Research, and William Lawrence of the Bowie State University for their advice and scientific input. We also thank Garik Gutman of NASA's Land Cover Land Use Change Program for funding this research.

## References

- Bishop, J. K. B., and W. B. Rossow (1991), Spatial and temporal variability of global surface solar irradiance, *J. Geophys. Res.*, **96**, 16,839.
- Cardoch, L. J., W. Day, and C. Ibanez (2002), Net primary productivity as an indicator of sustainability in the Ebro and Mississippi Deltas, *Ecol. Appl.*, **12**, 1044.
- Center for International Earth Science Information Network (2000), Gridded population of the world (GPW), version 2, Columbia Univ., Palisades, N. Y.
- Council for Agricultural Science and Technology (1999), Animal agriculture and global food supply, *Task Force Rep. R135*, Ames, Iowa. (Available at [http://www.cast-science.org/cast/pub/anag\\_nr.htm](http://www.cast-science.org/cast/pub/anag_nr.htm))
- Cramer, W., et al. (1999), Comparing global models of terrestrial net primary productivity (NPP): Overview and key results, *Global Change Biol.*, **5**, 1.
- Daily, G. C., et al. (1997), Ecosystem services: Benefits supplied to human societies by natural ecosystems, *Issues Ecol.*, **2**, 1.
- DePadua, D. B. (1978), Rice post-production and processing: Its significance to agricultural development, in *Accelerating Agricultural Development*, edited by J. D. Drilon and G. F. Sanguiguit, p. 135, SERCA Coll., Laguna, Philippines.
- Dykstra, D. P. (2001), The old and new of reduced impact logging, *Trop. For. Update*, **11**, 3.
- Enters, T. (2001), Trash or treasure? Logging and mill residues in Asia and the Pacific, *Rap. Publ. 2001/16*, Asia-Pac. For. Comm., FAO Reg. Off. for Asia and the Pac., Bangkok, Thailand. (Available at <http://www.fao.org/DOCREP/003/X6966E/X6966E00.htm>)
- Field, C. B. (2001), Sharing the garden, *Science*, **294**, 2490.
- Haberl, H. (1997), Human appropriation of net primary production as an environmental indicator: Implications for sustainable development, *Ambio*, **26**, 143.
- Haberl, H., F. Krausmann, K. H. Erb, and N. B. Schulz (2002), Human appropriation of net primary production, *Science*, **296**, 5575.
- Hall, F. G., G. Collatz, S. Los, E. Brown de Colstoun, and D. Landis (Eds.) (2005), ISLSCP Initiative II [DVD/CD-ROM], NASA, Washington D. C.
- Hansen, M., R. 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.
- Hasselmann, K., et al. (2003), The challenge of long-term climate change, *Science*, **302**, 1923.
- Hicke, J. A., G. P. Asner, J. T. Randerson, C. Tucker, S. Los, R. Birdsey, J. C. Jenkins, and C. Field (2002), Trends in North American net primary productivity derived from satellite observations, 1982–1998, *Global Biogeochem. Cycles*, **16**(2), 1018, doi:10.1029/2001GB001550.
- Holdren, J. P., and P. R. Ehrlich (1974), Human population and the global environment, *Am. Sci.*, **62**, 282.
- Imhoff, M. L., et al. (2004), Global patterns in human consumption of net primary production, *Nature*, **429**, 870.
- Ince, P. J. (2000), Industrial wood productivity in the United States 1900–1998, *Res. Note FPL-RN-0272*, For. Prod. Lab., For. Serv., U.S. Dep. of Agric., Madison, Wis.

- Kumar, M., and J. L. Monteith (1981), Remote sensing of crop growth, in *Plants and the Daylight Spectrum*, edited by H. Smith, p. 133, Elsevier, New York.
- Lieth, L. (1975), The measurement of caloric values, in *Primary Productivity of the Biosphere*, edited by H. Lieth and R. H. Whittaker, p. 119, Springer, New York.
- Los, S. O., G. J. Collatz, P. J. Sellers, C. M. Malmstrom, N. H. Pollack, R. S. DeFries, L. Bounoua, M. T. Parris, C. J. Tucker, and D. A. Dazlich (2000), A global 9-yr biophysical land surface dataset from NOAA AVHRR data, *J. Hydrometeorol.*, **1**, 183.
- Marland, G., and B. Schlamadinger (1997), Forests for carbon sequestration or fossil fuel substitution? A sensitivity analysis, paper presented at XI World Forestry Congress, U. N. Food and Agric. Organ., Antalya, Turkey. (Available at <http://www.fao.org/forestry/foda/wforcong/publi/v1/t4e3.HTM>)
- Monteith, J. L. (1977), Climate and efficiency of crop production in Britain, *Philos. Trans. R. Soc. London, Ser. B*, **281**, 277.
- Nemani, R. R., C. D. Keeling, H. Hashimoto, W. M. Jolly, S. C. Piper, C. J. Tucker, R. B. Myneni, and S. W. Running (2003), Climate-driven increases in global terrestrial net primary production from 1982 to 1999, *Science*, **300**, 1560.
- Oltjen, J. W., M. R. George, and D. J. Drake (1992), Dynamic allocation of multiple forage sources for beef cattle herds, in *Computers in Agricultural Extension Programs*, edited by D. G. Watson, F. S. Zazueta, and A. B. Bottcher, p. 58, Am. Soc. of Agric. Eng., St. Joseph, Mich.
- Pimm, S., and J. Gittleman (1992), Biological diversity: Where is it?, *Science*, **255**, 940.
- Postel, S. L., G. C. Daily, and P. R. Ehrlich (1996), Human appropriation of renewable fresh water, *Science*, **271**, 785.
- Potter, C. S., J. T. Randerson, C. B. Field, P. A. Matson, P. M. Vitousek, H. A. Mooney, and S. A. Klooster (1993), Terrestrial ecosystem production: A process model based on global satellite and surface data, *Global Biogeochem. Cycles*, **7**(4), 811–842.
- Pulkki, R. E. (1997), Literature synthesis on logging impacts in moist tropical forests, *Working Pap. GFSS/WP/06*, For. Prod. Div., U. N. Food and Agric. Organ., Rome.
- Rojstaczer, S., S. M. Sterling, and N. J. Moore (2001), Human appropriation of photosynthesis products, *Science*, **294**, 2549.
- Rosegrant, M. W., and S. A. Cline (2003), Global food security: Challenges and policies, *Science*, **302**, 1917.
- Sala, O. E., et al. (2000), Global biodiversity scenarios for the year 2100, *Science*, **287**, 1770.
- Schimel, D., et al. (2000), Contribution of increasing CO<sub>2</sub> and climate to carbon storage by ecosystems in the United States, *Science*, **287**, 2004.
- Schmincke, K. H. (1995), Forest industries: Crucial for overall socio-economic development, *Unasylva*, **182**, 46.
- Sellers, P. J. (1985), Canopy reflectance, photosynthesis and transpiration, *Int. J. Remote Sens.*, **8**, 1335.
- Sellers, P. J., D. A. Randall, G. J. Collatz, J. A. Berry, C. B. Field, D. A. Dazlich, C. Zhang, and L. Bounoua (1996a), A revised land surface parameterization (SiB2) for atmospheric GCMs. Part I: Model formulation, *J. Clim.*, **9**, 676.
- Sellers, P. J., S. O. Los, C. J. Tucker, C. O. Justice, D. A. Dazlich, G. J. Collatz, and D. A. Randall (1996b), A revised land surface parameterization (SiB2) for atmospheric GCMs. Part II: The generation of global fields of terrestrial biophysical parameters from satellite data, *J. Clim.*, **9**, 706.
- Sere, C., and H. Steinfeld (1996), World livestock production systems, *FAO Animal Prod. Health Pap.* **127**, U. N. Food and Agric. Organ., Rome. (Available at <http://www.fao.org/WAICENT/FAOINFO/AGRICULT/AGA/LSPA/Paper127/cover1.htm>)
- Shea, D. (1986), Climatological atlas: 1950–1979, surface air temperature, precipitation, sea level pressure, and sea surface temperature, Natl. Cent. for Atmos. Res., Boulder, Colo.
- Skog, K. E., P. J. Ince, and R. W. Haynes (1998), Wood fiber supply and demand in the United States, paper presented at Annual Meeting, Forest Products Study Group Workshop, N. Am. For. Comm., For. Prod. Soc., Merida, Mexico. (Available at <http://www.fpl.fs.fed.us/documnts/pdf2000/skog00a.pdf>)
- Smil, V. (1983), *Biomass Energies: Resources, Links, Constraints*, Springer, New York.
- Smith, H. J. (2003), Series introduction: The shape we're in, *Science*, **302**, 1171.
- Tucker, C. J., D. A. Slayback, J. E. Pinzon, S. O. Los, R. B. Myneni, and M. G. Taylor (2001), Higher northern latitude NDVI and growing season trends from 1982 to 1999, *Int. J. Biometeorol.*, **45**, 184.
- United Nations Food and Agriculture Organization (1997), Estimated post-harvest losses of rice in Southeast Asia, New York. (Available at <http://www.fao.org/NEWS/FACTFILE/IMG/FF9712-e.pdf>)
- United Nations Food and Agriculture Organization (2001), *FAOSTAT 2001*, Rome, Italy.
- United Nations Population Division (2002), UN-POP world population prospects: The 2000 revision, Dep. of Econ. and Soc. Affairs, New York.
- Vitousek, P. M., P. Ehrlich, A. Ehrlich, and P. M. Matson (1986), Human appropriation of the products of photosynthesis, *Bioscience*, **36**, 368.
- Vitousek, P. M., H. A. Mooney, J. Lubchenco, and J. M. Melillo (1997), Human domination of Earth's ecosystems, *Science*, **277**, 494.
- Wackernagel, M., et al. (2002), Tracking the ecological overshoot of the human economy, *Proc. Natl. Acad. Sci. U. S. A.*, **99**, 9266.
- Walsh, M. E., et al. (1999), Biomass feedstock availability in the United States: 1999 state level analysis, Bioenergy Inf. Network, Bioenergy Feedstock Dev. Program, Oak Ridge Natl. Lab., Oak Ridge, Tenn. (Available at <http://bioenergy.ornl.gov/resourcedata/>)
- World Resources Institutes (1998), *Disappearing Food: How Big Are Post-harvest Losses?*, Washington, D. C. (Available at <http://www.wri.org/trends/foodloss.html>)
- Zhao, M., F. Heinsch, R. R. Nemani, and S. W. Running (2005), Improvements of the MODIS terrestrial gross and net primary production global data set, *Remote Sens. Environ.*, **95**, 164.
- Zobler, L. (1986), A world soil file for global climate modeling, *NASA Tech. Memo.* **87802**.

L. Bounoua and M. L. Imhoff, Biospheric Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA. (marc.l.imhoff@nasa.gov)



# Use of ISLSCP II data to intercompare and validate the terrestrial net primary production in a land surface model coupled to a general circulation model

Li Dan,<sup>1</sup> Jinjun Ji,<sup>1</sup> and Yong He<sup>2</sup>

Received 30 June 2006; revised 2 November 2006; accepted 10 November 2006; published 24 January 2007.

[1] Using the global terrestrial NPP and climate data from International Satellite Land Surface Climatology Project Initiative II (ISLSCP II) and additional NPP data, we validated the NPP simulations and explored the relationship between NPP and climate variation in a global two-way coupled model AVIM-GOALS. The strength of this study is that the global simulations produced will enhance interactive climate and vegetation study; however, the weakness is that the NPP distribution is not fully reproduced in some regions, because of the coarse model resolution and climate biases. Global NPP is spatially consistent with IGBP NPP and MODIS data, though there is a discrepancy in NPP (significantly lower values) for boreal forests and tundra, due to the underestimated temperature. The NPP distribution in China indicates agreement with IGBP data, but the IGBP data in northeast China (around 48°N) seem to be slightly high in contrast with other modeled and estimated NPP. The spatial structure of NPP in USA and Australia roughly corresponds to the IGBP NPP data and GPPDI Gridded data, and a possible lower value of GPPDI data in central Australia exists, in contrast with other NPP data. The globally averaged NPP of 447.47 g C m<sup>-2</sup> year<sup>-1</sup> is close to the 450.42 g C m<sup>-2</sup> year<sup>-1</sup> from IGBP data. The global relative error of simulated NPP against IGBP data is about 20% and is comparable to other global biogeochemical models. The meridional variation of globally zonal mean NPP corresponds more to the meridional change of precipitation than temperature. The global NPP for all vegetation types is highly statistically significant in correlation with precipitation.

**Citation:** Dan, L., J. Ji, and Y. He (2007), Use of ISLSCP II data to intercompare and validate the terrestrial net primary production in a land surface model coupled to a general circulation model, *J. Geophys. Res.*, 112, D02S90, doi:10.1029/2006JD007721.

## 1. Introduction

[2] Net primary production (NPP) is an important variable for global carbon cycle and the feedback between terrestrial ecosystems and atmosphere, and many studies have addressed the view [Ciais *et al.*, 1995; Cramer *et al.*, 1999; Ruimy *et al.*, 1999; Running *et al.*, 1999; Nemani *et al.*, 2003; Williams *et al.*, 2005]. NPP integrates climatic, ecological, geochemical, and human influences on the biosphere [Nemani *et al.*, 2003]. NPP is the difference between total photosynthesis (Gross Primary Production, GPP) and total plant respiration in an ecosystem [Clark *et al.*, 2001]. The process regulating NPP at regional to global scales can only be addressed with observational studies and models [Williams *et al.*, 2005], and modeling is required when measurements cannot provide a complete view of

biospheric biogeochemical activity [Running *et al.*, 1999]. There have been many biogeochemical models designed to simulate NPP, so the observed or field data are needed to calibrate and validate the models. Scurlock *et al.* [1999] also cited the importance of obtaining high-quality data of NPP from around the globe for comparing various models. Terrestrial NPP data are more widely available than other estimates of biosphere-atmosphere exchanges of carbon such as GPP and net ecosystem exchange (NEE), but there are significant problems with inconsistency in measurement techniques between NPP studies separated in space and time [Scurlock and Olson, 2002]. As a result, although there are many site-level or point NPP field measurements, the data cannot be used as direct validation of NPP simulations at regional or global scale without the appropriate scale transformation [Running *et al.*, 1999; Scurlock and Olson, 2002; Zheng *et al.*, 2003; Ahl *et al.*, 2005]. A consistent NPP data set suitable for global NPP model validation is long overdue [Zheng *et al.*, 2003]. Compared to the more studied model validation of climate simulations using observation data, the simulated NPP needs more validations because of the relatively insufficient global data. Thus much work needs to be carried out at the regional and global scale for model validation of NPP.

<sup>1</sup>START Regional Center for Temperate East Asia and Key Laboratory of Regional Climate-Environment for Temperate East Asia, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China.

<sup>2</sup>National Climate Center, China Meteorological Administration, Beijing, China.

[3] Many studies have been implemented on the NPP simulations with the spatial scale ranging from some sites and local plots, to the entire country and continent for some ecosystem types, and even to global distribution like *Field et al.* [1998] and *Cao et al.* [2005]. The studied temporal scale of terrestrial carbon cycles can be arranged in order as weekly, seasonal, monthly, annual, decadal to geological time [Zeng, 2003] (geological time including such examples as the Holocene [Wang et al., 2005] and the last glacial maximum [He et al., 2005]). Projected simulations under several emission scenarios have been made, e.g., the work of *Cox et al.* [2000], to explore the long-term carbon absorbed by global terrestrial ecosystems from the year of 1850 to 2100. The NPP model intercomparison work [Cramer et al., 1999] can compare the simulated model results and reveal the spatial change of the global terrestrial carbon cycle, however, the same problem of the less model validation still exists because of the inadequate gridded NPP field data.

[4] In recent years, NPP, which is representative of the terrestrial carbon cycle has been incorporated into general circulation models (GCMs), and the terrestrial carbon cycle has been linked to climate change. The physical and biological processes at the land surface coupled to GCMs has been called two-way coupling [Dan et al., 2005]. The two-way coupled model AVIM-GOALS integrates the variation of the terrestrial carbon cycle and climatic change such that the interactive biosphere and atmosphere is revealed, compared to the prescribed structure of terrestrial vegetation like leaf area index (LAI) of some land surface models [Dan et al., 2005].

[5] The objectives of this paper are to (1) use modeled and estimated NPP data sets including the IGBP Global NPP Intercomparison Data and GPPDI Gridded NPP Data to intercompare and validate the NPP simulations at regional (the range of a country, such as China, USA and Australia) and global scales, respectively; (2) reveal the strength and weakness of the IGBP and GPPDI NPP data in the spatial distribution; and (3) explore the relationship between NPP and climate variation in the climate-vegetation coupled model.

