

THE GLOBAL LAND DATA ASSIMILATION SYSTEM

BY M. RODELL, P. R. HOUSER, U. JAMBOR, J. GOTTSCHALCK, K. MITCHELL, C.-J. MENG, K. ARSENAULT, B. COSGROVE, J. RADAKOVICH, M. BOSILOVICH, J. K. ENTIN, J. P. WALKER, D. LOHMANN, AND D. TOLL

This powerful new land surface modeling system integrates data from advanced observing systems to support improved forecast model initialization and hydrometeorological investigations.

Land and surface temperature and wetness conditions affect and are affected by numerous climatological, meteorological, ecological, and geophysical phenomena. Therefore, accurate, high-resolution estimates of terrestrial water and energy storages are

valuable for predicting climate change, weather, biological and agricultural productivity, and flooding, and for performing a wide array of studies in the broader biogeosciences. In particular, terrestrial stores of energy and water modulate fluxes between the land and atmosphere and exhibit persistence on diurnal, seasonal, and interannual time scales. Furthermore, because soil moisture, temperature, and snow are integrated states, biases in land surface forcing data and parameterizations accumulate as errors in the representations of these states in operational numerical weather forecast and climate models and their associated coupled data assimilation systems. That leads to incorrect surface water and energy partitioning, and, hence, inaccurate predictions. Reinitialization of land surface states would mollify this problem if the land surface fields were reliable and available globally, at high spatial resolution, and in near-real time.

A Global Land Data Assimilation System (GLDAS) has been developed jointly by scientists at the National Aeronautics and Space Administration (NASA) Goddard Space Flight Center (GSFC) and the National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Prediction (NCEP) in order to produce such fields. GLDAS makes use of the new generation of ground- and space-based observation systems, which provide

AFFILIATIONS: RODELL AND TOLL—Hydrological Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, Maryland; HOUSER AND BOSILOVICH—Hydrological Sciences Branch, and Data Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland; JAMBOR, GOTTSCHALCK, ARSENAULT, COSGROVE, RADAKOVICH, AND ENTIN*—Goddard Earth Science and Technology Center, University of Maryland, Baltimore County, Baltimore, Maryland; MENG—Goddard Earth Science and Technology Center, University of Maryland, Baltimore County, Baltimore, and NOAA/National Centers for Environmental Prediction, Camp Springs, Maryland; MITCHELL AND LOHMANN—NOAA/National Centers for Environmental Prediction, Camp Springs, Maryland; WALKER—Department of Civil and Environmental Engineering, University of Melbourne, Melbourne, Victoria, Australia
*Current affiliation: NASA, Washington, D.C.

CORRESPONDING AUTHOR: Dr. Matthew Rodell, Hydrological Sciences Branch, NASA Goddard Space Flight Center, Code 974.1, Greenbelt, MD 20771

E-mail: Matthew.Rodell@nasa.gov

DOI: 10.1175/BAMS-85-3-381

In final form 11 June 2003

© 2004 American Meteorological Society

data to constrain the modeled land surface states. Constraints are applied in two ways. First, by forcing the land surface models (LSMs) with observation-based meteorological fields, biases in atmospheric model-based forcing can be avoided. Second, by employing data assimilation techniques, observations of land surface states can be used to curb unrealistic model states.

Through innovation and an ever-improving conceptualization of the physics underlying earth system processes, LSMs have continued to evolve and to display an improved ability to simulate complex phenomena. Concurrently, increases in computing power and affordability are allowing global simulations to be run more routinely and with less processing time, at spatial resolutions that could only be simulated using supercomputers five years ago. GLDAS harnesses this low-cost computing power to integrate observation-based data products from multiple sources within a sophisticated, global, high-resolution land surface modeling framework.

What makes GLDAS unique is the union of all of these qualities: it is a global, high-resolution, offline (uncoupled to the atmosphere) terrestrial modeling system that incorporates satellite- and ground-based observations in order to produce optimal fields of land surface states and fluxes in near-real time. This article describes the major aspects of GLDAS and includes a sample of the output products. Subsequent scientific papers will present the results of several studies (now in various stages of completion) that are focusing on the data assimilation, validation, weather and climate model initialization, and other aspects of the project, in more detail than could be included in a single article.

BACKGROUND. *Modeling of the land surface.* Spurred by advances in the understanding of soil-water dynamics, plant physiology, micrometeorology, and the controls on atmosphere-biosphere-hydrosphere interactions, several LSMs have been developed in the past two decades with the goal of realistically simulating the transfer of mass, energy, and momentum between the soil and vegetation surfaces and the atmosphere. Currently, GLDAS drives three land surface models: Mosaic, Noah, and the Community Land Model (CLM). Additional models are slated for future incorporation, including the Variable Infiltration Capacity model (VIC; Liang et al. 1994) and the Catchment Land Surface Model (Koster et al. 2000). For a comparison of these and other LSMs, see results from the Project for Intercomparison of Land Surface Parameterization

Schemes (PILPS; Henderson-Sellers et al. 1995; Bowling et al. 2003) and the Global Soil Wetness Project (GSWP; Dirmeyer et al. 1999).

MOSAIC. Mosaic (Koster and Suarez 1996) is a well-established and theoretically sound LSM with roots in the Simple Biosphere model (SiB) of Sellers et al. (1986). The primary innovation of Mosaic was its treatment of subgrid-scale variability. It divides each model grid cell into a mosaic of tiles (after Avissar and Pielke 1989) based on the distribution of vegetation types within the cell. Surface flux calculations are similar to those described by Sellers et al. (1986).

CLM. The CLM is being developed by a grassroots collaboration of scientists who have an interest in making a general land surface model available for public use (Dai et al. 2003). The project is not controlled by any single organization or scientist, rather, the science is steered by the community. CLM includes superior components from each of three contributing models: the NCAR Land Surface Model (Bonan 1998), the Biosphere-Atmosphere Transfer Scheme (BATS; Dickinson et al. 1986), and the LSM of the Institute of Atmospheric Physics of the Chinese Academy of Sciences (Dai and Zeng 1997). Both of the first two “frozen” versions of CLM are included in the GLDAS.

NOAH. Since 1993, as a core project within the Global and Energy Water Cycle Experiment (GEWEX) Continental-Scale International Project (GCIP), NCEP has spearheaded a continuing collaboration of GCIP and other investigators from both public and private institutions to develop a modern LSM to be used for operations and research in NCEP weather and climate prediction models and their data assimilation systems, and also to be supported and distributed for community usage. The Noah LSM (Chen et al. 1996; Koren et al. 1999) was borne of that effort. Noah has been used operationally in NCEP models since 1996, and it continues to benefit from a steady progression of improvements (Betts et al. 1997; Ek et al. 2003).

Land data assimilation. Model predictions and observations are imperfect, and they contain different types of information. Observations may be highly accurate at discrete points in space and time, but they are subject to instrument failures, measurement drift, data stream interruptions, questions of spatial representativeness, and flaws in the algorithms used to derive useful quantities from measured signals. Models synthesize all of our knowledge of physical processes and

perform the thousands of calculations necessary to simulate a piece of the earth system, but they are limited by oversimplifications and misunderstandings of the myriad processes and feedback mechanisms that are active in the real world, as well as by errors in the data used to force the models. Data assimilation merges measurements with model predictions, with the goal of maximizing spatial and temporal coverage, consistency, resolution, and accuracy. Several data assimilation approaches are being tested for inclusion in GLDAS, including both the ensemble and the extended Kalman filters (Kalman 1960; Walker and Houser 2001), optimal interpolation, and hybrid insertion techniques.