## 2. Model and Data Descriptions

### 2.1. Atmosphere-Vegetation Interaction Model (AVIM)

[6] AVIM [Ji, 1995] is the land surface model incorporating the physical [Ji and Hu, 1989] and ecophysiological processes. The energy, momentum, and water exchange between land and air is linked with vegetation growth processes and terrestrial ecosystem carbon cycles, which enables the interaction between terrestrial ecosystems and climate. The model has one canopy layer and ten uneven soil layers, and the physical and chemical boundary conditions in the deepest soil layer are assumed constants. The surface hydrological processes include the interception of precipitation and drainage, evapotranspiration from the canopy and evaporation from the ground, surface runoff and infiltration, snowpackage and melting, and water transfer into atmosphere. The canopy photosynthesis, plant respiration and the soil carbon loss through respiration are related to the temperature of the canopy and soil. The vegetation morphology is affected by the accumulation

and consumption of dry matter, and the surface dynamical parameters (such as albedo, zero plane displacement and roughness) as well as resistance parameters, including the stomatal resistance, are affected by the physiological processes.

[7] The NPP equation can be expressed as follows:

$$NPP = GPP - R_m - R_g \quad (1)$$

where  $GPP$  is the gross photosynthesis rate excluding photorespiration,  $R_m$  is the maintenance respiration rate and  $R_g$  is growth respiration rate.  $GPP$  is a function of:  $CO_2$  concentration in the stomata, foliage temperature, leaf water potential, the photosynthetic active radiation at the top of canopy and the leaf area index.  $R_m$  is function of biomass and vegetation tissue temperature. Ji [1995] has described the details of all the functions.

[8] AVIM has simulated the spatial pattern of NPP and its temporal variation at a seasonal/interannual scale in many basic ecosystems including forests, grasslands, shrublands and croplands [Ji and Yu, 1999; Li and Ji, 2001; Lu and Ji, 2002, 2006]. The simulations reveal the satisfactory variation values of NPP related to changes of air temperature and precipitation. AVIM has participated in the Ecosystem Model-Data Intercomparison (EMDI) organized by IGBP project on Global Analysis, Interpretation and Modeling (GAIM), and the simulated NPP of AVIM agrees well with the measured NPP. The intercomparison result can be found in the figure "EMDI Initial Results: 11 models and field NPP data at 87 sites" and AVIM is the leftmost model, which is detailed at the following website: [http://gaim.unh.edu/Structure/Intercomparison/EMDI/phase1info/ESA\\_EMDI\\_p2.ppt](http://gaim.unh.edu/Structure/Intercomparison/EMDI/phase1info/ESA_EMDI_p2.ppt).

### 2.2. Global Ocean-Atmosphere-Land System (GOALS) and the Coupling Strategy

[9] The IAP/LASG GCM GOALS has 9 layers in the atmosphere and is truncated rhombically at the 15 wave numbers [Wu et al., 1996, 1997]. The reduction of a standard atmosphere proposed by Zeng [1963] and Phillips [1973] was used in the dynamical framework to improve the GCM performance. A new k-distribution radiation scheme [Wang et al., 2000] was introduced to make the cloud radiation processes more reasonable. The horizontal resolution is  $7.5^\circ$  longitude by  $4.5^\circ$  latitude. The 1979–1988 monthly mean climatologically observed SST and sea ice were taken from the Atmospheric Model Intercomparison Project (AMIP). The integration time step is 30 min and the land-air coupling is synchronous.

[10] The GOALS GCM developed at the Institute of Atmospheric Physics/State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics (IAP/LASG), performances well in global climate simulations, especially for the large-scale features and seasonal cycles [Liu et al., 2001]. It has been used to study such climate changes as the maintenance mechanism of the subtropical anticyclone [Liu et al., 2004]. This model was cited in the third IPCC report [Houghton et al., 2001] and used by other scientists to make the future climate projection [e.g., Coquard et al., 2004].

[11] For the purpose of terrestrial ecosystem simulations, the resolution of GOALS is coarse. The land surface model

**Table 1.** Global Vegetation Classifications [Dan et al., 2005]

Index	Types
1	tropical rainforest
2	broad leaf deciduous trees
3	broad leaf and needle leaf trees
4	needle leaf evergreen trees
5	needle leaf deciduous trees
6	broadleaf trees with ground cover
7	ground cover only
8	broadleaf shrub with ground cover
9	broadleaf shrubs with bare soil
10	dwarf trees with ground cover
11	bare soil
12	crops
13	ice
0	water

AVIM is designed at relatively fine grid of  $1.5^\circ$  by  $1.5^\circ$ , and the interface coupling adopts the nesting technique to link the two models. In other words, a unit grid of the GCM is subdivided into 15 grid cells down to the surface. The two-way coupled model AVIM-GOALS has run for 30 years, and we selected the last 15 years outputs for analysis.

[12] The initial values at land surface were derived from climatological averages of the offline run. The global terrestrial ecosystems were classified into 13 types listed in Table 1 [Dan et al., 2005], and the land cover data was derived from Dorman and Sellers [1989] with modifications for China according to the China Vegetation Map. The global soil texture data were taken from Zobler [1986] with 6 types of classification.

### 2.3. Climate Data, IGBP NPP Intercomparison Data, GPPDI Gridded NPP Data, and Other Modeled and Estimated NPP Data

[13] In this paper, the global terrestrial surface air temperature and precipitation data were provided by the International Satellite Land Surface Climatology Project Initiative II (ISLSCP II), and the data originated from the Climatic Research Unit (CRU), University of East Anglia. The data set is called CRU05 Mean Monthly Climatology (1961–1990) and its annual average is abbreviated to CRU temperature and CRU precipitation in this study (or CRU data). The data are  $0.5^\circ$  by  $0.5^\circ$  for global land areas excluding Antarctica and have been detailed by New et al. [1999].

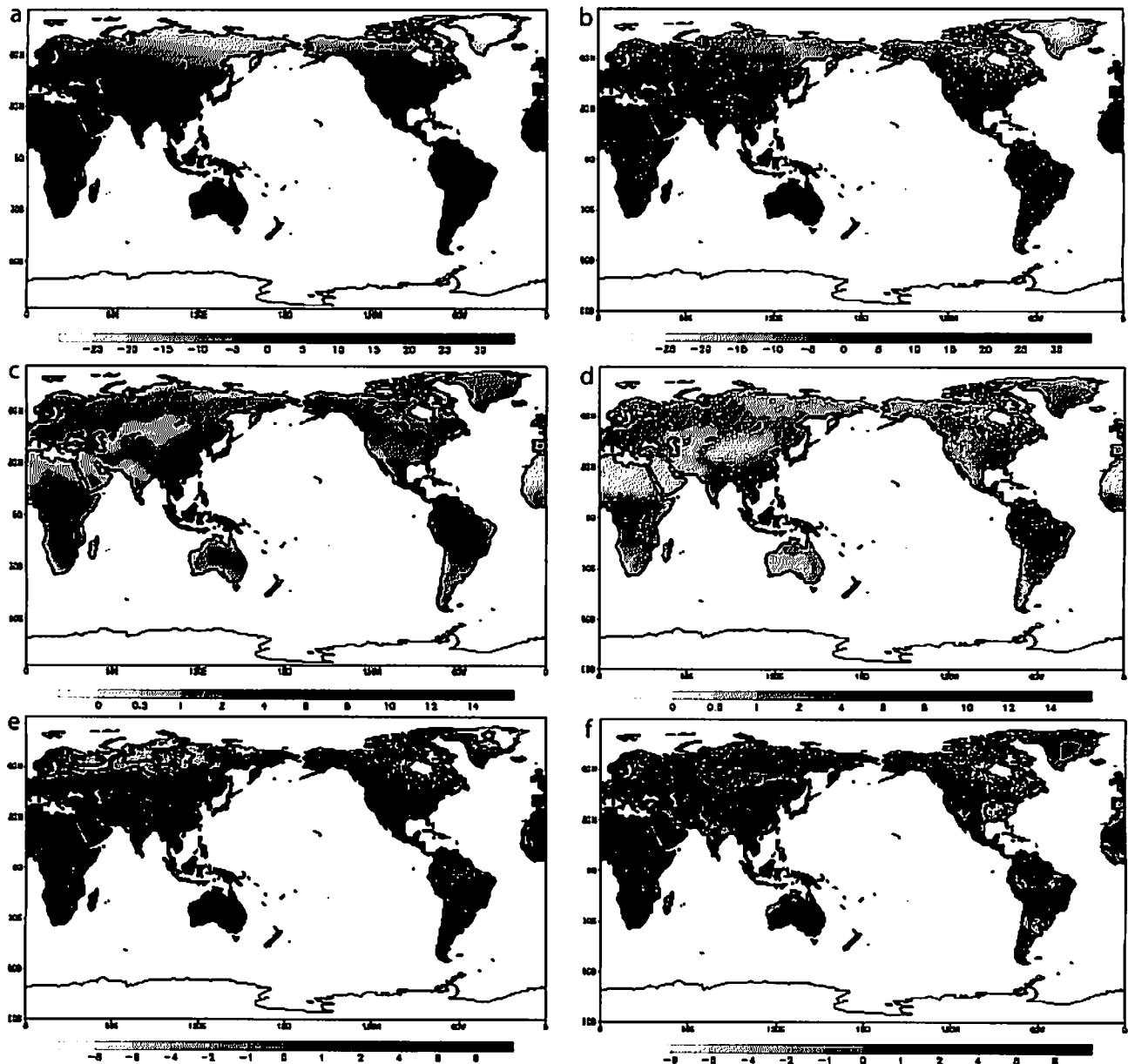
[14] The International Geosphere Biosphere Programme (IGBP) Global NPP Model Intercomparison Data (abbreviated to IGBP data here) were used to intercompare the simulated NPP of the two-way coupled model AVIM-GOALS. Global Primary Production Data Initiative (GPPDI) Gridded NPP Data (abbreviated to GPPDI data) were adopted to validate the NPP simulations of the global two-way coupled model. All of the annual mean IGBP and GPPDI NPP data were taken from the Website of ISLSCP II with the resolution of  $0.5^\circ$  by  $0.5^\circ$ . Global terrestrial half-degree IGBP data were derived from the original data containing gridded average NPP for 17 global models of biogeochemistry [Cramer et al., 1999]. GPPDI data [Olson et al., 2001; Zheng et al., 2003] contain 2335 half degree cells, and the field NPP data used to develop  $0.5^\circ$  grid cell estimates come from 15 sources worldwide.

[15] To compare and evaluate the simulated NPP spatial pattern, other NPP data from two global biogeochemical models (CASA and GloPEM), remote sensing data (MODIS NPP) and estimated NPP (ORNL NPP) were adopted in this study. CASA and GloPEM are two representative satellite-based models participating in IGBP NPP Model Intercomparison, and they are driven by NOAA/AVHRR data. The uniqueness is the entire model uses of satellite data of GloPEM without using any climatic driven variables observed on the ground [Cramer et al., 1999]. The multiyear averaged global NPP of CASA model at half degree resolution [Potter, 1999; Potter et al., 2003] can be downloaded at the Website [ftp://geo.arc.nasa.gov/pub/glemis/glemis\\_05/](ftp://geo.arc.nasa.gov/pub/glemis/glemis_05/). The global NPP of the GloPEM model at 8 km spatial resolution [Prince and Small, 2003] can be found at the Website <http://glcf.umiaccs.umd.edu/data/glopem/>. The multiyear mean GloPEM NPP has been aggregated to global  $0.5^\circ$  grid cell in this study. Improved global 1 km MODIS NPP data averaged during 2000–2003 [Zhao et al., 2006] can be downloaded at <ftp://ftp.ntsg.umd.edu/autofs/MODIS/5/MOD17A3.105.LATEST/>. The data were aggregated to global half degree by Dr. Zhao for our study. The estimated ORNL NPP data for global 2335  $0.5^\circ$ -grid cells [Zheng et al., 2004] were taken from Oak Ridge National Laboratory Distributed Active Archive Center (ORNL-DAAC), which is an update version of the GPPDI data and cover more areas of the three countries in this study.

## 3. Results and Analysis

### 3.1. Spatial Distribution of Global and Regional NPP at the Country Level

[16] Prior to presenting the global and the country-level NPP distribution, we make a short validation between the CRU data and the relevant simulations of the global climate. Figure 1 shows the temperature and precipitation of CRU data and the two-way coupled model, and the difference between simulated climate interpolated to a half degree using the bilinear interpolation method and CRU data is also presented. The globally annual mean temperature of AVIM-GOALS (Figure 1a) agrees with CRU temperature (Figure 1b), however, some marked regional biases exist across the globe, such as the cold bias of up to  $8^\circ\text{C}$  at northern high latitudes, especially in Eurasia and Greenland (Figure 1e). The lower surface air temperature can be attributed to the insufficient incident radiation in the region with underestimated downward shortwave radiation [Zhang et al., 2002]. A strong warm bias up to  $8^\circ\text{C}$  (Figure 1e) occurs along Tibetan Plateau because of the lower surface albedo of thin snow coverage. The simulated annual mean precipitation (Figure 1c) is generally consistent with CRU data (Figure 1d) in the global spatial distribution, but some regional biases are present (Figure 1f). The simulation in the tropical rain forests of Africa, Southeast Asia and South America is higher than CRU precipitation and magnitudes possibly exceeding 4 mm/day. The 2 mm/day overestimated precipitation occurs in eastern China and Australia, and a slightly lower simulation in eastern USA. The area-averaged temperature and precipitation at global and regional scale excluding Antarctica (Tables 2 and 3) shows the overall agreement with CRU data in magnitude despite



**Figure 1.** (a) Annual mean surface air temperature of the coupled model; (b) annual mean CRU air temperature from ISLSCP II, with units of °C; (c) annual mean precipitation of the coupled model; (d) annual mean CRU precipitation from ISLSCP II, with units of mm/day; (e) annual mean simulated temperature minus CRU temperature; and (f) annual mean simulated precipitation minus CRU precipitation. (Correlation coefficients of global climatological state are 0.96 for simulated and observed temperature and 0.73 for precipitation, at 99.9% confidence level of t-test.) See color version of this figure in the HTML.

the discrepancy of generally higher simulated temperature and precipitation. The global and regional differences between simulated climate and CRU data are comparable to other GCM simulations in the magnitude of the discrepancy with observations, such as the global climate simulation of *Foley et al.* [1998]. The spatial correlation coefficient between simulation and observation is another method to evaluate the performance of the GCM or RCM [Zhou and Qian, 1995; Lau and Nath, 2004] because it can represent the degree of agreement between two spatial patterns of temperature or precipitation [Haywood et al., 1997]. The global spatial correlation coefficient is 0.96 for

temperature and 0.73 for precipitation, which accounts for the fact that precipitation is inherently much noisier spatially than temperature. The values are similar to the coefficients of temperature and precipitation in previous studies, e.g., 0.92/0.5 for GCM and 0.94/0.7 for RCM of *Ju and Wang* [2006].

[17] NPP means the amount of net fixed carbon by vegetation per unit area and per unit time. The vegetation utilizes the photosynthetic active radiation from sunlight to convert the CO<sub>2</sub> into dry matter, which contains the matter and energy for the basic needs of humans. It is a major determinant of carbon sink on land and in the ocean, and a



**Table 2.** Area-Averaged Surface Air Temperature<sup>a</sup>

Time	Area			
	Global T Sim/Obs	China T Sim/Obs	USA T Sim/Obs	Australia T Sim/Obs
DJF mean	6.38/5.70	-1.71/-4.17	-1.09/0.30	32.72/27.58
JJA mean	20.89/19.52	21.41/19.62	25.79/21.96	17.44/15.07
Annual mean	14.29/12.96	11.34/8.55	12.97/11.36	26.11/21.76

<sup>a</sup>Units are °C. T, temperature; Sim, simulation; Obs, CRU observation data; DJF, December-January-February; JJA, June-July-August.

key regulator of ecological processes [Field *et al.*, 1998]. Consequently, many studies [Tian *et al.*, 1999; Cox *et al.*, 2000; Nemani *et al.*, 2003; Lu and Ji, 2006] have been done to determine a quantitative estimate of global and regional NPP, especially under the background of global warming.