North American LDAS. GLDAS has its basis in the North American Land Data Assimilation System (NLDAS) project (Mitchell et al. 1999; Cosgrove et al. 2003). NLDAS was initiated in 1998 as a multiinstitution collaboration led by NCEP, with the goal of producing land surface states and fluxes via an uncoupled land-only system that relied as much as possible on observation-based forcing fields, in order to avoid known biases in atmospheric model-based forcing fields. The study region for NLDAS encompasses the conterminous United States and parts of Mexico and Canada. Four LSMs are run by the contributing groups at 0.125° resolution and are subject to the same forcing fields, thus enabling unambiguous intercomparison of the LSM simulations. NLDAS models execute in real time at NCEP and, retrospectively, at NASA GSFC, Princeton University, the University of Washington, and NOAA's Office of Hydrologic Development. NLDAS states, fluxes, and forcings are being validated by researchers at Rutgers University and the University of Maryland as well as the aforementioned institutions. Much of the GLDAS driver code was derived from GSFC's NLDAS driver code, and many of the project goals are similar, albeit over different spatial domains.

METHODS AND SPECIFICATIONS. One of the primary objectives of the GLDAS project was to develop a driver that allows users to run multiple LSMs without specific knowledge of the models' architectures or physics. Designing a simulation with any of the models only requires modification of a single, simple inter-

face file. GLDAS routines adapt the forcing data to the individual input requirements of each LSM. Run-time options are provided in the GLDAS user interface file (these are summarized in Table 1). One is the forcing data source, as described in the section titled "Surface forcing fields." Another is the method of LSM state variable initialization: 1) the user can declare a globally uniform value for each variable; 2) the values can be taken from a restart file produced by a prior run; 3) GLDAS can input the surface state variable fields produced by the land model coupled to the atmospheric forecast and analysis system that produced the baseline forcing data (forcing data initialization option). Although time to spin up a model to a self-consistent state is still required with the third approach, it is greatly reduced. GLDAS static parameter fields, subgrid-scale variability, and data assimilation options are described below.

Land surface parameters. **VEGETATION.** A high-quality vegetation classification map is critical to GLDAS for three reasons. First, the Mosaic, CLM, and Noah models incorporate soil-vegetation-atmosphere transfer schemes (SVATS), so that the fluxes and storages of energy and water at the land surface are strongly tied to the properties of the vegetation. Second, the vegetation type dictates other parameters such as albedo and roughness height. Third, GLDAS simulates subgrid-scale variability based on vegetation, as described in the section titled "Subgrid-scale variability." GLDAS uses a static, 1-km resolution, global dataset of land cover class that was produced at the University of Maryland (UMD) based on observations from the Advanced Very High Resolution Radiometer (AVHRR) aboard the NOAA-15 satellite (Hansen et al. 2000). As shown in Fig. 1, the UMD

TABLE 1. Basic options available in the GLDAS user interface.

Spatial resolution	0.25°; 0.5°; 1.0°; 2.0° × 2.5°
Temporal resolution	Adjustable model time step and output interval
Land surface model	Mosaic; CLM; Noah
Forcing	Various analysis- and observation-based products
Initialization	None (constant value); restart file; forcing data
Subgrid variability	1–13 tiles per grid cell (constant or fractional cutoff)
Elevation adjustment	Temperature; pressure; humidity; longwave radiation
Data assimilation	Surface temperature; snow cover
Soil classification	Lookup table; Reynolds et al. (1999)
Leaf area index	Lookup table; AVHRR derived

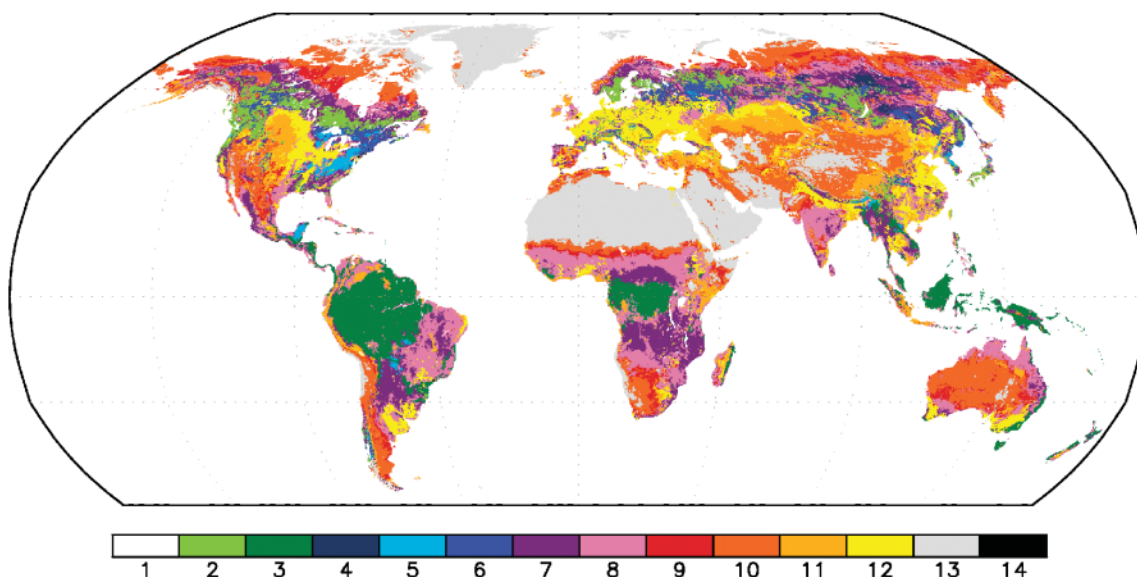


Fig. 1. Predominant UMD vegetation type in each 0.25° grid cell. Key: 1 = water, 2 = evergreen needleleaf forest, 3 = evergreen broadleaf forest, 4 = deciduous needleleaf forest, 5 = deciduous broadleaf forest, 6 = mixed cover, 7 = woodland, 8 = wooded grassland, 9 = closed shrubland, 10 = open shrubland, 11 = grassland, 12 = cropland, 13 = bare ground, 14 = urban and build-up.

classification includes 11 vegetation types in addition to water, bare ground, and urban covers.

GLDAS also ingests a satellite-based, 1-km resolution time series (when available) or climatology of leaf area index (LAI). These were generated using three information sources (Gottschalck et al. 2002): 1) a 16-km resolution time series of LAI (currently spanning July 1981–July 2001), which was derived by scientists at Boston University (BU; Myneni et al. 1997) from AVHRR measurements of normalized difference vegetation index (NDVI) and other satellite observations; 2) a climatology based on the 16-km time series, in which LAI is indexed by 10° latitude zones, month of the year, and vegetation type; and 3) the 1-km UMD vegetation-type classification. The information is blended so that the resulting 1-km pixel values vary by vegetation type while the BU 16-km average LAIs are maintained. GLDAS scales the 1-km data to the selected model resolution and adjusts for fractional vegetation cover.

SOILS. The soil parameter maps used in GLDAS were derived from the 5' resolution global soils dataset of Reynolds et al. (2000). Porosity and the percentages of sand, silt, and clay were horizontally resampled to the 0.25° GLDAS grid and linearly interpolated to 0–2-, 2–150-, and 150–350-cm depths from the original 0–30- and 30–100-cm depths. Those depths were chosen to match the set of depths most commonly assigned in the original version of Mosaic and to facili-

tate future assimilation of surface soil moisture fields derived from Advanced Microwave Scanning Radiometer (AMSR) satellite measurements. Certain parameters employed by the LSMs are indexed to the U.S. Department of Agriculture (USDA) soil texture class. Therefore, GLDAS includes a routine to classify the texture based on the percentages of sand, silt, and clay in a given grid cell, resulting in the map shown in Fig. 2. CLM also employs soil color as a parameter. GLDAS soil colors are interpolated from a 2° × 2.5° global map produced at NCAR.

ELEVATION AND SLOPE. GLDAS uses a global 30-arc-s resolution topographic map (GTOPO30; Verdin and Greenlee 1996) as its standard. GTOPO30 elevations were averaged onto the 0.25° GLDAS grid. By default, GLDAS corrects the modeled temperature, pressure, humidity, and longwave radiation forcing fields based on the difference between the GLDAS elevation definition and the elevation definition of the model that created the forcing data, following Cosgrove et al. (2003). Because some LSMs, including Mosaic, ingest surface or bedrock slope as a parameter, geographic information systems software was used to assess the slope at each GTOPO30 pixel, and from those values the mean slope within each GLDAS grid cell was computed.

Subgrid-scale variability. The standard operational GLDAS model runs are performed on a 0.25° × 0.25°

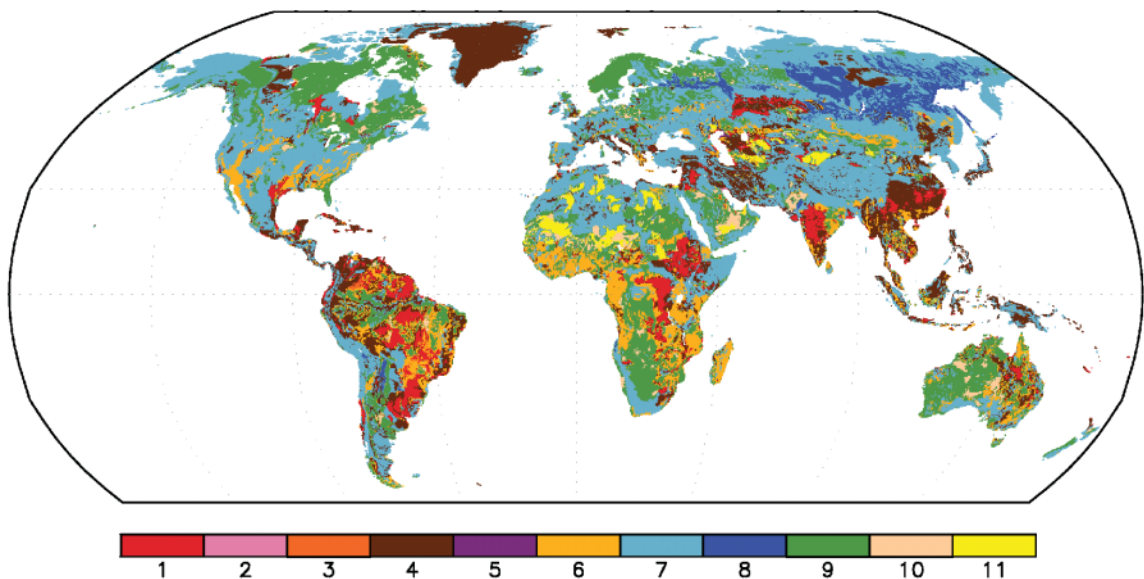


FIG. 2. USDA soil texture class in each 0.25° grid cell, derived from Reynolds et al. (1999). Key: 1 = clay, 2 = silty clay, 3 = sandy clay, 4 = clay loam, 5 = silty clay loam, 6 = sandy clay loam, 7 = loam, 8 = silty loam, 9 = sandy loam, 10 = loamy sand, 11 = sand.

grid, which is nearly global, covering all of the land north of 60°S latitude. GLDAS also is able to run on $0.5^\circ \times 0.5^\circ$, $1.0^\circ \times 1.0^\circ$, and $2.0^\circ \times 2.5^\circ$ global grids. Following Koster and Suarez (1992), subgrid variability is simulated by running the LSMs on a series of independent soil columns, or tiles, where each tile represents one vegetation class within a grid cell. Gridded output is returned by weighting each tile by its fractional coverage within the cell. The aim of this approach is to encapsulate subgrid-scale water and energy flux variability, which may be significant at model scales (e.g., Vukovich et al. 1997). Users select a maximum whole number of tiles to be defined per grid cell. This can be as many as 13, which is the number of land cover types in the UMD vegetation classification. In addition, users select the smallest percentage of a cell for which to create a tile. For example, given settings of three tiles and 12%, a tile would be created for any vegetation type that covers at least 12%

of a grid cell, for up to three tiles per cell. Any vegetation type that covers less than 12% of the cell or is not among the three most common types would be omitted, and the resulting tile areas would be normalized to 100%. Figure 3 illustrates this example. Tests have shown that constraining the number of tiles per grid using the percentage cutoff is more efficient (i.e., the sum of all the subgrid areas containing a vegetation type that was omitted is smaller given the same global-total number of tiles) than setting a maximum

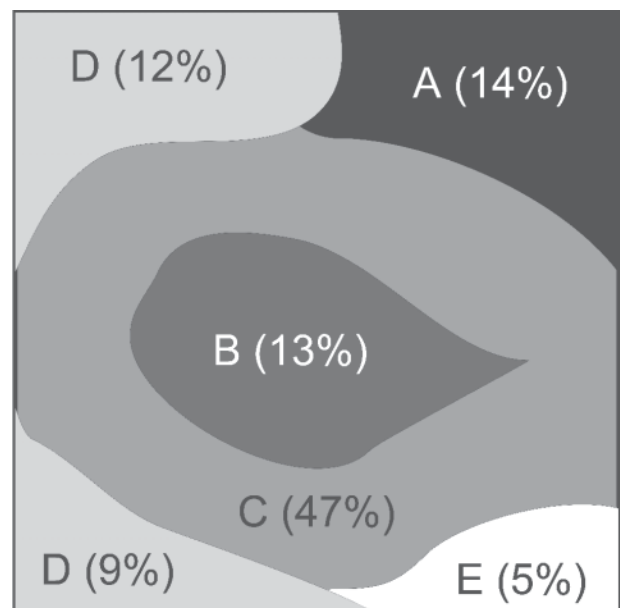


FIG. 3. The spatial coverage of five vegetation types (A–E) within a fictional grid cell is shown. Given this information along with user-defined cutoffs, e.g., 3 = maximum tiles per grid and 12% = minimum grid coverage to define a tile, GLDAS would 1) eliminate vegetation-type E (less than 12%), 2) eliminate vegetation-type B (not among three predominant vegetation types), and 3) normalize the weights of the three remaining tiles to 100%, resulting in three tiles: A (17%), C (57%), D (26%).

number of tiles per grid (Toll et al. 2001). Based on these tests, the standard for GLDAS operational runs is to set 10% as the minimum tile area percentage and 13 (i.e., no constraint) as the maximum tiles per grid. This results in 505,680 tiles at 0.25° grid resolution, or an average of just over two tiles per cell.

Data assimilation. GLDAS includes an algorithm for assimilating remotely sensed, 6-hourly skin temperature observations (Ottle and Vidalmadjar 1992) from the Television Infrared Observation Satellite (TIROS) Operational Vertical Sounder (TOVS) instrument. The algorithm relies on an optimal interpolation routine to calculate an analysis increment, which is used to perform an incremental, semidaily, or daily skin temperature bias correction. The bias correction continually steers the modeled state toward the observation. The observed-minus-forecast value for the skin temperature is first calculated by GLDAS and passed into the optimal interpolator to retrieve the analysis increment. This is then relayed, along with a bias correction term, to the LSM code for proper energy budget considerations. Testing has shown that this technique effectively constrains the modeled surface temperature (Radakovich et al. 2001), though the results are limited by the temporal sparseness of the TOVS data. The same technique could potentially be applied using 3-hourly skin temperature data from the International Satellite Cloud Climatology Project (ISCCP).