[18] Global terrestrial NPP of AVIM-GOALS, IGBP data and MODIS data are shown in Figures 2a–2c. The model intercomparison NPP data (IGBP data) averaged for all 17 models is considered to a comprehensive representation of NPP fluxes [Cramer *et al.*, 1999]. MODIS NPP is the first continuous satellite-driven data set for repeated monitoring of global vegetation productivity at 1-km resolution over vegetated land at an 8-day interval [Zhao *et al.*, 2005, 2006]. The simulated NPP is similar to the IGBP data and MODIS NPP, with the highest NPP ( $>1000 \text{ g C m}^{-2} \text{ year}^{-1}$ ) in the tropical rain forests, the intermediate NPP ( $500\text{--}700 \text{ g C m}^{-2} \text{ year}^{-1}$ ) in temperate regions, and the lowest NPP ( $<200 \text{ g C m}^{-2} \text{ year}^{-1}$ ) in the cold or arid regions. However, the simulated NPP in boreal forests and tundra exhibits an obviously lower NPP belt (same hereafter), which is mainly due to the cold bias (Figure 1a) limiting the vegetation growth especially in boreal summer. For instance in northwest Europe ( $10\text{--}20^\circ\text{E}$ ), the simulated NPP can be  $200 \text{ g C m}^{-2} \text{ year}^{-1}$  lower than IGBP data south of  $63^\circ\text{N}$ . When validating the simulated NPP further in Sweden using GPPDI data (figure not shown here), we found that the simulated NPP in southern Sweden ( $56\text{--}63^\circ\text{N}$  and centered  $15^\circ\text{E}$ ) appears to be around  $120 \text{ g C m}^{-2} \text{ year}^{-1}$  lower because of the cold bias, and IGBP data are at least  $150 \text{ g C m}^{-2} \text{ year}^{-1}$  overestimated, given that the GPPDI NPP from field data is regarded as an actual distribution. In the Southern Hemisphere, the simulated NPP south of  $15^\circ\text{S}$  in Africa is larger than IGBP data mainly because of the vegetation type classification of broadleaf trees in contrast with the shrublands or grassland in other models such as C3 and C4 grasslands of BIOME-BGC [Running and Hunt, 1993; Hunt *et al.*, 1996]. The broadleaf trees under overestimated temperature and precipitation can lead to the higher NPP simulation.

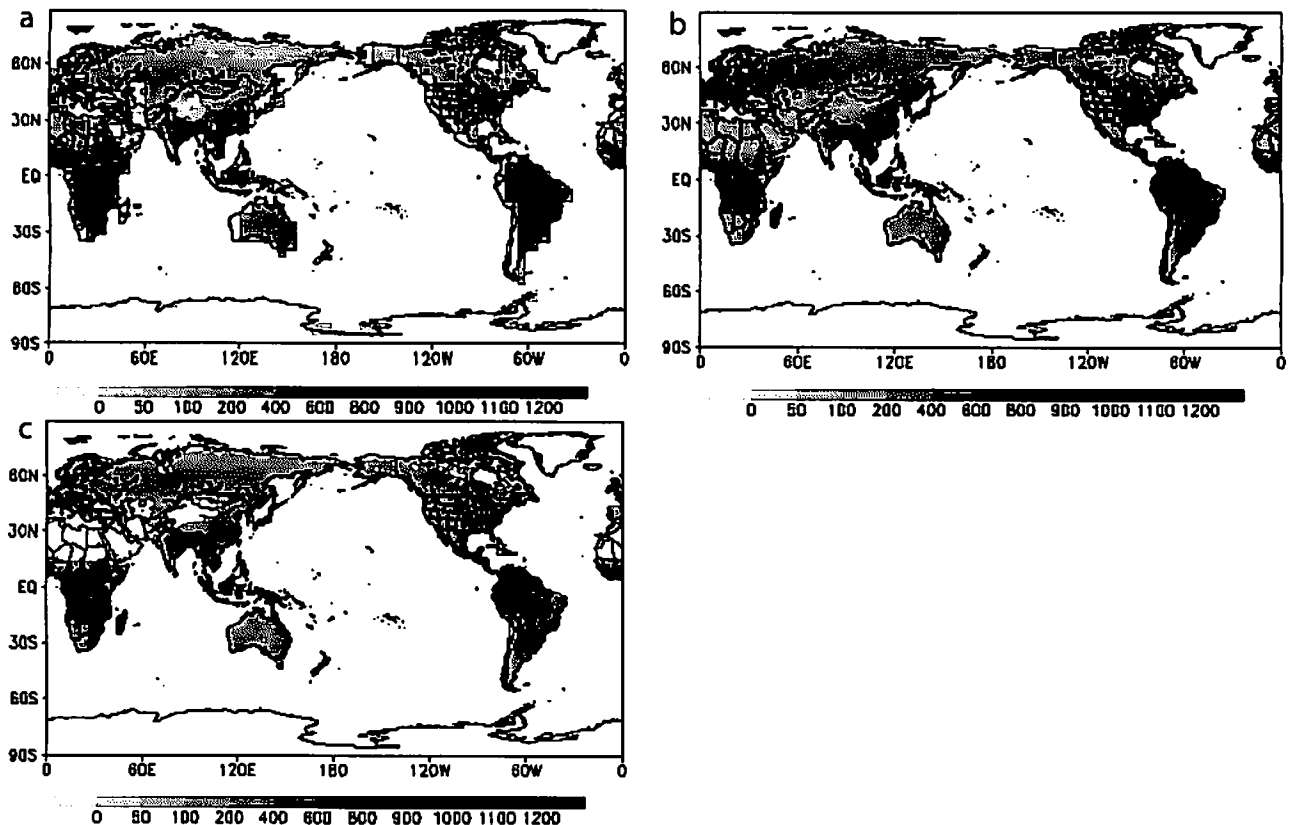
[19] For the sake of deep validation of NPP in spatial pattern, we present the simulated NPP of China, USA and Australia because GPPDI data do indeed have measured values in these three countries (available data is sparse over

China). Figure 3 shows the simulated NPP (Figure 3a), IGBP data (Figure 3d) and MODIS NPP (Figure 3e) in China. To compare the simulated NPP to other models, regional NPP of CASA and GloPEM is also presented in Figures 3b and 3c. The simulated NPP in eastern China south of  $35^\circ\text{N}$  is consistent with IGBP and MODIS data in that the magnitude is larger than  $500 \text{ g C m}^{-2} \text{ year}^{-1}$  and reaches  $700 \text{ g C m}^{-2} \text{ year}^{-1}$  more in southern China, which reflects the relatively large carbon flux in this area corresponding to the largest water and heat conditions for vegetation growth over China. The CASA and GloPEM model shows the similar spatial pattern with the exception of generally higher value of GloPEM and lower of CASA in southern China. The spatial pattern is reasonable compared to the modeled result of Tao *et al.* [2003] using the NOAA NDVI in the CEVSA model [Cao and Woodward, 1998] to calculate the climatological NPP distribution of China. Tao *et al.* [2003] shows a similar pattern south of  $35^\circ\text{N}$  in China, and other studies [e.g., Sun and Zhu, 2001; Chen *et al.*, 2002] use remote sensing data of vegetation to estimate the terrestrial NPP of China, presenting the same NPP magnitude south of  $35^\circ\text{N}$  in China. However, the NPP in northeast China simulated by AVIM-GOALS is much lower than IGBP and MODIS data, especially north of  $48^\circ\text{N}$ , which can be attributed to underestimated air temperature reaching  $4^\circ\text{C}$ . However, it is interesting to explore the actual NPP in this area further. Jiang *et al.* [1999] used the ground-based measurements of NPP from Forestry Ministry of China [1994] to present the spatial distribution of NPP in China. Their subdivided areas (9A, 9B and 10B) sum equivalent to the area north of  $48^\circ\text{N}$  in this study shows the NPP ranging from  $301.5$  to  $342 \text{ g C m}^{-2} \text{ year}^{-1}$ . The GPPDI data in this area is mainly  $350\text{--}400 \text{ g C m}^{-2} \text{ year}^{-1}$ . The simulated NPP of Tao *et al.* [2003] and the calculation of Wang *et al.* [2001] using the model of Zhou and Zhang [1995] show also the similar ranges. Consequently, we consider that the value  $350\text{--}400 \text{ g C m}^{-2} \text{ year}^{-1}$  seems to be the actual NPP north of  $48^\circ\text{N}$  in northeast China, which is reflected in MODIS data. The new calculated NPP of Cao *et al.* [2005] using AVHRR land data and observational climate to drive the GloPEM model over China (Figure 3f) and the NPP of ORNL data (Figure 3g), present the same NPP magnitude in this area. For the NPP distribution in western China, the

**Table 3.** Area-Averaged Precipitation<sup>a</sup>

Time	Area			
	Global P Sim/Obs	China P Sim/Obs	USA P Sim/Obs	Australia P Sim/Obs
DJF mean	2.64/2.01	1.68/0.44	1.93/1.71	2.86/2.41
JJA mean	2.51/2.45	3.37/4.33	1.97/2.30	1.02/0.68
Annual mean	2.70/2.15	2.57/2.00	1.95/1.98	1.89/1.31

<sup>a</sup>Units are mm/day. P, precipitation; Sim, simulation; Obs, CRU observation data; DJF, December-January-February; JJA, June-July-August.



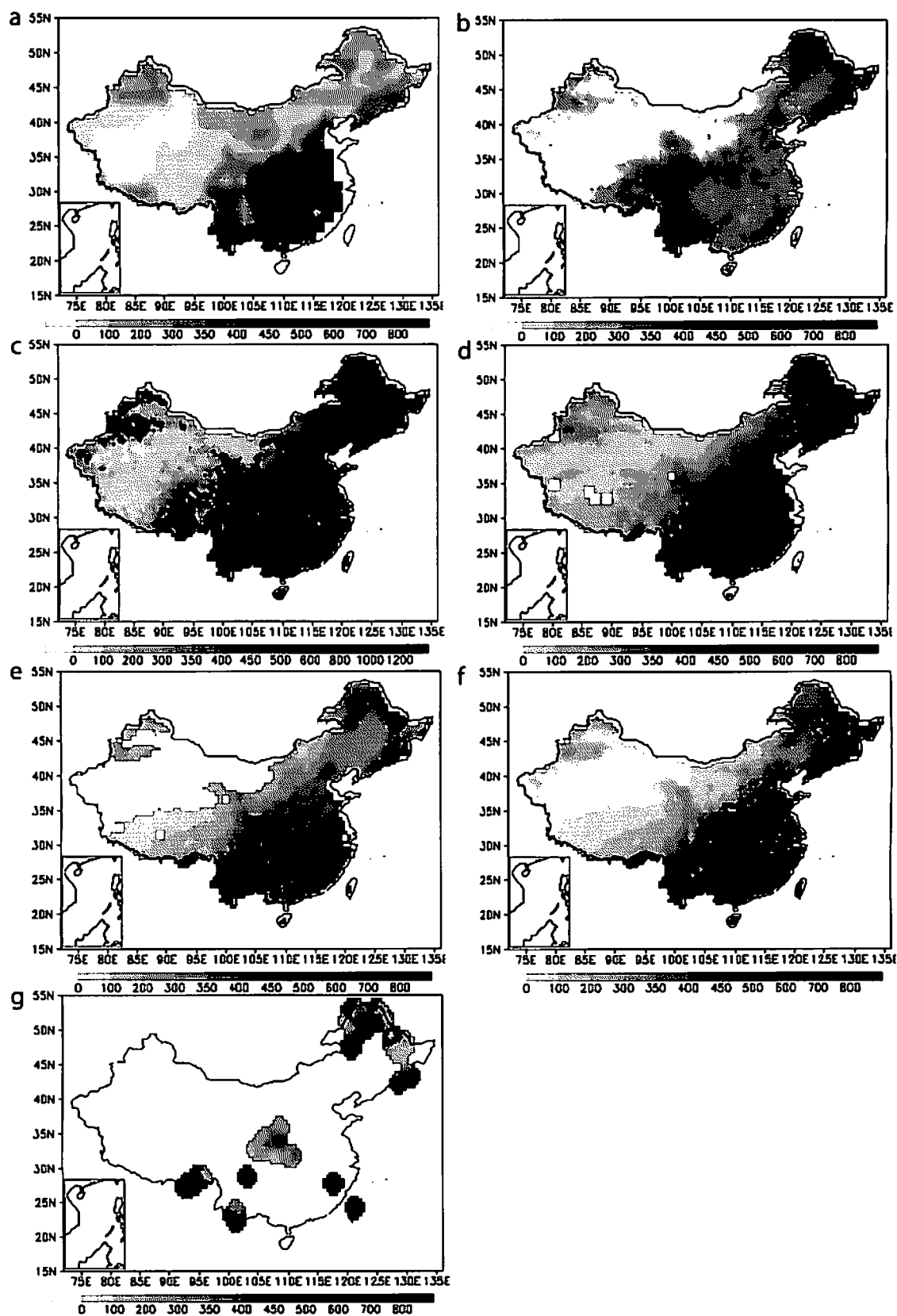
**Figure 2.** Globally annual mean NPP of (a) the coupled model AVIM-GOALS, (b) IGBP Global NPP Intercomparison data from ISLSCP II, and (c) MODIS NPP data. Units are  $\text{g C m}^{-2} \text{ year}^{-1}$ . See color version of this figure in the HTML.

simulated NPP west of  $95^{\circ}\text{E}$  agrees well with IGBP data and the agreement is consistent with the work of *Tao et al.* [2003].

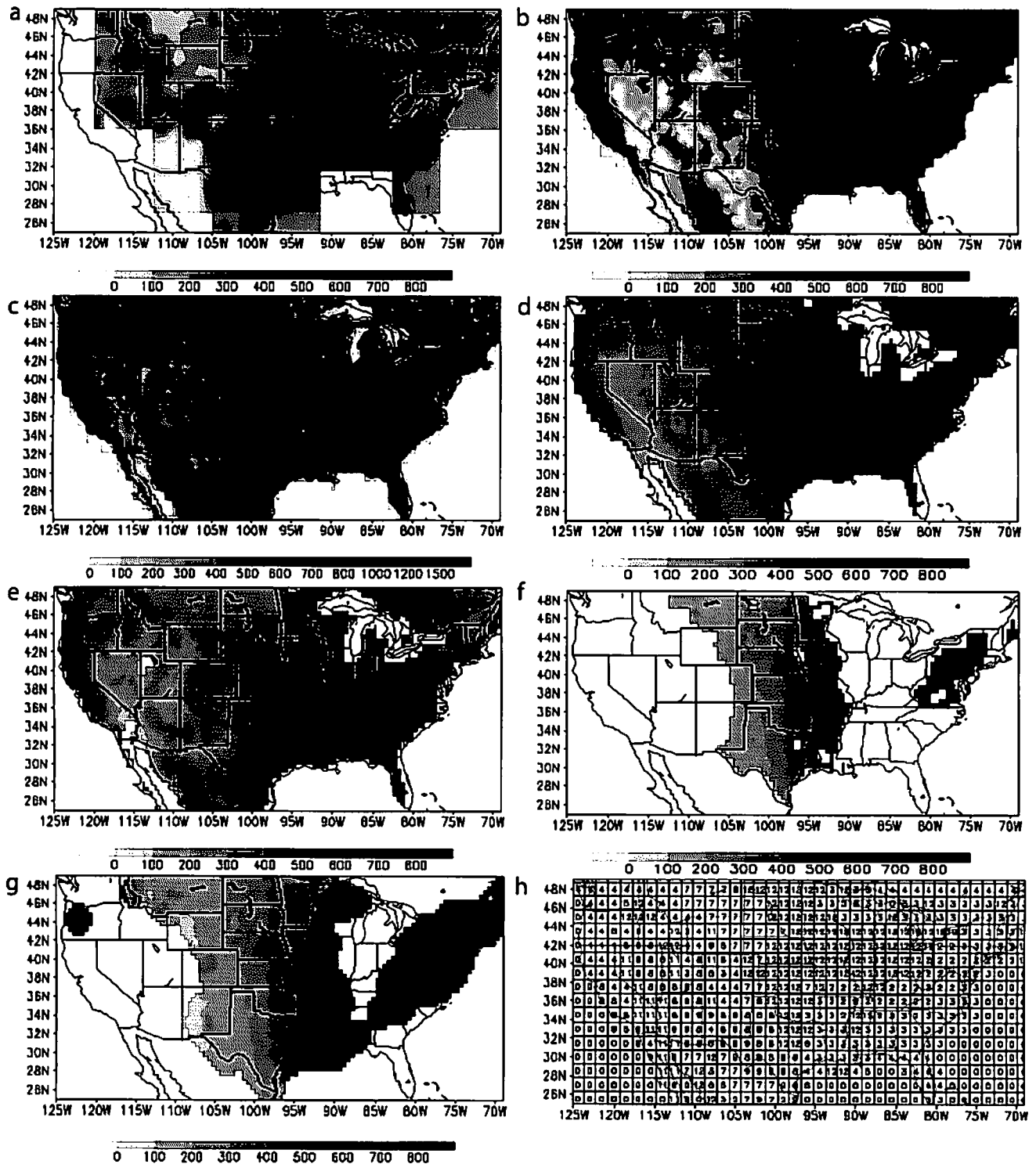
[20] Figures 4a–4g present the simulated annual mean NPP, modeled NPP (CASA and GloPEM), IGBP data, MODIS data, GPPDI data and ORNL data of USA respectively. Compared to the estimated and modeled NPP data (CASA and GloPEM), the simulated NPP reproduces the high NPP in the southeastern USA and low values in the northwest. However, the simulated NPP in the central USA south of  $40^{\circ}\text{N}$  is overestimated because of the high temperature and precipitation. The maximum discrepancy is located in northeast USA north of  $36^{\circ}\text{N}$ , where the simulation is much lower than the IGBP data, GPPDI data and ORNL data. The reason for the markedly lower magnitude up to  $400 \text{ g C m}^{-2} \text{ year}^{-1}$  in the northeastern corner of USA is complicated, since the annual temperature and precipitation of the two-way coupled models agrees generally with the CRU data. We assumed that one of the potential causes is the radiation (not analyzed in this study) affecting the photosynthetic active radiation (PAR) at the top of the canopy, because the solar radiation in the eastern USA is one potential climatic constraint to plant growth according to *Nemani et al.* [2003]. It is interesting that the MODIS data in northeastern corner (Figure 4e) are also lower than GPPDI and ORNL data, which demonstrates the complicated NPP distribution for different data sets and the necessity of validating NPP at a regional scale.