Methods for assimilating surface soil moisture data using forms of the ensemble and extended Kalman filters have been designed and tested by members of the GLDAS science team (Walker and Houser 2001; Zhan et al. 2002), but these have not yet been implemented in GLDAS due to a lack of global observations of soil moisture. The AMSR instrument, aboard NASA's Earth Observing System (EOS) *Aqua* satellite, has begun to deliver global C-band microwave observations. Under certain land cover conditions these may soon be used to derive near-surface (0–2 cm) soil moisture (Owe et al. 2001).

GLDAS can assimilate snow cover information derived from measurements made by the Moderate Resolution Imaging Spectroradiometer (MODIS) aboard NASA's *Terra* satellite (Rodell et al. 2002). The MODIS science team at GSFC provides a daily, 0.05° gridded global snow cover dataset (D. K. Hall et al. 2002). For each grid cell over land, this product reports the percentages of snow-covered, non-snow-covered, and nonvisible (i.e., obscured by cloud cover, night, or otherwise) subgrid pixels. The GLDAS snow correction algorithm determines the snow-cover state

(snow covered or bare) at each model grid cell based on the ratio of snow-covered pixels to visible pixels and uses the number of visible pixels as a measure of observation reliability. If the observation is deemed unreliable or if the observation and model agree on the snow-cover state in a grid cell, the modeled snow condition is not changed. Otherwise, if MODIS indicates that the grid cell is bare but the model shows snow, then the model's snow is removed. If MODIS reveals snow but the model shows no snow, then a thin cover of snow is added to the model grid cell. The albedo changes automatically. Testing is currently under way to improve the technique by adjusting near-surface temperature, soil moisture, and/or the precipitation type (rain or snow) of prior events to reflect the observed snow-cover state, and hence to minimize unintended effects on the water balance.

SURFACE FORCING FIELDS. Users of GLDAS select, as the baseline land surface forcing, output from the atmospheric data assimilation system (ADAS) component of a weather forecast and analysis system, or, for long-term retrospective runs, from a reanalysis product (next section). Observation-derived fields, including precipitation and longwave and shortwave radiation, are optionally specified to replace the corresponding ADAS forcing fields when and where possible. This capability reduces reliance on forcing fields that are largely model generated, while ensuring continuity of forcing. However, in choosing a forcing option, the strengths and weaknesses of the analysis and observation-based fields should be weighed, in particular, the lower biases of the observation-based field against the consistency and quality control inherent to the analysis fields. Variables required to force the LSMs are listed in Table 2.

Operational meteorological forecast and data assimilation system output. GEOS. The Goddard Earth Observing System (GEOS) Data Assimilation System (Pfaendtner et al. 1995) supports level-4 (model analysis) product generation for the NASA *Terra* satellite (Atlas and Lucchesi 2000). GEOS "first look" output is obtained for operational GLDAS simulations and then replaced later in the forcing archive with "late look" output as available, in order to improve the quality of GLDAS retrospective forcing. GEOS uses the physical-space statistical analysis system (Cohn et al. 1998) to assimilate NCEP's rawinsonde reports, TOVS retrievals, and SSM/I total precipitable water retrievals. It utilizes NCEP's operational sea surface temperature (SST) and sea

ice boundary conditions. GEOS 3-hourly fields are produced on a $1^\circ \times 1.25^\circ$ global grid.

GDAS. The Global Data Assimilation System (GDAS) is the operational global atmospheric data assimilation system of NCEP (Derber et al. 1991). GDAS executes on a thinned Gaussian grid with 768 grid points in the zonal direction and 364 grid points in the meridional direction (about 0.47° resolution). There are 64 model atmospheric levels. GDAS assimilates a variety of conventional data (radiosonde, buoy, ship, and airborne) and satellite-derived observations, using a four-dimensional multivariate approach, and produces operational, global analyses for four synoptic hours: 0000, 0600, 1200, and 1800 UTC. GLDAS makes use of the latter analyses, as well as the GDAS 3-h and (as needed) 6-h background forecast from each analysis.

ECMWF. The European Centre for Medium-Range Weather Forecasts (ECMWF) also produces operational, global analyses and forecasts for four synoptic hours: 0000, 0600, 1200, and 1800 UTC (Persson 2001). Their atmospheric assimilation system is run on a linear reduced Gaussian grid that has 553,384 surface grid points (approximately 39-km resolution). It includes 60 model atmospheric levels. The analysis also incorporates both conventional and satellite-derived data using a four-dimensional multivariate assimilation approach (Klinker et al. 2000).

Observation-derived data. **DOWNWARD SOLAR RADIATION.** GLDAS estimates global, downward shortwave and longwave radiation fluxes at the land surface using a procedure and cloud and snow products from the Air Force Weather Agency's (AFWA) Agricultural Meteorology modeling system (AGRMET). The snow product is the AFWA daily, 48-km global snow depth analysis (Kopp and Kiess 1996). Prior to July 2002, the cloud product was the AFWA Real Time Nepha-

TABLE 2. GLDAS forcing and output fields.

Required forcing fields	Summary of output fields
Precipitation	Soil moisture in each layer
Downward shortwave radiation	Snow depth, fractional coverage, and water equivalent
Downward longwave radiation	
Near-surface air temperature	Plant canopy surface water storage
Near-surface specific humidity	Soil temperature in each layer
Near-surface <i>U</i> wind	Average surface temperature
Near-surface <i>V</i> wind	Surface and subsurface runoff
Surface pressure	Bare soil, snow, and canopy surface water evaporation
	Canopy transpiration
	Latent, sensible, and ground heat flux
	Snow phase change heat flux
	Snowmelt
	Snowfall and rainfall
	Net surface shortwave and longwave radiation
	Aerodynamic conductance
	Canopy conductance
	Surface albedo

lysis (RTNEPH) 3-hourly, 48-km global cloud analysis (Hamill et al. 1992), thereafter replaced by AFWA's new hourly, 24-km World Wide Merged Cloud Analysis (WWMCA). The cloud and snow products are used to calculate surface downward shortwave radiation and longwave radiation based, respectively, on the AFWA-supplied algorithms of Shapiro (1987) and Idso (1981), which implement the shortwave and longwave procedures on a three-layer (high, middle, and low) plane-parallel atmosphere. For each layer, atmospheric transmissivity and reflectivity, with respect to shortwave radiation, and longwave radiation emitted by clouds are calculated as functions of cloud type and amount. The cloud and snow products are derived primarily from observations made by Defense Meteorological Satellite Program and NOAA satellites. RTNEPH incorporated only polar-orbiting observations, while WWMCA incorporates both geostationary and polar-orbiting observations. The strength of the resulting radiation fields is their use of satellite-derived cloud cover, as opposed to the model-based cloud cover used in the radiation calculations of the atmospheric data assimilation systems.