[21] The simulated and other NPP data over Australia (not including Tasmania) are shown in Figure 5. All of the NPP values show the decreasing trend from east, north and southwest to central Australia. NPP of the coupled model shows intermediate values, with magnitudes ranging between CASA and GloPEM in most regions. The simulated NPP in eastern part can reach  $800\text{--}1100 \text{ g C m}^{-2} \text{ year}^{-1}$  similar to the IGBP and MODIS data, which reflects the sufficient heat and water conditions for vegetation growth, but the simulation band is much wider than the IGBP and MODIS data with the highest NPP located along the eastern coasts. According to the document of ISLSCP II for the GPPDI data, there are possible lower estimates of belowground NPP (consequently for total NPP as well) for shrubland cells in Australian data because of application of belowground/aboveground carbon allocation method [*Raich and Nadelhoffer*, 1989] that is more suitable for forest ecosystems. We can see that the GPPDI data in central Australia are almost below  $100 \text{ g C m}^{-2} \text{ year}^{-1}$  (Figure 5f); the simulation, IGBP and MODIS data can range from 200 to  $300 \text{ g C m}^{-2} \text{ year}^{-1}$ , and this discrepancy maybe reflects the possible underestimate of the GPPDI NPP in this area.

[22] To compare the NPP magnitude at global and national levels, area-averaged NPP for all models and estimated data is presented in Table 4. Compared to the global and regional averaged NPP of IGBP and MODIS data, the simulation shows the agreement in global NPP, lower NPP



**Figure 3.** Annual mean NPP over China of (a) AVIM-GOALS, (b) CASA, (c) GloPEM, (d) IGBP Global NPP data from ISLSCP II, (e) MODIS NPP data, (f) GloPEM from *Cao et al.* [2005], and (g) ORNL estimated data. Units are  $\text{g C m}^{-2} \text{ year}^{-1}$ . See color version of this figure in the HTML.



**Figure 4.** Annual mean NPP over USA of (a) AVIM-GOALS, (b) CASA, (c) GLoPEM, (d) IGBP Global NPP data from ISLSCP II, (e) MODIS data, (f) GPPDI Gridded data from ISLSCP II, and (g) ORNL estimate data. Units are  $\text{g C m}^{-2} \text{ year}^{-1}$ . (h) Land cover classifications over USA (see Table 1 for the specification). See color version of this figure in the HTML.

magnitudes in China and USA, and higher magnitudes in Australia. The reason is the underestimated NPP of boreal forests in northeast China and northeast USA, and the wider distribution of high NPP in eastern Australia affected mainly by overestimated precipitation. The simulated NPP

lies between CASA and GLoPEM modeled values at global and regional scale. The values of *Cao et al.* [2005] over China and ORNL data over USA demonstrate the slightly high NPP of China and  $90 \text{ g C m}^{-2} \text{ year}^{-1}$  high in USA for IGBP NPP data, whereas MODIS NPP shows an approx-



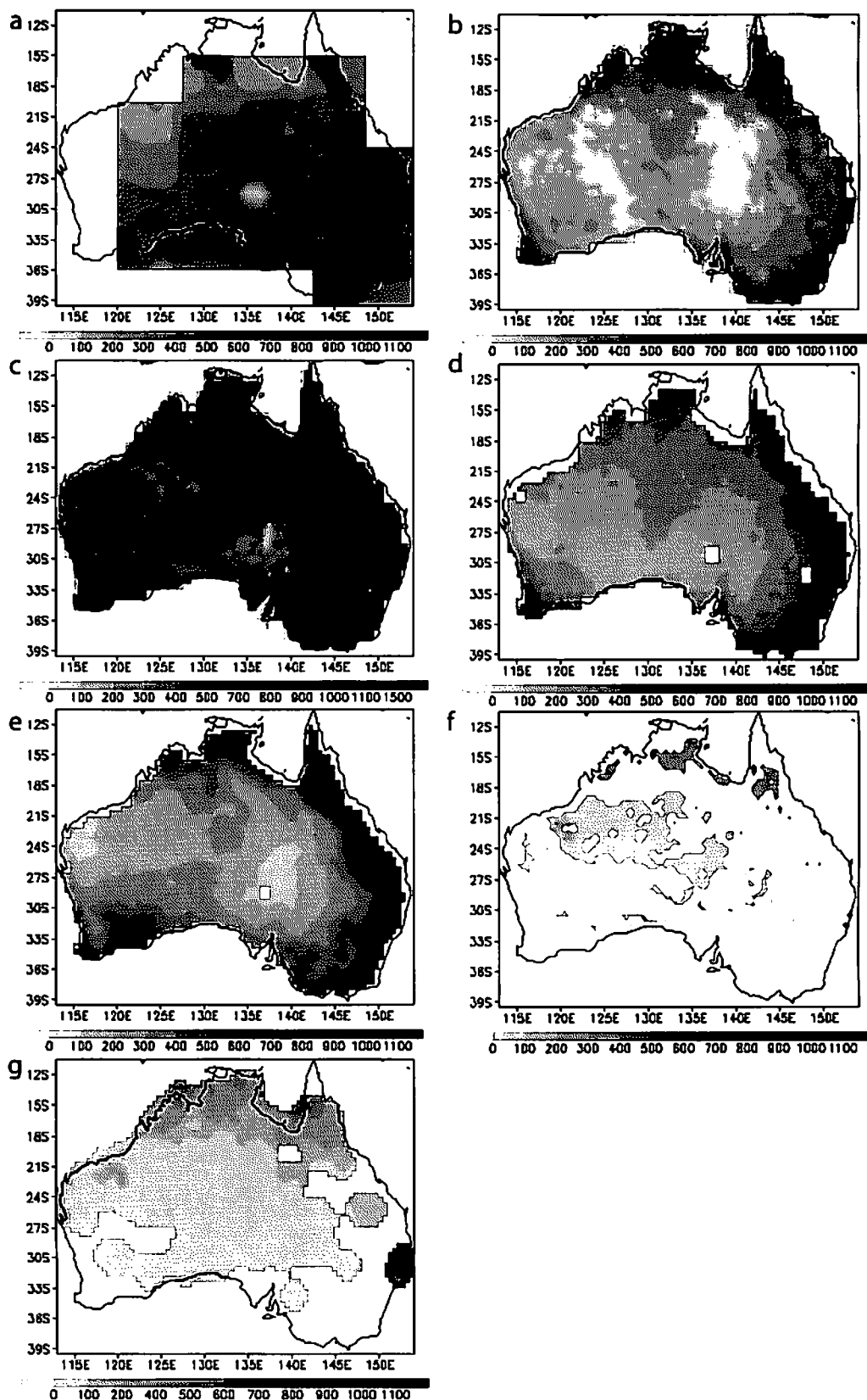


Figure 5. Legend as Figures 4a–4g but for Australia. See color version of this figure in the HTML.

**Table 4.** Area-Averaged Annual NPP at Global and Country-Level Scale<sup>a</sup>

NPP Origin	Global Average	Average Over China	Average Over USA	Average Over Australia
AVIM-GOALS	447.47	330.83	309.13	459.33
CASA NPP	313.99	331.79	387.57	251.54
GloPEM NPP	744.84	745.66	814.78	664.49
IGBP NPP data	450.42	417.58	497.73	300.26
MODIS NPP data	505.85	405.50	409.36	354.27
Other NPP		391.14 <sup>b</sup>	402.96 <sup>b</sup>	

<sup>a</sup>Units are g C m<sup>-2</sup> year<sup>-1</sup>.<sup>b</sup>NPP value of *Cao et al.* [2005] over China using GloPEM model and ORNL data over USA.

imate magnitude in the two countries. Global higher averaged MODIS NPP in comparison to IGBP data results in part from the high NPP of 5°S–30°N due to the lower vapor pressure deficit (VPD) [*Zhao et al.*, 2006]. To study the spatial agreement between the simulated NPP and other data, MODIS NPP was dealt with the referenced NPP to calculate the correlation coefficients of NPP values. The correlation coefficients have a high statistical significance of 99.9% in a student t-test and are shown in Table 5. For all NPP values, the lowest correlation occurs in the USA. For the three countries, AVIM-GOALS shows the largest correlation in China, whereas CASA and GloPEM have largest values in Australia. The correlation coefficient of AVIM-GOALS is located between CASA and GloPEM for global average. IGBP NPP shows the largest correlation at global and the country-level scale in contrast with other NPP sources. The coefficients reflect the agreement of AVIM-GOALS NPP with MODIS data lies between CASA and GloPEM for global average, and IGBP NPP agrees spatially well with MODIS data.

[23] In this study, a relative error term (RE), as adopted by *Zhao et al.* [2006], was used to evaluate the simulated NPP uncertainty at regional and global scale against IGBP data, and RE equation is:

$$RE = \frac{\sum_{i=1}^n \frac{P_i - P_r}{P_r}}{n} \times 100\% \quad (2)$$

where  $P_i$  is global and regional NPP of models,  $P_r$  is IGBP data,  $n$  is the number of grid cells. Table 6 presents the RE values of AVIM-GOALS, CASA and GloPEM at global and regional scale. A significant discrepancy exists for the three models: AVIM-GOALS has positive RE values of globe and Australia and negative in China and USA, whereas CASA is negative and GloPEM is positive at regional and global scales. The simulated RE values are largest in Australia (>60%) and relatively small in other regions. The global RE value is a quantitative measure of the deviation between AVIM-GOALS and IGBP NPP. This is comparable to CASA and GloPEM and the degree of departure from global IGBP NPP in the coupled model is located between the two models.

**Table 5.** Correlation Coefficients Between NPP Values and MODIS NPP Data in the Global and Country-Level Spatial Pattern

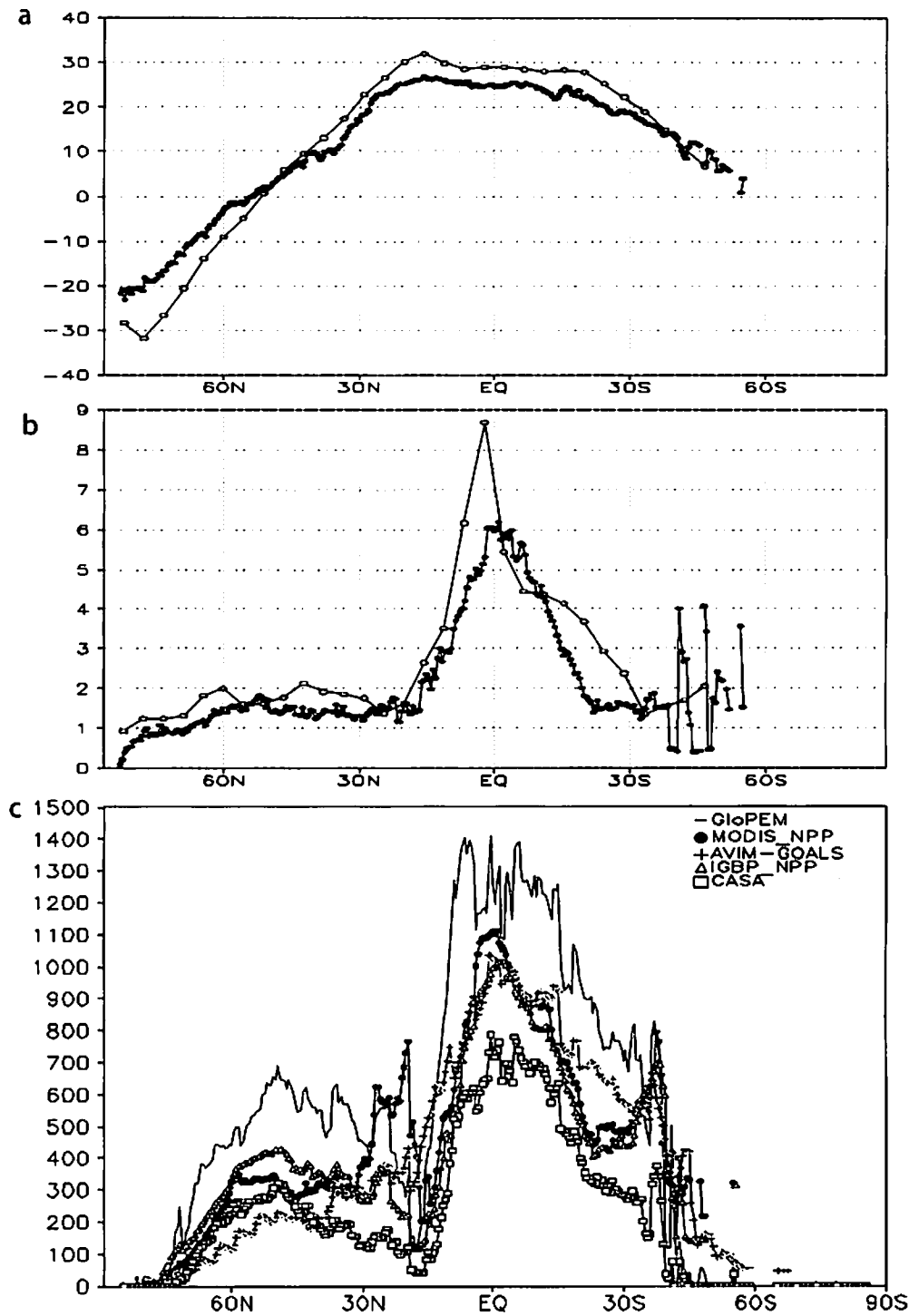
NPP Origin	Globe	China	USA	Australia
AVIM-GOALS	0.77	0.76	0.57	0.63
CASA NPP	0.82	0.64	0.64	0.80
GloPEM NPP	0.72	0.65	0.45	0.70
IGBP NPP data	0.86	0.81	0.75	0.90

### 3.2. Zonal Average and Latitudinal-Seasonal Variation of Global NPP

[24] Figure 6 is the globally zonal average for surface air temperature, precipitation and NPP distribution. All zonal mean NPP values (Figure 6c) show the similar latitudinal variation pattern despite the different magnitude. The simulated NPP is close to IGBP and MODIS data except for north of 45°N, and the magnitude varies between CASA and GloPEM in 45°N–50°S. The zonal mean temperature over land south of 50°N is above 0°C, which is reflected in the simulation and CRU temperature. The climatic factors of temperature, precipitation and radiation (not analyzed here) tend to be colimiting the plant growth [*Nemani et al.*, 2003]. This study only considers precipitation and temperature at regional and global scales, which are the two major climatic factors that govern the NPP of the biosphere [*Cramer et al.*, 1999; *Tian et al.*, 1999]. In boreal regions, the vegetation is mainly controlled by air temperature; in temperate regions, the vegetation growth is affected by temperature and precipitation; in subtropical and tropical regions, the precipitation is the principal limiting factor for the vegetation growth because of the sufficient surface heat content. Consequently, the NPP discrepancy in Figure 6c is related to the climate difference (Figures 6a and 6b). The simulated zonal mean NPP north of 50°N is lower than the IGBP and MODIS data, which reflects the lower temperature in this region. For the NPP distribution in the Arctic Circle, the ecosystem is mainly tundra and the absolute value of NPP is rather low, thus the magnitude of the NPP discrepancy is not large despite the lower temperature up to 10°C. However, the question remains as to why the 30–50°N zonal mean NPP is obviously smaller than the IGBP and MODIS data since the simulated temperature is 1–2° higher than the CRU data and the simulated precipitation is slightly overestimated. Through careful analysis, we found the cause originating over the Tibetan Plateau, where the simulated strong warming bias more than 8°C (Figure 1e) leads to the slight overestimate of zonal mean temperature between 30 and 50°N, with the explanation being the thin snow cover [*Dan et al.*, 2005]. However, the obvious warming in the Tibetan Plateau cannot cause the corresponding NPP increase in full, because the land cover

**Table 6.** Relative Error Term (RE) of the Three Global Models With Respect to IGBP NPP Data

Global/Regional RE	AVIM-GOALS, %	CASA, %	GloPEM, %
Globe	20.01	–16.47	106.27
China	–20.71	–18.37	96.09
USA	–31.96	–16.78	107.32
Australia	64.41	–10.58	161.36



**Figure 6.** Zonal average of the annual means: (a) surface air temperature, with units of  $^{\circ}\text{C}$ ; (b) precipitation, with units of  $\text{mm/day}$ , open circle indicates outputs of AVIM-GOALS, and solid circle indicates CRU climate; and (c) NPP of three global models, IGBP NPP data and MODIS data, with units of  $\text{g C m}^{-2} \text{ year}^{-1}$ .

is bare soil or sparse vegetation in arid regions. So the underestimated temperature in other regions along this latitude results in the well-defined decrease in NPP (e.g., northeast China for conifer forests, see Figure 3a). For the zonal mean NPP between  $30^{\circ}\text{N}$ – $30^{\circ}\text{S}$ , the NPP variation is mainly controlled by precipitation change, and the overestimated NPP simulation (in contrast with IGBP and

MODIS data) generally corresponds to the increasing precipitation magnitude compared to CRU data (Figure 6b). The  $5^{\circ}\text{S}$ – $30^{\circ}\text{N}$  MODIS NPP is overestimated because of the lower vapor pressure deficit [Zhao *et al.*, 2006], thus it presents NPP values which are too high between  $20^{\circ}$  and  $30^{\circ}\text{N}$ . By comparing the overall trend of zonal mean curves of NPP with that of air temperature and precipitation, we

Table 7. Zonal Mean of Annual NPP for Latitudinal Bands With 30° Interval<sup>a</sup>

NPP Origin	60–30°S Average	30°S–0° Average	0–30°N Average	30–60°N Average	60–90°N Average
AVIM-GOALS	470.76	791.89	493.18	216.83	70.92
CASA NPP	311.71	564.86	318.93	253.68	146.73
GloPEM NPP	848.71	1182.55	846.09	626.59	387.35
IGBP NPP data	453.58	672.40	451.53	380.43	214.78
MODIS NPP data	506.32	737.38	678.61	325.78	180.13

<sup>a</sup>Units are g C m<sup>-2</sup> year<sup>-1</sup>.

found that the NPP variation across latitudes is more similar to precipitation change than it is to temperature.