PRECIPITATION. Near-real-time satellite-derived precipitation data are obtained from groups at the U.S. Naval Research Laboratory (NRL) and at GSFC for forcing GLDAS. NRL produces precipitation fields based on geostationary satellite infrared (IR) cloud-top temperature measurements and microwave observation techniques (Turk et al. 2000). The microwave product merges data from the Special Sensor Microwave/Imager (SSM/I), NASA–NASDA’s Tropical Rainfall Measuring Mission (TRMM), and the Advanced Microwave Sounding Unit (AMSU) instruments. Both NRL products have a spatial resolution of 0.25° and a temporal resolution of 6 h and both cover an area from 60°S to 60°N. Archiving of those data for GLDAS began in April 2001. Scientists at GSFC began to provide near-real-time precipitation fields based on the optimal merging of the more accurate microwave data with the more frequent IR data in February 2002 (Huffman et al. 2003). When the satellite-derived precipitation forcing option is chosen, GLDAS overlays the baseline ADAS field with one of these two products as available. Another precipitation forcing option is being tested that makes use of the NOAA Climate Prediction Center’s (CPC’s) operational global 2.5° 5-day Merged Analysis of Precipitation (CMAP), which is a blending of satellite (IR and microwave) and gauge observations. GDAS modeled precipitation fields are used to disaggregate the CMAP fields spatially and temporally to match the GLDAS resolutions.

OPERATIONAL SIMULATION RESULTS.

Daily, parallel Mosaic and Noah model simulations began on 1 January 2001, initialized by GDAS land surface states. Currently, GEOS is set as the baseline forcing source, and the precipitation and radiation fields are overwritten with the observation-derived fields described in the section titled “Surface forcing fields,” when and where these are available. The model spatial resolution is 0.25°, and 10% is set as the minimum tile area percentage, as described in the section titled “Subgrid-scale variability.” The model time step is 15 min, and output is 3-hourly. Typically the daily runs are complete within 36–48 h of real time.

Results are presented from two of the operational Mosaic simulations: a control run and a derived forcing run. The control run relied on GEOS forcing exclusively. The derived forcing run used the NRL (prior to 30 June 2002) and GSFC observation-based precipitation fields and the AGRMET observation-based downward shortwave and longwave radiation fields. Figure 4 compares the root zone soil moisture output from the two simulations at the end of each sea-

son from spring 2002 to winter 2002/03. The broad seasonal patterns of wetness generally match, although some regional differences are apparent. The deserts of North Africa are more extensive in the control run. Central North America is wetter in the derived forcing run, while Southeast Asia tends to be wetter in the control run. (The effects of the latter two differences on modeled evapotranspiration can be seen in Fig. 6, described below.) Discrepancies between the modeled and satellite-derived precipitation forcing (Fig. 5) correspond to the differences in soil moisture. The sparseness of in situ observations of soil moisture will complicate the task of validating these results and choosing the better product, but AMSR-derived soil moisture fields may provide some guidance in the near future. Results over Greenland should be ignored, as the LSMs do not include ice sheet models and reliable forcing data and soil parameter information are not available.

The greater finescale variability of the observation-based precipitation (Fig. 5) is reflected in the finescale patterns of soil moisture in the derived forcing run (Fig. 4). Because soil moisture shows a high degree of variability at all scales (e.g., Famiglietti et al. 1999), the finescale soil moisture variability evident in the derived forcing run results are more realistic in regard to spatial variability. This may be preferable from a weather forecasting perspective, in that overly homogeneous surface conditions can deter the development of atmospheric instabilities. However, the broader patterns of soil moisture are not necessarily more accurate in the derived forcing run. Adler et al. (2001) found that the GSFC satellite-based precipitation product had a smaller bias and smaller root-mean-square error (rmse) than GDAS and GEOS over land between 30°S and 30°N, but a larger bias and rmse over land between 30° and 60°N. Prior to March 2003 when controls were instituted, problems with the satellite algorithm over snow caused overestimation of precipitation (G. Huffman 2003, personal communication), as seen in the satellite-derived December–February (DJF) plot (bottom right) in Fig. 5 over central Asia. The CMAP satellite-gauge product described in the subsection titled “Precipitation,” a disaggregated version of which is being tested as an additional forcing option, had a smaller bias and rmse than all of the modeled and satellite-only products over land in both zones.

Due to the high quality of the cloud observations that contribute to the AGRMET radiation fields (Hamill et al. 1992), their finescale patterns are likely to be realistic. Meng et al. (2003) has demonstrated that AGRMET shortwave radiation estimates are typi-

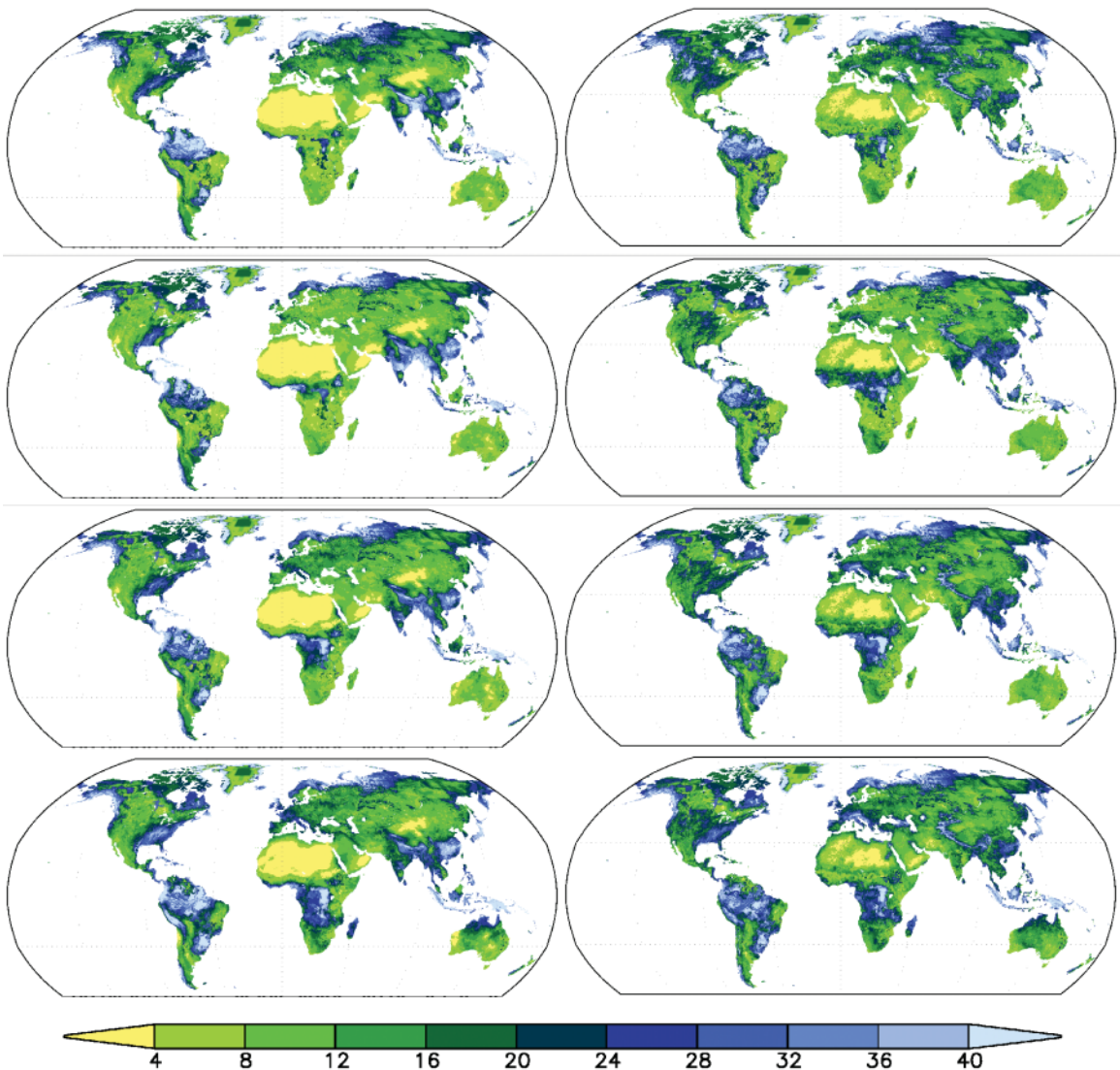


FIG. 4. Daily mean volumetric soil water content (%) in the root zone from two operational GLDAS simulations: (left) control run; (right) derived forcing run; (from top to bottom) 31 May 2002, 31 Aug 2002, 30 Nov 2002, 28 Feb 2003.