[25] The zonal mean NPP in the latitudinal band is generally consistent with IGBP and MODIS data (Table 7), however, the values above 60°N is markedly underestimated for the same reason for China and USA. The largest NPP distribution was found between 30°S–30°N, whereas the NPP of 0–30°N is smaller than that of 30°S–0 because of more arid regions or deserts in the 0–30°N belt. The simulated value is located between CASA and GloPEM excluding the northern high latitude. This suggests again that the simulated NPP magnitude varies generally between the two global biogeochemical models and the latitudinal NPP variation of the coupled model is reasonable in comparison to IGBP and MODIS data. The high NPP of MODIS data in the regions 30°S–0 and 0–30°N has an identical explanation to that of the 5°S–30°N.

[26] Figure 7 exhibits the month-to-month change of climate and simulated NPP across latitudes. The latitudinal-seasonal change of simulated air temperature agrees well with the CRU temperature, with a discrepancy of 5° overestimate between 30°S and 30°N. During June to August in each semihemisphere, a temperature ridge exists representing a cold peak in the Southern Hemisphere and a warm peak in the Northern Hemisphere, especially poleward of 30°S or 30°N. The simulated and CRU precipitation shows the latitudinal-seasonal pattern similar to the temperature and the main rain belt is located in the tropical regions. The simulated monthly NPP with the magnitude above 60 g C m<sup>-2</sup> is mainly located in the 10°S–5°N tropical belt, which corresponds to the largest air temperature and precipitation in this latitudinal zone. The NPP in northern latitudes above 30°N shows the maximum value during boreal summer related to the climate ridge especially to temperature. The region poleward of 30°S has the lowermost NPP during June to August, but it is not as obvious as the peak of the Northern Hemisphere because of the fact that most landmasses in the Southern Hemisphere are mainly distributed within the relatively warm conditions of 35°S. The NPP in the tropical latitudes exhibits no peaks or conspicuous low values, which can be attributed to the fact that the different seasonal patterns (but of similar magnitudes) of NPP in Northern and Southern Hemisphere oppose each other, acting as buffers. Also, the latitudinal-seasonal variation of globally monthly NPP should be correct when compared with the results of the 15 models participating in the Potsdam NPP model Intercomparison Workshop [Kicklighter et al., 1999].

### 3.3. Global NPP and Climate for Vegetation Type

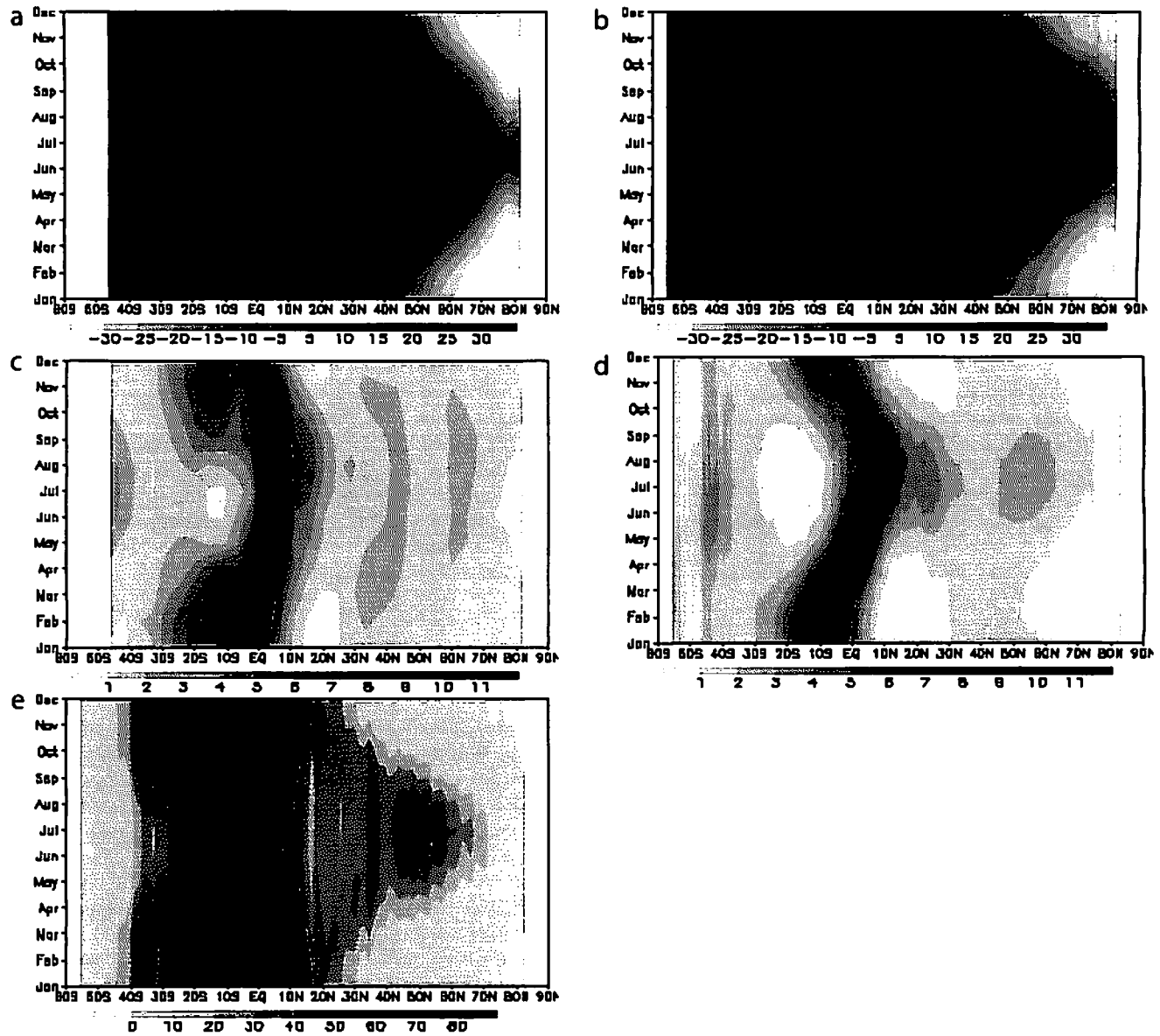
[27] It is rather difficult to obtain global field data of NPP to validate global or regional models, but use of the

averaged NPP by biome type is a valuable method to evaluate the terrestrial NPP [Zhao et al., 2005]. The intercomparison of NPP by vegetation types has been used in many previous studies [Scurlock and Olson, 2002; Dan et al., 2005; Zhao et al., 2005] and it proves to be a useful validation method for global NPP distribution. Global NPP averaged by biome types is shown in Figure 8a for 11 vegetation types excluding bare soil or desert, and the vegetation classification can be found in Table 1. Similar to NPP, the simulated and observed climate (temperature and precipitation) for the 11 vegetation types is also presented in Figures 8b and 8c.

[28] NPP of all sources show high values in tropical rain forest (vegetation 1), broadleaf trees with ground cover (vegetation 6) and crops (vegetation 12), which can be attributed to the favorable climate conditions for vegetation growth in the two former types and the human effects (e.g., irrigation) in the latter. The lower NPP occurs in arid or cold regions (vegetation 9 and 10) with unfavorable growth conditions of climate. The simulated NPP of AVIM-GOALS in boreal forest (types 3–5) and tundra (type 10) is lower than IGBP and MODIS NPP because of the underestimated temperature, especially for vegetation type 5, and the simulated NPP of other vegetation types lies generally between CASA and GloPEM. The averaged NPP for all NPP sources (Figure 8a) is highly correlated to mean climate of precipitation (Figure 8c) with a correlation coefficient of 0.96, which is higher in magnitude and statistical significance than those of temperature (0.59). This shows that the NPP variation across global vegetation types depends highly on precipitation and temperature is generally the second climate constraint. The correlation coefficients between the simulated and observed climate (0.98 for temperature and 0.93 for precipitation) reflect the good agreement of actual climate data to simulated climate data across terrestrial ecosystem types excluding bare soil. All the coefficients are conducted in a student t-test and the significance level is shown in the Figure 8 legend.

[29] To evaluate further the uncertainty of NPP for biome types in the variation amplitude, the maximum, minimum and range (Max-Min-Ran) of NPP are presented in Figure 9. The NPP range of AVIM-GOALS is higher than IGBP and MODIS data, and the range of MODIS NPP corresponds closely to that of mean NPP averaged in all NPP sources (Figure 9a). The maximum and minimum simulations are located in the intermediate position for the three global models. The Max-Min-Ran of each NPP source minus that of mean NPP is shown in Figure 9b. The simulated range difference of AVIM-GOALS can be more than 100 g C m<sup>-2</sup> year<sup>-1</sup>, and the Max-Min difference lies again between CASA and GloPEM. The comparison reflects that the smallest discrepancy with respect to averaged NPP occurs





**Figure 7.** Month-to-month variation for globally zonal mean of annual NPP and climate: (a) surface air temperature ( $T_s$ ) of AVIM-GOALS; (b) CRU  $T_s$  data, with units of  $^{\circ}\text{C}$ ; (c) precipitation of AVIM-GOALS; (d) CRU precipitation data, with units of  $\text{mm/day}$ ; and (e) NPP of AVIM-GOALS, with units of  $\text{g C m}^{-2} \text{ month}^{-1}$ . See color version of this figure in the HTML.

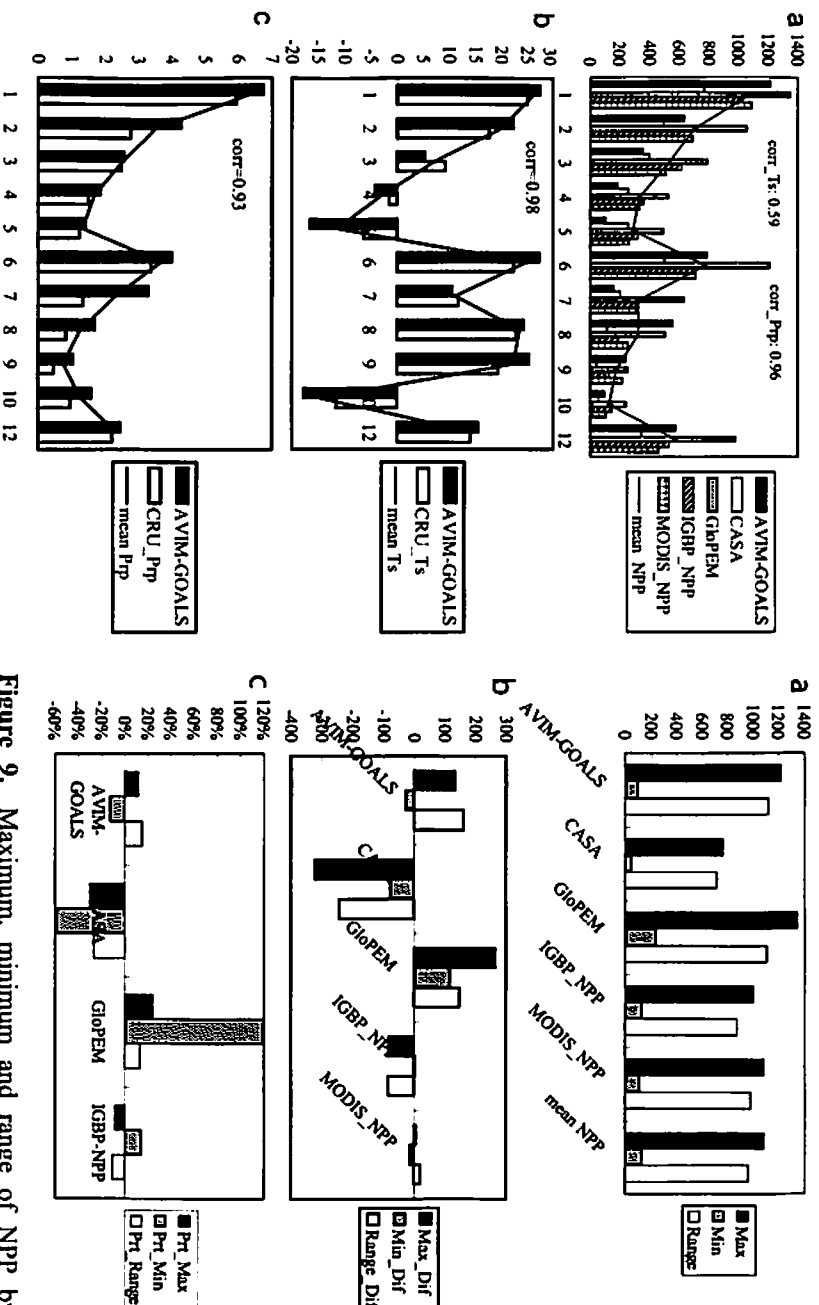
in MODIS NPP, and IGBP NPP is second to MODIS data. Thus the Max-Min-Ran relative percentage (RP) with respect to MODIS data is calculated as:

$$RP = \frac{P_i - P_m}{P_m} \times 100\% \quad (3)$$

where  $P_i$  is the Max-Min-Ran of three global models and IGBP NPP,  $P_m$  is the Max-Min-Ran of MODIS NPP data. The RP values of AIVM-GOALS (Figure 9c) lie within  $\pm 20\%$  and resemble those of IGBP NPP data, which reflects the relative agreement with IGBP and MODIS data. The absolute RP value of GloPEM range is the smallest in the three global models and agrees well with IGBP and MODIS data. The Max-Min RP value of AVIM-GOALS is located between CASA and GloPEM. The absolute RP values of

IGBP data are lower than 15%, which shows the closest agreement with MODIS data.