cally more consistent with station observations than ADAS results. Using these data as forcing should have a positive impact on the GLDAS simulations. Figure 6 zooms in on North America and Southeast Asia so that the small-scale patterns of input shortwave radiation and output evapotranspiration can be discerned. The greater finescale variability of the AGRMET than the GEOS shortwave radiation fields is apparent, while the broadscale patterns and quantities generally agree. Inspection of *GOES-8* and *GOES-10* visible imagery (not shown) for 1800 UTC 31 July 2002 reveals patterns of cloud cover over the southeastern, southwestern, and northwestern United States that are more consistent with the AGRMET shortwave radiation field. Evapotranspiration rates over the

western United States are lower in the control run than in the derived forcing run, which reflect the drier soils (Fig. 4). The opposite is true over India.

VALIDATION AND SCIENCE. Development of GLDAS is ongoing, but the emphasis of the project is shifting toward validation, scientific studies, and applications. Results from simulations using various combinations of LSMs, modeled and observation-derived forcing fields, data assimilation, and other options listed in Table 1, are being compared and analyzed. Participation in the Global Soil Wetness Project phase 2 will enable comparisons with results from tens of other land surface modeling groups. GLDAS output over North America also will be com-

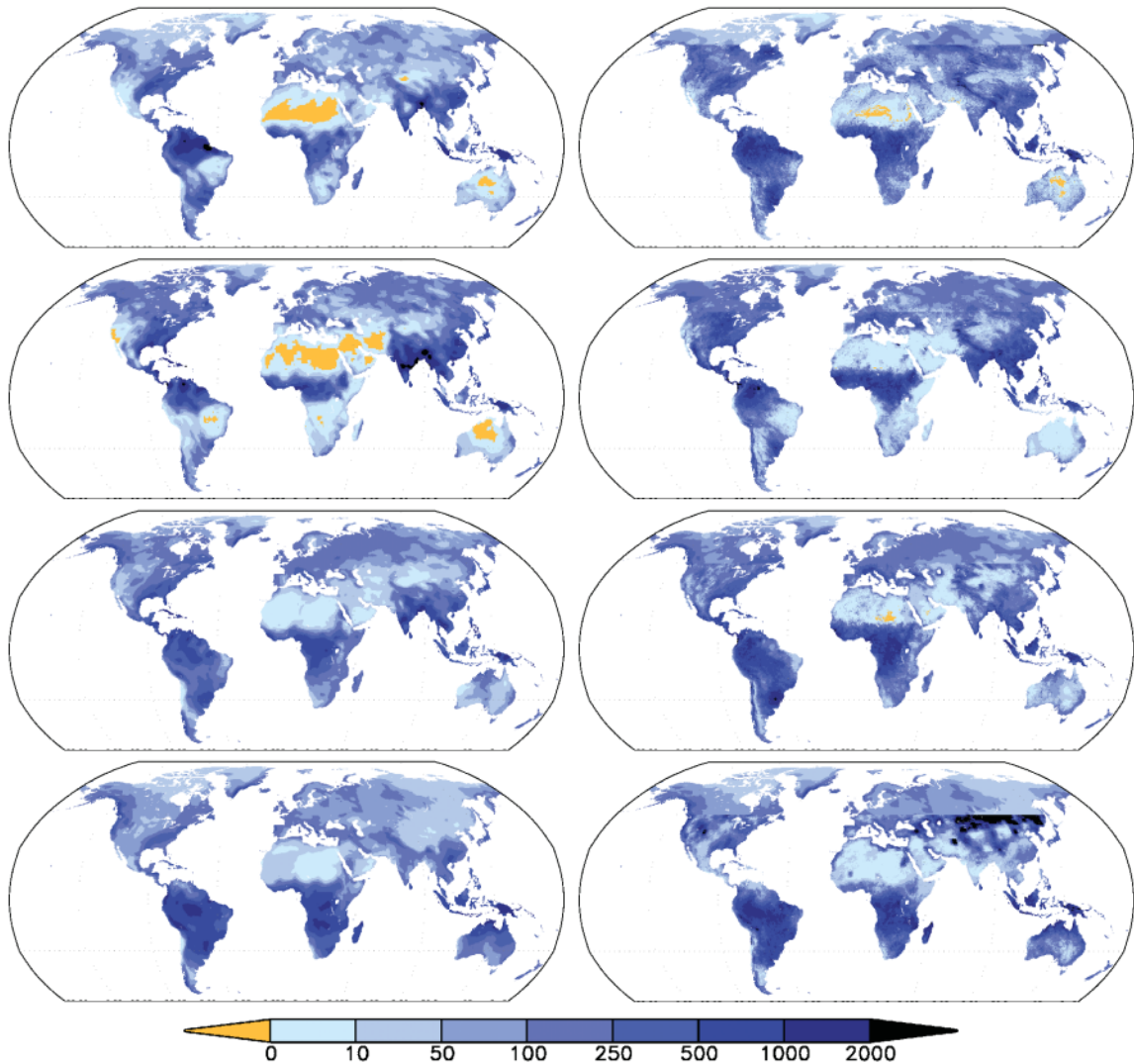


FIG. 5. Total precipitation (mm) forcing from two operational GLDAS simulations; (left) control run: GEOS-modeled precipitation; (right) derived forcing run: satellite-derived precipitation; (from top to bottom) Mar-Apr-May 2002, Jun-Jul-Aug 2002, Sep-Oct-Nov 2002, Dec-Jan-Feb 2002/03.

pared with 0.125° output from NLDAS, which makes use of several high-quality data sources that are not available globally. A study is under way to use assimilation techniques to validate modeled soil moisture using data from AMSR (Zhan et al. 2002). Data from ground-based monitoring stations, including those archived at the Global Soil Moisture Data Bank (Robock et al. 2000), and field experiments, including Soil Moisture Experiments (SMEX) in 2002 and 2003, the Cold Land Processes Experiment (CLPX) series, and others affiliated with the Coordinated Enhanced Observing Period initiative, will be compared with model location time series (MOLTS) from GLDAS. The MOLTS will be generated using the subgrid tile results, so that the vegetation type can be chosen to match that of the measurement location.

Validation and intercomparison of the various forcing data options will also be important.