[30] Finally, the correlation coefficient (Corr) between NPP and the observation climate is calculated to explore the relationship of global NPP with temperature and precipitation for vegetation types. All the coefficients and their significance level in a student t-test are presented in Table 8. The Corr of all NPP with precipitation is statistical high compared to that of temperature. The Corr of temperature for the simulated NPP is highest in all NPP sources, and the Corr of precipitation is smallest compared to that of other NPP sources. This reflects that the NPP variation of global vegetation types in the coupled model is more related to temperature than other NPP data, which helps to understand the lower simulated NPP at northern high latitudes due to



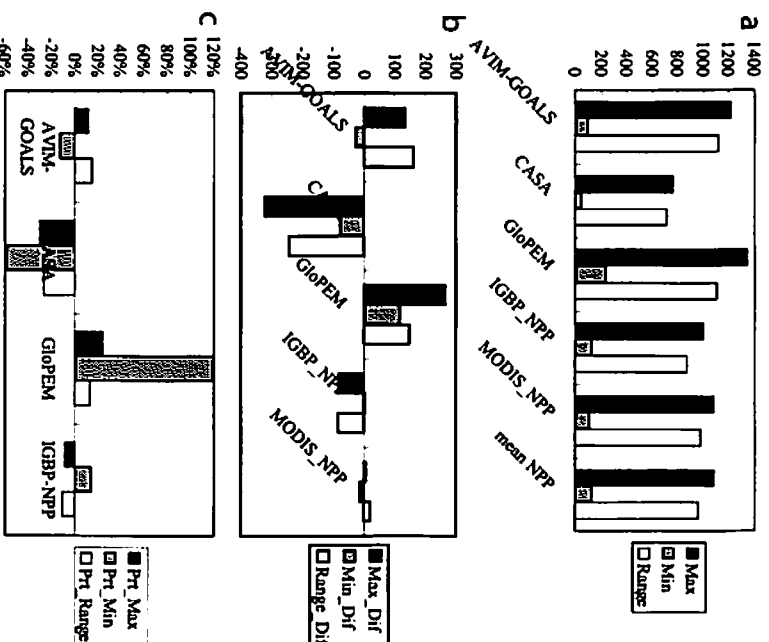
**Figure 8.** Annual mean NPP and climate based on global vegetation type: (a) NPP, with units of  $\text{g C m}^{-2} \text{ year}^{-1}$ ; (b) surface air temperature, with units of  $^{\circ}\text{C}$ ; and (c) precipitation, with units of  $\text{mm/day}$ . (corr, correlation coefficient between the simulated climate and CRU data at 99.9% significance level; corr\_Ts, correlation coefficient of mean NPP with mean temperature at 90% significance level; corr\_Prp, correlation coefficient of mean NPP with mean precipitation at 99.9% significance level.)

the underestimated temperature. The Corr of MODIS data shows the highest correlation with temperature and precipitation and equals nearly to the Corr of mean NPP, which exhibits the high dependence on temperature and precipitation across global biome types. IGBP Corr shows stronger correlation with precipitation than temperature, as opposed to MODIS data.

[31] For all NPP sources, vegetation type 8 (broadleaf shrub with ground cover) is the main type leading to the lower Corr with climate. To show this effect, the Corr removing type 8 is shown in Table 9. The simulated Corr for precipitation increases to 0.95, and the Corr of other NPP data increases markedly for temperature, especially in CASA model. This reflects the different mechanism of vegetation type 8 related to climate: controlled by precipitation and not simulated well in the coupled model; controlled mainly by temperature and possibly not captured well in other two global models, IGBP and MODIS data.

#### 4. Conclusions and Discussion

[32] The global model NPP, ranging from 39.9 to 80.5  $\text{Pg C year}^{-1}$  [Cramer *et al.*, 1999], indicates a large uncertainty



**Figure 9.** Maximum, minimum and range of NPP by global vegetation type for all data sets: (a) NPP and their average; (b) the difference of NPP minus the mean values in Figure 9a, with units of  $\text{g C m}^{-2} \text{ year}^{-1}$ ; and (c) relative percentage of NPP with respect to MODIS data in the equation of  $(P_i - P_m)/P_m$ , multiplying by 100%,  $P_m$  is NPP from MODIS data. (Max, maximum; Min, Minimum; Dif, difference; Prt, percentage.)

[Mhl *et al.*, 2005]. Consistent data on terrestrial NPP are urgently needed to constrain model estimates of carbon fluxes and hence to refine our understanding of ecosystem responses to climate change [Scurlock and Olson, 2002]. However, because of the long-term deficiency of global data [Zheng *et al.*, 2003; Zhao *et al.*, 2005], the model NPP validation is not implemented fully and maturely at global and regional scales as are the validations of climatic simulations. Many discrepancies exist for global NPP estimates, such as in the tropical rain forests, the maximum of  $850 \text{ g C m}^{-2} \text{ year}^{-1}$  for Field *et al.* [1998],  $964 \text{ g C m}^{-2} \text{ year}^{-1}$  for Cao and Woodward [1998],  $1500 \text{ g C m}^{-2}$

**Table 8.** Correlation Coefficients Between All NPP and CRU Climate by Global Vegetation Type<sup>a</sup>

NPP Origin	Corr Ts and the Significance Level	Corr Prp and the Significance Level
AVIM-GOALS	0.76 (99%)	0.88 (99.9%)
CASA_NPP	0.46 (80%)	0.97 (99.9%)
GloPEM_NPP	0.60 (99%)	0.90 (99.9%)
IGBP_NPP data	0.47 (80%)	0.97 (99.9%)
MODIS_NPP data	0.60 (95%)	0.97 (99.9%)
Mean NPP	0.61 (95%)	0.96 (99.9%)

<sup>a</sup>Ts, air temperature; Prp, precipitation. The significance level in the parentheses is obtained in a student t-test.

**Table 9.** Correlation Coefficients Between All NPP and Climate as Table 8 but for Excluding Vegetation Type 8 of Broadleaf Shrub With Ground Cover

NPP Origin	Corr_Ts and the Significance Level	Corr_Prp and the Significance Level
AVIM-GOALS	0.77 (99%)	0.95 (99.9%)
CASA NPP	0.61 (90%)	0.96 (99.9%)
GloPEM NPP	0.71 (98%)	0.90 (99.9%)
IGBP NPP data	0.63 (95%)	0.97 (99.9%)
MODIS NPP data	0.73 (98%)	0.97 (99.9%)
Mean NPP	0.71 (99%)	0.97 (99.9%)

year<sup>-1</sup> for Goetz *et al.* [2000], 1672 g C m<sup>-2</sup> year<sup>-1</sup> for Bonan *et al.* [2003] and 1000 g C m<sup>-2</sup> year<sup>-1</sup> for Zeng [2003]. There is a need for improved understanding of the factors that influence the variability of NPP model estimates at different scales so the accuracy of NPP estimated at the global scale can also be improved [Ahl *et al.*, 2005]. Consequently, the model validation for NPP estimates and the exploration of the NPP variation related to climate change is very important.

[33] In this study, we used the modeled and estimated NPP (including IGBP and GPPDI) data to compare and validate the simulated global NPP with the two-way coupled model between biosphere and atmosphere. By comparing the spatial pattern of NPP at global and the country-level scale (China, USA and Australia), we found the general consistency of the simulation with IGBP and MODIS data. The global NPP distributions of the model simulations from low, mid to high latitudes were reasonable, compared to the spatial pattern of previous studies [Field *et al.*, 1998; Kicklighter *et al.*, 1999; Running *et al.*, 1999; Bonan *et al.*, 2003; Zeng, 2003]. The underestimated air temperature limiting the vegetation growth in the northern high latitudes leads to the lower simulated NPP. The simulated NPP south of 15°S in Africa is larger than the IGBP and MODIS data and it results from the classification of different ecosystem types; that is, the land cover type is mainly broadleaf trees as the SSiB model [Xue *et al.*, 2004], whereas the shrublands or grasslands are distributed in many other models. The zonal mean NPP values of the simulation in each latitudinal band with 30° interval show the consistency with IGBP and MODIS data except the obviously lower estimate poleward of 60°N.

[34] The country-level NPP was compared and validated using IGBP data, GPPDI data, two global model NPP and ORNL estimate data with the global 0.5° cell. In China, the simulated NPP corresponds well to the IGBP and MODIS data in the spatial pattern with NPP >500 g C m<sup>-2</sup> year<sup>-1</sup> south of 35°N in eastern China and up to 700 g C m<sup>-2</sup> year<sup>-1</sup> more in southern China, which is reproduced in the new calculation using GloPEM [Cao *et al.*, 2005]. However, in northeast China, the simulated NPP is underestimated (caused by a cold bias) even reaching lower 200 g C m<sup>-2</sup> year<sup>-1</sup> north of 48°N compared to IGBP and other modeled and estimated data. By carefully comparing the simulation, IGBP data and GPPDI data north of 48°N in northeast China to previous relevant studies [e.g., Tao *et al.*, 2003; Jiang *et al.*, 1999], we found that the actual NPP in this region may be 350–400 g C m<sup>-2</sup> year<sup>-1</sup>. The NPP value of Cao *et al.* [2005], MODIS and ORNL data also supports

those values in this area. In the USA, the simulated NPP shows roughly an overall agreement with the IGBP and GPPDI data, but the simulation cannot capture the west-east gradients between 95 and 105°W because the simulated precipitation cannot capture the 1–2.5 mm/day gradient in this area, unlike with CRU data. As a result, along with the north-south gradient impact of air temperature, the west-east NPP band is not shown in the simulation. In Australia, the simulated NPP corresponds to the modeled and estimated data with the large values in the eastern, northern and southwestern belts. The simulated NPP, IGBP and MODIS data in central Australia (200–300 g C m<sup>-2</sup> year<sup>-1</sup>) reflect the possible underestimate of NPP in GPPDI data (<100 g C m<sup>-2</sup> year<sup>-1</sup>) depicted in the document from ISLSCP II, but whether or not the NPP value of 200–300 g C m<sup>-2</sup> year<sup>-1</sup> is more actual is a decision not to be taken lightly, considering the NPP distribution modeled by Wang and Barrett [2003], which shows less than 200 g C m<sup>-2</sup> year<sup>-1</sup> in central Australia.

[35] Every single model simulates the NPP with the discrepancies from other models' result and limited field measurement, which demonstrates again the necessity of model NPP validation against field data at regional and global scales. As for the current data availability, besides field or ground-based measured data, the multimodel ensemble result is another effective and reliable choice for model validation because the ensemble result (e.g., IGBP NPP data in this study) produces better estimates of the real value of NPP than any one particular model taken individually; this is similar to the ensemble results of climate simulations (chapter 9 of the third IPCC report) and surface hydrological simulations [Gordon *et al.*, 2004]. The relative error term (RE) is one quantified criteria to evaluate the model NPP discrepancy against the ensemble result or field data. Compared to IGBP NPP data, the RE values of the coupled model lie around ±20% across the globe and over China, respectively, and more than –30% in the USA and 60% in Australia. This reflects that the best agreement with IGBP data is over the entire globe and China, and the large deviation occurs in the USA and Australia. However, the overall RE values can be compared to those of CASA and GloPEM and are located in the intermediate position at global scale. The area-averaged NPP and correlation coefficients with MODIS data at global and regional scales again reveal that the values of AVIM-GOALS lie generally between the two global models. All the values above show the reasonable overall simulation capacity of the coupled model to capture spatially global and regional NPP in contrast with modeled and estimated NPP.

[36] The response of ecosystem processes at regional and global scales to changing climate is the central theme in the analysis of global change [Tian *et al.*, 1999]. We also compared the globally zonal mean NPP using IGBP data, MODIS data and NPP of two global models and studied the relationship between NPP and climate. The simulated NPP is consistent overall with the NPP data and is lower than IGBP and MODIS data north of 30°N, which is due to the underestimated air temperature (excluding the 30–50°N warm bias of the Tibetan Plateau). In view of the overall latitudinal variation, the global terrestrial NPP corresponds more closely to the precipitation than the air temperature for all NPP data. The month-to-month variation of NPP across

latitudes reflects the seasonal variation of temperature and precipitation, and the seasonal variation in Northern Hemisphere is more obvious than that of the Southern Hemisphere. The global NPP of all data for vegetation types is more strongly correlated to precipitation (higher correlation coefficients at 99.9% significance level) than to temperature, which shows that the climatic effect of moisture is generally the principal limiting factor for global vegetation types. The work of *Lu and Ji* [2006] using AVIM in Northeast China Transect (NECT) presents also the high dependence of moisture including precipitation for spatial distribution of NPP.

[37] To quantify the global and regional NPP is a challenge, however, though it is one useful and reliable way to validate the NPP values with the synthesized NPP as suggested by *Zhao et al.* [2005]. The globally averaged NPP of AVIM-GOALS is  $447.47 \text{ g C m}^{-2} \text{ year}^{-1}$  and agrees with the  $450.42 \text{ g C m}^{-2} \text{ year}^{-1}$  of IGBP data, but the regional mean NPP is underestimated over China and USA and overestimated over Australia. Global total NPP is meaningful for the global carbon cycle because carbon sinks result from an increase in global terrestrial NPP [*Nemani et al.*, 2003; *Zhao et al.*, 2005]. The simulated global terrestrial total NPP of  $53.4 \text{ Pg C year}^{-1}$  is close to the  $54.9 \text{ Pg C year}^{-1}$  of IGBP data and it approximates the value of other literature [e.g., *Cao and Woodward*, 1998; *Nemani et al.*, 2003]. The simulated and IGBP data averaged over China are  $330.83/417.58 \text{ g C m}^{-2} \text{ year}^{-1}$ , whereas other studies [*Xiao et al.*, 1998; *Sun and Zhu*, 2001; *Cao et al.*, 2003, 2005; *He et al.*, 2005] present different values below  $400 \text{ g C m}^{-2} \text{ year}^{-1}$ . Thus the IGBP value appears to be overestimated to an extent, partly because of the high NPP in northeast China, and the mean value of China below  $400 \text{ g C m}^{-2} \text{ year}^{-1}$  may be closer to the truly correct result. Compared to the regional estimate of forest NPP from Resource Planning Act/Forest Inventory and Assessment (RPA/FIA) and CASA model driven with satellite data [*Hicke et al.*, 2002], the NPP of IGBP data averaged over USA seems reasonable with the value of  $497.73 \text{ g C m}^{-2} \text{ year}^{-1}$ . However, compared to the longer time (1900–1993) mean NPP from the supplementary information of *Nemani et al.* [2002], the IGBP data of USA appear to be about  $90 \text{ g C m}^{-2} \text{ year}^{-1}$  high. The ORNL NPP averaged over the USA reveals the higher IGBP data as well. The simulated NPP over the USA in the coupled model AVIM-GOALS is underestimated because of the lower NPP north of  $42^\circ\text{N}$ . For the NPP averaged over Australia, it is difficult for us to draw a definite conclusion because the national-level NPP varies considerably, similar to that in the work of *Roxburgh et al.* [2004] (who presented twelve model estimates of long-term annual total NPP ranging from  $0.67$  to  $3.31 \text{ Pg C year}^{-1}$ ). We make a roughly hypothetical calculation with the ratio of the areas of Australia to global land areas (0.0502) and using the method of *Raupach and Moran* [1998] we found that the global terrestrial total NPP for the simulation and IGBP data multiplied by 0.0502 is  $2.68$  and  $2.76 \text{ Pg C year}^{-1}$ , and the two total NPP values of the Australian continent seem to be within the range presented by *Roxburgh et al.* [2004]. On the basis of the analysis above, we can conclude that the global and national-level mean NPP should be validated further in future with more field data as suggested by *Cao et al.*

[2003]. Also, the regional mean NPP shows more uncertainty than the global average.

[38] Generally speaking, the simulated NPP is accurate in the global distribution. However, at a national scale, it shows the regional bias due to the different impacts of climate, such as the underestimate in northeast China caused by cold bias and not reproducing the  $95\text{--}105^\circ\text{W}$  west-east gradient in the USA from the spatial weakness of precipitation. Correlation and relative error term (RE) were used to evaluate the simulated NPP at global and regional scale and show the similar spatial strength and weakness of simulation. In contrast with the regional correlation with MODIS data, the coupled model simulates NPP better over China than USA and Australia. This can also be reflected in the RE values of AVIM-GOALS with respect to IGBP data.

[39] IGBP and GPPDI data show the good spatial pattern for the distribution of NPP, even at the national scale. Compared to previous literature and estimated data, IGBP data capture the reasonable NPP magnitude from the regional scale in China, USA and Australia, up to global estimate. Its highest correlation coefficient with MODIS data (0.75 of USA to 0.90 of Australia) in all modeled NPP shows the best reasonable global and regional spatial structure in contrast with the three global models. The relative percentage of maximum, minimum and range for NPP across global vegetation types with respect to MODIS data is between  $\pm 15\%$ , reflecting the best and smallest variation amplitude in all modeled NPP values. However, the deficiency appears to be the slight overestimate of NPP in northeast China. Compared to data of *Cao et al.* [2005], previous literature, MODIS data and ORNL data, IGBP NPP north of  $48^\circ\text{N}$  in northeast China seems to be at least a  $50 \text{ g C m}^{-2} \text{ year}^{-1}$  overestimate. For IGBP NPP across global vegetation type, vegetation type 8 affects the correlation with CRU temperature markedly and lowers the correlation coefficient significantly; however the explanation for this effect is complicated and will not be discussed here. GPPDI data obtained from field data exhibit finer distribution at regional scales such as in the USA. The field data can guarantee the high quality of the GPPDI data from the perspective of data source, but the carbon allocation method in central Australia leads to the possible underestimate of approximately  $100 \text{ g C m}^{-2} \text{ year}^{-1}$  as described by the ISLSCP II document, and the lower NPP may be true in contrast to that of MODIS data, IGBP data and CASA model.