Numerous sensitivity and scaling studies are possible given all of the options available in GLDAS, such as the effect of improved soil and vegetation parameterizations on model output. Initialization of weather and climate prediction models with GLDAS land surface fields is being tested in order to assess the effect on prediction skill. Preliminary results indicate that seasonal precipitation forecasts improve when GLDAS soil moisture conditions are used to initialize NASA's Seasonal to Interannual Prediction Project climate model (R. D. Koster et al. 2003, unpublished manuscript). Preparations are being made for similar testing with NCEP's seasonal and weather forecast systems. GLDAS also will provide a priori knowledge

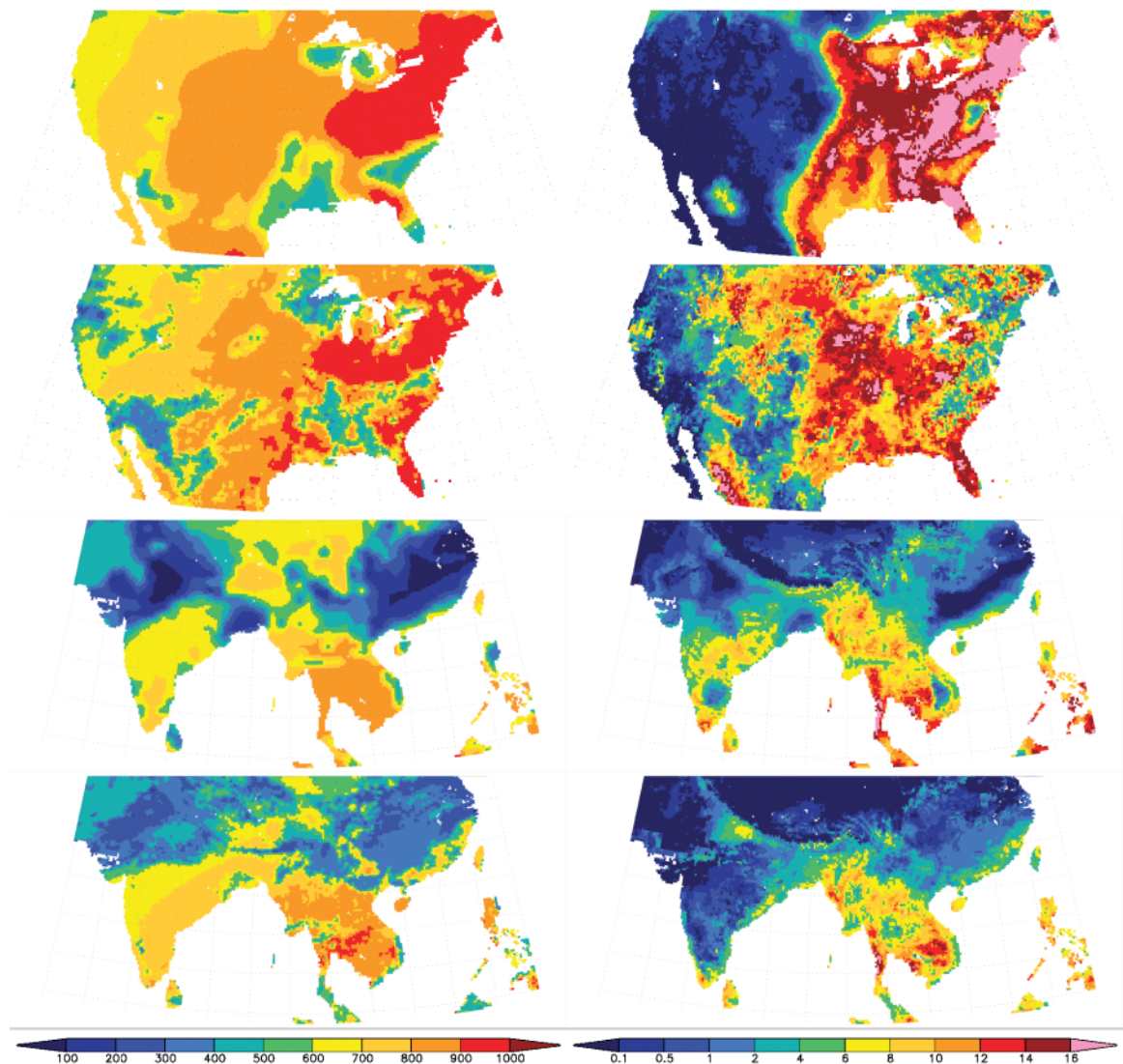


FIG. 6. (left) Downward shortwave radiation forcing (W m^{-2}) and (right) output total evapotranspiration rate (mm day^{-1}) from two operational GLDAS simulations: (from top to bottom) control run 1800–2100 UTC 31 Jul 2002, derived forcing run 1800–2100 UTC 31 Jul 2002, control run 0600–0900 UTC 31 Jan 2003, derived forcing run 0600–0900 UTC 31 Jan 2003; (top four) central North America; (bottom four) Southeast Asia.

of variations in the global distribution of terrestrial water mass, which will contribute to the optimization of global gravity field solutions based on observations from NASA's Gravity Recovery and Climate Experiment (GRACE). GRACE gravity observations will be used to estimate monthly changes in terrestrial water storage over regions larger than $200,000 \text{ km}^2$ (Rodell and Famiglietti 1999). Innovative data assimilation techniques to make use of these estimates in GLDAS are being explored.

DATA DISTRIBUTION. One of the most ambitious activities of the GLDAS project has been the assemblage of an archive of global, operational ADAS

output and observation-based data fields for parameterizing and forcing LSMs. Most of the time series begin around January 2001 and continue in the present. The most recent fields are downloaded daily from forecast centers and groups that process satellite data. Many of these fields become inaccessible through their providers not long after initial release, being archived to tape or not saved at all. However, the GLDAS active archive is public to the extent allowed by the providers, so that these data, once lost for all intents and purposes, are preserved for scientific usage. The parameter fields, including vegetation, soils, and elevation fields, are also accessible.

Output fields of land surface states and fluxes from

GLDAS model simulations, including those described in the section titled “Operational simulation results,” likewise are freely available to the public. The GLDAS Web site includes a real-time image generator that allows users to view the most recent output fields. Time series are currently available by request and will soon be available through an automated data server. In addition to the operational simulations, multiyear retrospective simulations are ongoing at lower resolutions ($0.5^\circ \times 0.5^\circ$ and $2.0^\circ \times 2.5^\circ$) using bias-corrected atmospheric reanalysis data (Berg et al. 2003) and data from the International Satellite Land Surface Climatology Project, Initiative II (F. G. Hall et al. 2002). Broad use of GLDAS results is encouraged for education, policy making, and social, agricultural, and natural hazards planning, as well as for scientific research. (More details on data availability are provided at <http://ldas.gsfc.nasa.gov>.)

ACKNOWLEDGMENTS. This work was supported by NASA’s Earth Observing System (EOS) Interdisciplinary Science Program. The authors wish to thank the Air Force Weather Agency for 1) supplying the global daily snow depth analysis and the global hourly USAF Worldwide Merged Cloud Analysis (WWMCA, formerly known as “RTNEPH”) in real time; and 2) providing, through the support of George Gayno, the surface radiation algorithms. We are also grateful for data products and advice provided by Joe Turk at NRL, Pedro Viterbo and many others at ECMWF, Ranga Myneni at BU, the PERSIANN precipitation group at The University of Arizona, and George Huffman and Randy Koster at NASA GSFC.