[40] Admittedly, MODIS data are also not the global observation NPP. However, the NPP estimate from satellite data is potentially accurate for global NPP distribution [*Zhao et al.*, 2006] and can provide good quality data for the validation of model NPP. Thus we can use MODIS NPP as a reference to compare the spatial distribution between all modeled NPP and IGBP data, especially the relationship for global vegetation types. MODIS NPP used in this paper is averaged during 2000–2003, and strictly speaking, the 4-year mean has inconsistency to a degree compared with the multiyear mean NPP of other global models. However, for current global NPP data available, use of MODIS data is a good choice to validate the NPP simulation from global climate-vegetation models, especially in global and regional spatial distributions. This comparison method with insufficient duration is similar to the work of *Zhao et al.* [2006]



using 2- or 3-year mean MODIS NPP from the year of 2000 to compare against EMDI NPP data. On the other hand, considering the increase of NPP in northeast China since the 1980s [Cao *et al.*, 2003, 2005], MODIS data in this region should be higher than NPP in the 1980s. However, MODIS NPP in northeast China is lower than IGBP data and this discrepancy again supports the overestimated IGBP NPP in this region. Global models of GloPEM and CASA with different moisture schemes for NPP controlling may estimate high and low global NPP, respectively [Cramer *et al.*, 1999]. The discussion of the discrepancy in the two models is out of the scope of this study. In other words, the models were just used in a comparison for global and regional NPP simulation in spatial pattern to show the model performance of AVIM-GOALS in contrast with other representative global biogeochemical models.

[41] The coarse resolution of GOALS GCM due to time-consuming calculation exerts some limit on the NPP simulation resulting from the land-sea border control (e.g., in the western coast areas of USA and Australia), but it does not influence much of the study as presented in this paper at global and regional scales. The higher resolution of GCM should be coupled to simulate terrestrial NPP distribution and variation at global/regional scales. Fortunately, the R42 L9 GOALS GCM with the grid approximately 2.815° longitude by 1.67° latitude has been developed and is in progress [Wu *et al.*, 2003], and we think the finer resolution of a two-way coupled model in the future will improve the simulated NPP distribution at the regional scale to an extent, such as the closer capture of spatial heterogeneity in contrast to other model and estimated data, with the finer resolution of AVIM (better for 0.5° by 0.5°).

[42] In fact, the actual global observation of NPP distribution data is still unavailable at present. For instance, the GPPDI data derived from field measured NPP can provide the data for model validation only in some regions such as the three countries in this study (scattered, sparse data in China), and there are no records with 0.5° in many regions of the globe. As a result, the NPP data, which incompletely covers the surface of the globe, causes some limitations on our model validation, and hence the validation should be carried out further with more field data aggregated to the appropriate spatial scale suitable for model validation in the future.

[43] **Acknowledgments.** This work was supported by National Basic Research Program of China "the Development and Application of a Regional Model of the Earth System" (grant 2006CB400506), the Innovation Project of Chinese Academy of Sciences (grant KZCX3-SW-229), and the project of National Natural Science Foundation of China (grant 40205013). J. Ji was funded by NKBRF under grant 2002CB412503. We thank ISLSCP II for providing the CRU data, IGBP data and GPPDI data. We also acknowledge Christopher Michael Brodowski and Liza P. Koshy for the English editing of this paper.

## References

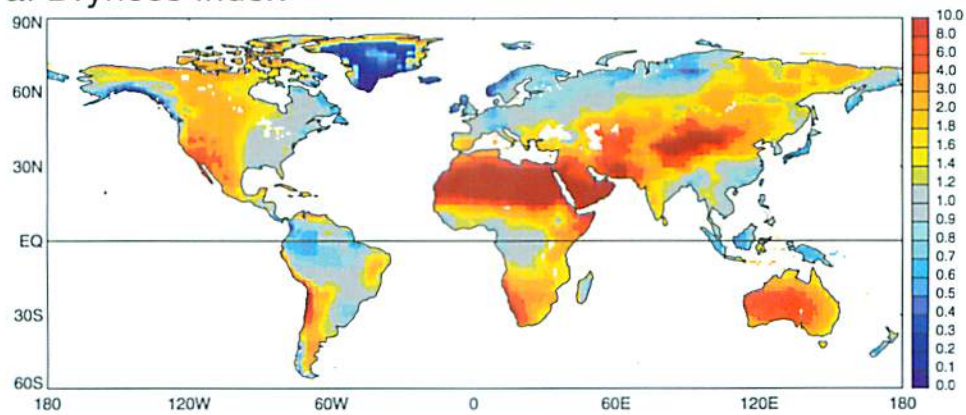
- Ahl, D. T., S. T. Gower, D. S. Mackay, S. N. Burrows, J. M. Norman, and G. R. Diak (2005), The effects of aggregated land cover data on estimating NPP in northern Wisconsin, *Remote Sens. Environ.*, **97**, 1–14.
- Bonan, G. B., S. Levis, S. Sitch, M. Versteijn, and K. W. Oleson (2003), A dynamical global vegetation model for use with climate models: Concepts and description of simulated vegetation dynamics, *Global Change Biol.*, **9**, 1543–1566.
- Cao, M. K., and F. I. Woodward (1998), Net primary and ecosystem production and carbon stocks of terrestrial ecosystems and their responses to climate change, *Global Change Biol.*, **4**, 185–198.
- Cao, M. K., B. Tao, K. R. Li, X. M. Shao, and S. D. Prince (2003), Interannual variation in terrestrial ecosystem carbon fluxes in China from 1981 to 1998, *Acta Bot. Sin.*, **45**, 552–560.
- Cao, M. K., S. D. Prince, B. Tao, J. Small, and K. R. Li (2005), Regional pattern and interannual variations in global terrestrial carbon uptake in response to changes in climate and atmospheric CO<sub>2</sub>, *Tellus, Ser. B*, **57**, 210–217.
- Chen, L. J., G. H. Liu, and H. G. Li (2002), Estimating net primary productivity of terrestrial vegetation in China using remote sensing (in Chinese), *J. Remote Sens.*, **6**, 129–135.
- Ciais, P., P. P. Tans, M. Trolier, J. W. C. White, and R. J. Francey (1995), A large Northern Hemisphere terrestrial CO<sub>2</sub> sink indicated by the <sup>13</sup>C/<sup>12</sup>C ratio of atmospheric CO<sub>2</sub>, *Science*, **269**, 1098–1102.
- Clark, D. A., S. Brown, D. W. Kicklighter, J. Q. Chambers, J. R. Thomlinson, J. Ni, and E. A. Holland (2001), Net primary production in tropical forests: An evaluation and synthesis of existing field data, *Ecol. Appl.*, **11**, 371–384.
- Coquard, J., P. B. Buffy, K. E. Taylor, and J. P. Iorio (2004), Present and future surface climate in the western USA as simulated by 15 global climate models, *Clim. Dyn.*, **23**, 455–472.
- Cox, P. M., R. A. Betts, C. D. Jones, S. A. Spall, and L. J. Totterdell (2000), Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model, *Nature*, **408**, 184–187.
- Cramer, W., D. W. Kicklighter, A. Bondeau, B. Moore, G. Churkina, B. Nemry, A. Ruimy, A. L. Schloss, and the participants of the Potsdam NPP Model Intercomparison (1999), Comparing global models of terrestrial net primary productivity (NPP): Overview and key results, *Global Change Biol.*, **5**, suppl. 1, 1–15.
- Dan, L., J. J. Ji, and Y. P. Li (2005), Climatic and biological simulations in a two-way coupled atmosphere-biosphere model (CABM), *Global Planet. Change*, **47**, 153–169.
- 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 (SiB), *J. Appl. Meteorol.*, **28**, 833–855.
- Field, C. B., M. J. Behrenfeld, J. T. Randerson, and P. Falkowski (1998), Primary production of the biosphere: Integrating terrestrial and oceanic components, *Science*, **281**, 237–240.
- Foley, J. A., S. Levis, I. C. Prentice, D. Pollard, and S. L. Thompson (1998), Coupling dynamical models of climate and vegetation, *Global Change Biol.*, **4**, 561–579.
- Forestry Ministry of China (1994), *Statistics on Forest Resources of China (1989–1993)* (in Chinese), 385 pp., For. Publ., Beijing, China.
- Goetz, S. J., S. D. Prince, J. Small, and A. C. R. Gleason (2000), Interannual variability of global terrestrial primary production: Results of a model driven with satellite observations, *J. Geophys. Res.*, **105**, 20,077–20,091.
- Gordon, W. S., J. S. Famiglietti, N. L. Fowler, T. G. F. Kittel, and K. A. Hibbarde (2004), Validation of simulated runoff from six terrestrial ecosystem models: Results from VEMAP, *Ecol. Appl.*, **14**, 527–545.
- Haywood, J. M., R. J. Stouffer, R. T. Wetherald, S. Manabe, and V. Ramaswamy (1997), Transient response of a coupled model to estimated changes in greenhouse gas and sulphate concentrations, *Geophys. Res. Lett.*, **24**, 1335–1338.
- He, Y., L. Dan, W. J. Dong, J. J. Ji, and D. H. Qin (2005), The terrestrial NPP simulations in China since Last Glacial Maximum, *Chin. Sci. Bull.*, **50**, 2074–2079.
- Hicke, J. A., G. P. Asner, J. T. Randerson, C. Tucker, S. Los, R. Birdsey, J. C. Jenkins, C. Field, and E. Holland (2002), Satellite-derived increases in net primary productivity across North America, 1982–1998, *Geophys. Res. Lett.*, **29**(10), 1427, doi:10.1029/2001GL013578.
- Houghton, J. T., Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, and D. Xiaosu (Eds.) (2001), Model evaluation, in *Climate Change 2001: The Scientific Basis—Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, pp. 471–523, Cambridge Univ. Press, New York.
- Hunt, E. R., S. C. Piper, R. Nemani, C. D. Keeling, R. D. Otto, and S. W. Running (1996), Global net carbon exchange and intra-annual atmospheric CO<sub>2</sub> concentrations predicted by an ecosystem process model and three-dimensional atmospheric transport model, *Global Biogeochem. Cycles*, **10**, 431–456.
- Ji, J. J. (1995), A climate-vegetation interaction model: Simulating physical and biological processes at the surface, *J. Biogeogr.*, **22**, 445–451.
- Ji, J. J., and Y. C. Hu (1989), A simple land surface process model for use in climate studies, *Acta Meteorol. Sin.*, **3**, 344–353.
- Ji, J. J., and L. Yu (1999), A simulation study of coupled feedback mechanism between physical and biogeochemical processes at the surface (in Chinese), *Chin. J. Atmos. Sci.*, **23**, 439–448.
- Jiang, H., M. J. Apps, Y. L. Zhang, C. H. Peng, and P. M. Woodare (1999), Modelling the spatial pattern of net primary productivity in Chinese forests, *Ecol. Modell.*, **122**, 275–288.

- Ju, L. X., and H. J. Wang (2006), A regional climate model nested in a global gridpoint general circulation model, *Chin. J. Geophys.*, **49**, 49–58.
- Kicklighter, D. W., A. Donbeau, A. L. Schloss, J. Kaduk, A. D. McGuire, and the participants of the Potsdam NPP Model Intercomparison (1999), Comparing global models of terrestrial net primary productivity (NPP): Global pattern and differentiation by major biomes, *Global Change Biol.*, **5**, suppl. 1, 16–24.
- Lau, N. C., and M. J. Nath (2004), Coupled GCM simulation of atmosphere-ocean variability associated with zonally asymmetric SST changes in the tropical Indian Ocean, *J. Clim.*, **17**, 245–265.
- Li, Y. P., and J. J. Ji (2001), Model estimates of global carbon flux between vegetation and the atmosphere, *Adv. Atmos. Sci.*, **18**, 807–818.
- Liu, Y. M., G. X. Wu, H. Liu, and P. Liu (2001), Condensation heating of the Asian summer monsoon and the subtropical anticyclone in the Eastern Hemisphere, *Clim. Dyn.*, **19**, 327–338.
- Liu, Y. M., G. X. Wu, and R. C. Ren (2004), Relationship between the subtropical anticyclone and diabatic heating, *J. Clim.*, **17**, 682–698.
- Lu, J. H., and J. J. Ji (2002), A simulation study of atmosphere-vegetation interactions over the Tibetan Plateau, part II: Net primary productivity and leaf area index (in Chinese), *Chin. J. Atmos. Sci.*, **26**, 254–262.
- Lu, J. H., and J. J. Ji (2006), A simulation and mechanism analysis of long-term variations at land surface over arid/semi-arid area in north China, *J. Geophys. Res.*, **111**, D09306, doi:10.1029/2005JD006252.
- Nemani, R., M. White, P. Thornton, K. Nishida, S. Reddy, J. Jenkins, and S. Running (2002), Recent trends in hydrologic balance have enhanced the terrestrial carbon sink in the United States, *Geophys. Res. Lett.*, **29**(10), 1468, doi:10.1029/2002GL014867.
- Nemani, R. R., C. D. Keeling, H. Hashimoto, W. M. Jolly, S. C. Piper, C. J. Tucker, R. B. Myneni, and S. W. Running (2003), Climate-driven increases in global terrestrial net primary production from 1982 to 1999, *Science*, **300**, 1560–1563.
- 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.
- Olson, R. J., K. Johnson, D. Zheng, and J. M. O. Scurlock (2001), Global and regional ecosystem modeling: Databases of model drivers and validation measurements, *ORNL/TM-2001/196*, 84 pp., Environ. Sci. Div., Oak Ridge Natl. Lab., Oak Ridge, Tenn.
- Phillips, N. A. (1973), Principles of large scale numerical weather prediction, in *Dynamic Meteorology*, edited by P. Morel, pp. 3–96, Springer, New York.
- Potter, C. S. (1999), Terrestrial biomass and the effects of deforestation on the global carbon cycle, *Bioscience*, **49**, 769–778.
- Potter, C. S., S. A. Klooster, R. Myneni, V. Genovesi, P.-N. Tan, and V. Kumar (2003), Continental-scale comparisons of terrestrial carbon sinks estimated from satellite data and ecosystem modeling 1982–1998, *Global Planet. Change*, **39**, 201–213.
- Prince, S., and J. Small (2003), AVHRR Global Production Efficiency Model, 1981–2000, Global Land Cover Facil., College Park, Md.
- Raich, J. W., and K. J. Nadelhoffer (1989), Belowground carbon allocation in forest ecosystems; global trends, *Ecology*, **70**, 1346–1354.
- Raupach, M. R., and C. Moran (1998), Material budgets as an organising framework: Contributions to methods development for the national and water resources audit, *Tech. Rep. 27/98*, 23 pp., CSIRO Land and Water, Canberra, A. C. T., Australia.
- Roxburgh, S. H., et al. (2004), A critical overview of model estimates of net primary productivity for the Australian continent, *Functional Plant Biol.*, **31**, 1043–1059.
- Ruimy, A., L. Kergoat, A. Bondeau, and the participants of the Potsdam NPP Model Intercomparison (1999), Comparing global models of terrestrial net primary productivity (NPP): Analysis of differences in light absorption and light-use efficiency, *Global Change Biol.*, **5**, suppl. 1, 56–64.
- Running, S. W., and E. R. Hunt (1993), Generalization of a forest ecosystem process model for other biomes, BIOME-BGC, and an application for global-scale models, in *Scaling Physiological Processes: Leaf to Globe*, pp. 141–158, edited by J. R. Ehleringer and C. Field, Elsevier, New York.
- Running, S. W., D. D. Baldocchi, D. P. Turner, S. T. Gower, P. S. Bakwin, and K. A. Hibbard (1999), A global terrestrial monitoring network integrating tower fluxes, flask sampling, ecosystem modeling and EOS satellite data, *Remote Sens. Environ.*, **70**, 108–127.
- Scurlock, J. M. O., and R. J. Olson (2002), Terrestrial net primary productivity—A brief history and new worldwide database, *Environ. Rev.*, **10**, 91–109.
- Scurlock, J. M. O., W. Cramer, R. J. Olson, W. J. Parton, and S. D. Prince (1999), Terrestrial NPP: Towards a consistent data set for global model evaluation, *Ecol. Appl.*, **9**, 913–919.
- Sun, R., and Q. J. Zhu (2001), Estimation of net primary productivity in China using remote sensing data, *J. Geogr. Sci.*, **11**, 14–23.
- Tao, B., K. R. Li, X. M. Shao, and M. K. Cao (2003), Temporal and spatial pattern of net primary production of terrestrial ecosystems in China (in Chinese), *Acta Geogr. Sin.*, **58**, 372–380.
- Tian, H., J. M. Melillo, D. W. Kicklighter, A. D. McGuire, and J. Helfrich (1999), The sensitivity of terrestrial carbon storage to historical climate variability and atmospheric CO<sub>2</sub> in the United States, *Tellus, Ser. B*, **51**, 414–452.
- Wang, B., H. Liu, and G. Y. Shi (2000), Radiation and cloud scheme, in *IAP Global Ocean Atmosphere-Land System Model*, edited by X. H. Zhang et al., chap. 3, pp. 28–49, Sci. Press, Beijing.
- Wang, S. Q., C. H. Zhou, J. Y. Liu, K. R. Li, and X. M. Yang (2001), Simulation analyses of terrestrial carbon cycle balance model in northeast China (in Chinese), *Acta Geogr. Sin.*, **56**, 390–400.
- Wang, Y., L. A. Mysak, and N. T. Roulet (2005), Holocene climate and carbon cycle dynamics: experiments with the “green” McGill Paleoclimate Model, *Global Biogeochem. Cycles*, **19**, GB3022, doi:10.1029/2005GB002484.
- Wang, Y. P., and D. J. Barrett (2003), Estimating regional terrestrial carbon fluxes for the Australian continent using a multiple-constraint approach, part I: Using remotely sensed data and ecological observations of net primary production, *Tellus, Ser. B*, **55**, 270–289.
- Williams, J. W., E. W. Seabloom, D. Slayback, D. M. Stoms, and J. H. Viers (2005), Anthropogenic impacts upon plant species richness and net primary productivity in California, *Ecol. Lett.*, **8**, 127–137.
- Wu, G. X., H. Liu, Y. C. Zhao, and W. P. Li (1996), A nine-layer atmospheric general circulation model and its performance, *Adv. Atmos. Sci.*, **13**, 1–18.
- Wu, G. X., X. H. Zhang, H. Liu, Y. Q. Yu, X. Z. Jin, Y. F. Guo, S. F. Sun, W. P. Li, B. Wang, and G. Y. Shi (1997), Global ocean-atmosphere-land system model of LASG (GOALS/LASG) and its performance in simulation study (in Chinese), *Q. J. Appl. Meteorol.*, **8**, 15–28.
- Wu, T. W., P. Liu, Z. Z. Wang, Y. M. Liu, R. C. Yu, and G. X. Wu (2003), The performance of atmospheric component model R42L9 of GOALS/LASG, *Adv. Atmos. Sci.*, **20**, 726–742.
- Xiao, X. M., J. M. Melillo, D. W. Kicklighter, Y. Pan, A. D. McGuire, and J. Helfrich (1998), Net primary production of terrestrial ecosystems in China and its equilibrium responses to changes in climate and atmospheric CO<sub>2</sub> concentration, *Acta Phytocool. Sin.*, **22**, 97–118.
- Xue, Y., H. M. H. Juang, W. P. Li, S. Prince, R. DeFries, Y. Jiao, and R. Vasic (2004), Role of land surface processes in monsoon development: East Asia and West Africa, *J. Geophys. Res.*, **109**, D03105, doi:10.1029/2003JD003556.
- Zeng, N. (2003), Glacial-interglacial atmospheric CO<sub>2</sub> change—the glacial burial hypothesis, *Adv. Atmos. Sci.*, **20**, 677–693.
- Zeng, Q. C. (1963), Characteristic parameters and dynamical equations of atmospheric motions (in Chinese), *Acta Meteorol. Sin.*, **33**, 472–498.
- Zhang, T., G. X. Wu, and Y. F. Guo (2002), Energy budget bias in global coupled ocean-atmosphere-land model (in Chinese), *Acta Meteorol. Sin.*, **60**, 278–289.
- Zhao, M. S., F. A. Heinsch, R. R. Nemani, and S. W. Running (2005), Improvements of the MODIS terrestrial gross and net primary production global data set, *Remote Sens. Environ.*, **95**, 164–176.
- Zhao, M. S., S. W. Running, and R. R. Nemani (2006), Sensitivity of Moderate Resolution Imaging Spectroradiometer (MODIS) terrestrial primary production to the accuracy of meteorological reanalyses, *J. Geophys. Res.*, **111**, G01002, doi:10.1029/2004JG000004.
- Zheng, D. L., S. Prince, and R. Wright (2003), Terrestrial net primary production estimates for 0.5° grid cells from field observations—A contribution to global biogeochemical modeling, *Global Change Biol.*, **9**, 46–64.
- Zheng, D. L., S. D. Prince, and R. Wright (2004), NPP multi-biome: Gridded estimates for selected regions worldwide, 1989–2001, revision 2, <http://www.daac.ornl.gov>, Oak Ridge Natl. Lab. Distrib. Active Arch. Cent., Oak Ridge, Tenn.
- Zhou, G. S., and X. S. Zhang (1995), A natural vegetation NPP model (in Chinese), *Acta Phytocool. Sin.*, **19**, 193–200.
- Zhou, T. J., and Y. F. Qian (1995), The design and forecast verification of an one-way nested fine-mesh limited area numerical model (in Chinese), *J. Trop. Meteorol.*, **11**, 342–353.
- Zobler, L. (1986), A world soil file for global climate modeling, *NASA Tech. Memo. 87802*, 33 pp.

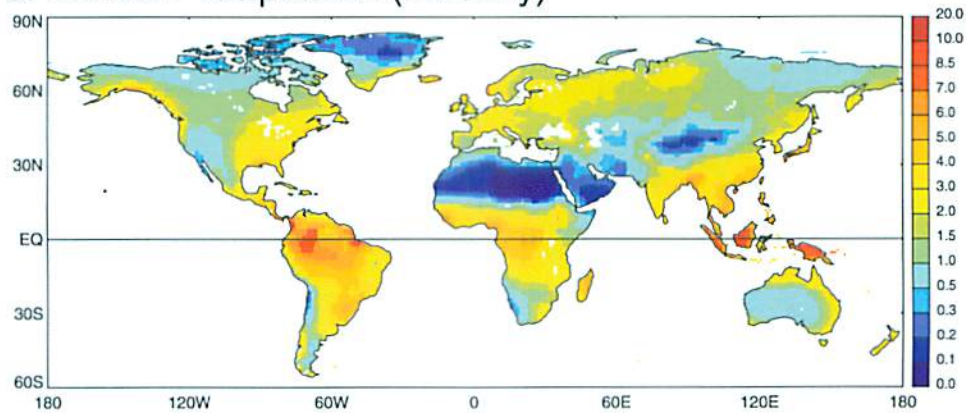
L. Dan and J. Ji, START Regional Center for Temperate East Asia and Key Laboratory of Regional Climate-Environment for Temperate East Asia, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China. (danli@tea.ac.cn)

Y. He, National Climate Center, China Meteorological Administration, Beijing 100081, China.

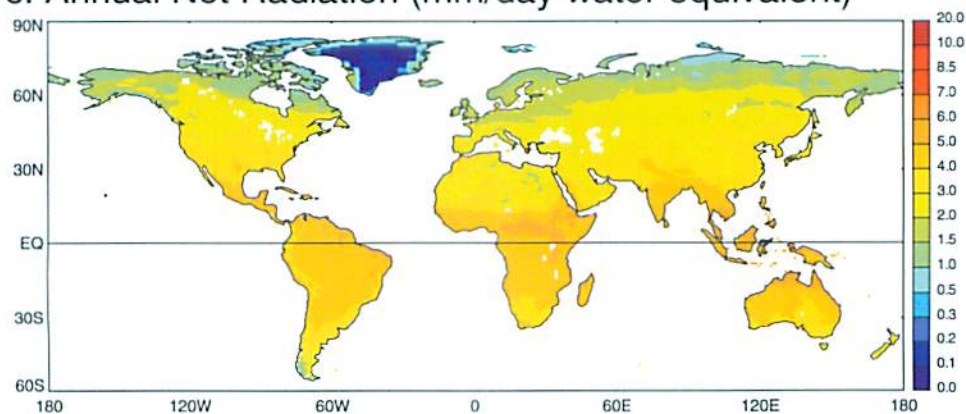
## a. Dryness Index



## b. Annual Precipitation (mm/day)

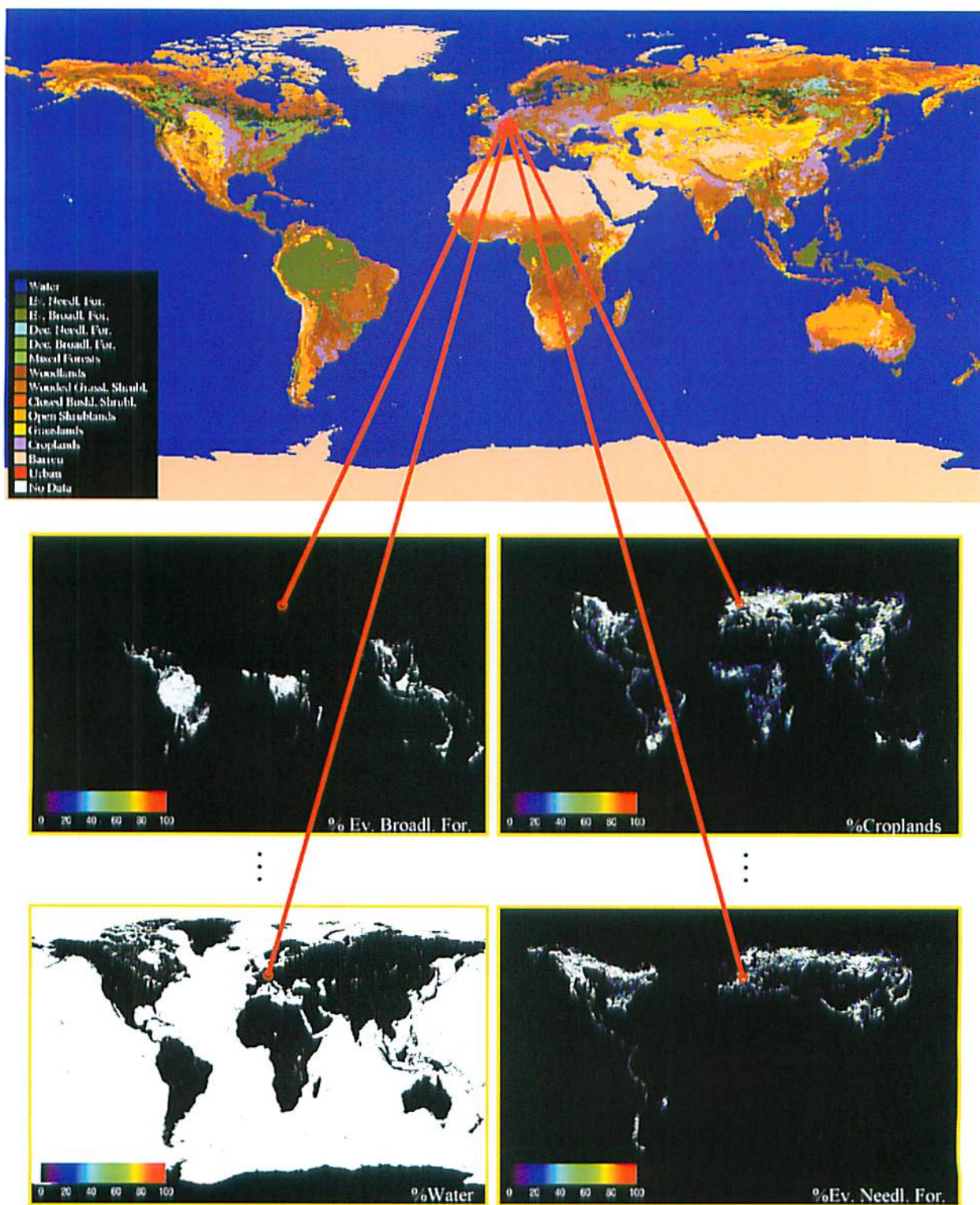


## c. Annual Net Radiation (mm/day water equivalent)



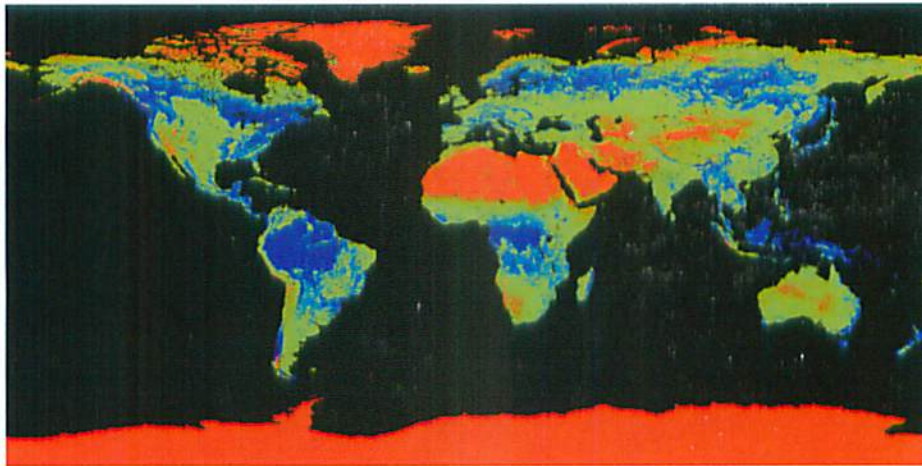
**Figure 4.** (a) Dryness index, as computed from the long-term annual means of precipitation and net radiation in the ISLSCP-2 data set. (b) Long-term annual mean of precipitation, from ISLSCP-2. (c) Long-term annual mean of net radiation (in terms of water equivalent), from ISLSCP-2.



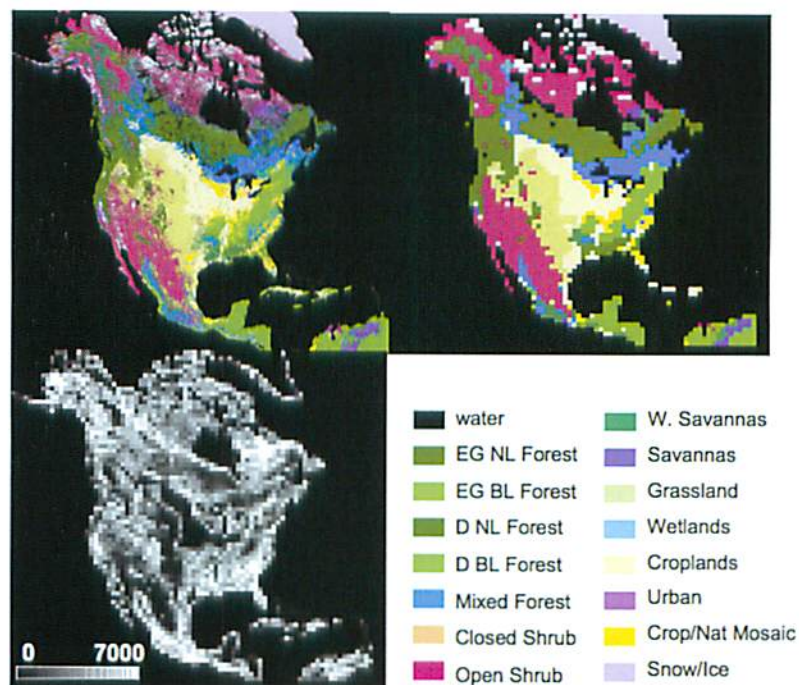


**Figure 1.** Dominant land cover type determined from the proportions of the individual land cover types at the original resolution of the products. The example here is for the University of Maryland (UMD) land cover product at 1/4° resolution. Only 4 of the 15 layers with subcell percentages of land cover types are shown for clarity. The dominant land cover type in each cell in the top map is determined from a combination of the percentage of each cover type within that same cell.





**Figure 2.** UMD vegetation continuous fields product at  $1/4^\circ$  spatial resolution. The % bare, herbaceous, and woody cover for each cell has been coded as red, green, and blue, respectively, to create this global representation. Other vegetation continuous fields provided in ISLSCP Initiative II include leaf type (needleleaf, broadleaf) and longevity (deciduous, broadleaf) for tree cover.



**Figure 1.** Differences between 1 km and 100 km ( $\sim 1^\circ$ ) land cover representations for North America. (top left) MODIS 1 km IGBP land cover classification, (top right) MODIS land cover map aggregated to  $1^\circ$  resolution using dominant type, and (bottom left) the difference between the area of the dominant type in the  $1^\circ$  cell (e.g., 10,000 km<sup>2</sup>) and the actual area of that type within the cell from the 1 km representation. In heterogeneous landscapes, the area of the dominant type is overestimated by up to 7000 km<sup>2</sup>, or 70% of the cell area.