REFERENCES

- Adler, R. F., C. Kidd, G. Petty, M. Morissey, and H. M. Goodman, 2001: Intercomparison of global precipitation products: The third precipitation intercomparison project (PIP-3). *Bull. Amer. Meteor. Soc.*, **82**, 1377–1396.
- Atlas, R. M., and R. Lucchesi, 2000: File Specification for GEOS-DAS Gridded Output. DAO-1001v4.3, 41 pp. [Available online at <http://gmao.gsfc.nasa.gov/operations/filespec4.3.pdf>.]
- Avissar, R., and R. Pielke, 1989: A parameterization of heterogeneous land surfaces for atmospheric numerical models and its impact on regional meteorology. *Mon. Wea. Rev.*, **117**, 2113–2136.
- 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**, 4490, doi:10.1029/2002JD003334.
- Betts, A., F. Chen, K. Mitchell, and Z. Janjic, 1997: Assessment of the land surface and boundary layer models in two operational versions of the NCEP Eta model using FIFE data. *Mon. Wea. Rev.*, **125**, 2896–2916.
- Bonan, G. B., 1998: The land surface climatology of the NCAR Land Surface Model coupled to the NCAR Community Climate Model. *J. Climate*, **11**, 1307–1326.
- Bowling, L. C., and Coauthors, 2003: Simulation of high latitude hydrological processes in the Torne-Kalix basin: PILPS Phase 2e. 1: Experimental description and summary intercomparisons. *Global Planet. Change*, **38**, 1–30.
- Chen, F., and Coauthors, 1996: Modeling of land-surface evaporation by four schemes and comparison with FIFE observations. *J. Geophys. Res.*, **101** (D3), 7251–7268.
- Cohn, S. E., A. da Silva, J. Guo, M. Sienkiewicz, and D. Lamich, 1998: Assessing the effects of data selection with the DAO Physical-space Statistical Analysis System. *Mon. Wea. Rev.*, **126**, 2913–2926.
- Cosgrove, B. A., and Coauthors, 2003: Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. *J. Geophys. Res.*, **108**, 8842, doi:10.1029/2002JD003118.
- Dai, Y., and Q. Zeng, 1997: A land surface model (IAP94) for climate studies. Part I: Formulation and validation in off-line experiments. *Adv. Atmos. Sci.*, **14**, 443–460.
- , and Coauthors, 2003: The Common Land Model (CLM). *Bull. Amer. Meteor. Soc.*, **84**, 1013–1023.
- Derber, J. C., D. F. Parrish, and S. J. Lord, 1991: The new global operational analysis system at the National Meteorological Center. *Wea. Forecasting*, **6**, 538–547.
- 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, 69 pp.
- 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. Amer. Meteor. Soc.*, **80**, 851–878.
- Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley, 2003: Implementation of the upgraded Noah land-surface model in the NCEP operational mesoscale Eta model. *J. Geophys. Res.*, **108** (D22), 8851, doi: 10.1029/2002JD003296.
- Famiglietti, J. S., and Coauthors, 1999: Ground-based investigation of soil moisture variability within re-

- mote sensing footprints during the Southern Great Plains 1997 (SGP97) Hydrology Experiment. *Water Resour. Res.*, **35**, 1839–1851.
- Gottschalck, J., P. Houser, and X. Zeng, 2002: Impact of remotely sensed leaf area index on a global land data assimilation system. Preprints, *16th Conf. on Hydrology*, Orlando, FL, Amer. Meteor. Soc., J1.15.
- Hall, D. K., G. A. Riggs, V. V. Salomonson, N. DiGiromamo, and K. J. Bayr, 2002: MODIS snow-cover products. *Remote Sens. Environ.*, **83**, 181–194.
- Hall, F. G., B. Meeson, S. Los, L. Steyaert, E. de Colstoun, and D. Landis, Eds., 2002: *ISLSCP Initiative II*. NASA, DVD/CD-ROM. [Available online at http://islscp2.sesda.com/ISLSCP2_1/html_pages/islscp2_home.html.]
- Hamill, T. M., R. P. d'Entremont, and J. T. Bunting, 1992: A description of the Air Force real-time nephalanlysis model. *Wea. Forecasting*, **7**, 288–306.
- Hansen, M. C., R. S. DeFries, J. R. G. Townshend, and R. Sohlberg, 2000: Global land cover classification at 1km spatial resolution using a classification tree approach. *Int. J. Remote Sens.*, **21**, 1331–1364.
- Henderson-Sellers, A., A. J. Pitman, P. K. Love, P. Irannejad, and T. Chen, 1995: The Project for Intercomparison of Land Surface Parameterization Schemes (PILPS): Phases 2 and 3. *Bull. Amer. Meteor. Soc.*, **76**, 489–503.
- Huffman, G. J., R. F. Adler, E. F. Stocker, D. T. Bolvin, and E. J. Nelkin, 2003: Analysis of TRMM 3-hourly multi-satellite precipitation estimates computed in both real and post-real time. Preprints, *12th Conf. on Satellite Meteorology and Oceanography*, Long Beach, CA, Amer. Meteor. Soc., P4.11.
- Idso, S., 1981: A set of equations for the full spectrum and 8- and 14-micron and 10.5- to 12.5 thermal radiation from cloudless skies. *Water Resour. Res.*, **17**, 295–304.
- Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *Trans. ASME, Ser. D, J. Basic Eng.*, **82**, 35–45.
- Klinker, E., F. Rabier, G. Kelly, and J. F. Mahfouf, 2000: The ECMWF operational implementation of four dimensional variational assimilation. Part III: Experimental results and diagnostics with operational configuration. *Quart. J. Roy. Meteor. Soc.*, **126**, 1191–1215.
- Kopp, T. J., and R. B. Kiess, 1996: The Air Force Global Weather Central cloud analysis model. Preprints, *15th Conf. on Weather Analysis and Forecasting*, Norfolk, VA, Amer. Meteor. Soc., 220–222.
- Koren, V., J. Schaake, K. Mitchell, Q. Y. Duan, F. Chen, and J. M. Baker, 1999: A parameterization of snow-pack and frozen ground intended for NCEP weather and climate models. *J. Geophys. Res.*, **104**, 19 569–19 585.
- 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**, 2697–2715.
- , and —, 1996: Energy and water balance calculations in the Mosaic LSM. NASA Tech. Memo. 104606, Vol. 9, 76 pp.
- , —, A. Ducharne, M. Stiglitz, and P. Kumar, 2000: A catchment-based approach to modeling land surface processes in a GCM. Part 1: Model structure. *J. Geophys. Res.*, **105** (D20), 24 809–24 822.
- Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges, 1994: A simple hydrologically based model of land surface water and energy fluxes for GSMs. *J. Geophys. Res.*, **99** (D7), 14 415–14 428.
- Meng, C. J., and Coauthors, 2003: Global land surface radiation budget and its impact on water and energy cycles. Preprints, *17th Conf. on Hydrology*, Long Beach, CA, Amer. Meteor. Soc., 2.8–2.9.
- Mitchell, K., and Coauthors, 1999: GCIP Land Data Assimilation System (LDAS) project now underway. *GEWEX News*, **9** (4), 3–6.
- Myneni, R. B., R. R. Nemani, and S. W. Running, 1997: Algorithm for the estimation of global land cover, LAI and FPAR based on radiative transfer models. *IEEE Trans. Geosci. Remote Sens.*, **35**, 1380–1393.
- Ottle, C., and D. Vidalmadjar, 1992: Estimation of land surface temperature with NOAA9 data. *Remote Sens. Environ.*, **40**, 27–41.
- Owe, M., R. De Jeu, and J. P. Walker, 2001: A methodology for surface soil moisture and optical depth retrieval using the microwave polarization difference index. *IEEE Trans. Geosci. Remote Sens.*, **39**, 1643–1654.
- Persson, A., 2001: User guide to ECMWF forecast products. Meteorological Bull. M3.2, ECMWF, Reading, United Kingdom, 115 pp.
- Pfaendtner, J., S. Bloom, D. Lamich, M. Seablom, M. Sienkiewicz, J. Stobie, and A. da Silva, 1995: Documentation of the Goddard Earth Observing System (GEOS) Data Assimilation System—Version 1. NASA Tech. Memo. 104606, Vol. 4, 44 pp.
- Radakovich, J. D., P. R. Houser, A. da Silva, and M. G. Bosilovich, 2001: Results from global land-surface data assimilation methods. Preprints, *Fifth Symp. on Integrated Observing Systems*, Albuquerque, NM, Amer. Meteor. Soc., 132–134.
- Reynolds, C. A., T. J. Jackson, and W. J. Rawls, 2000: Estimating soil water-holding capacities by linking the Food and Agriculture Organization soil map of the world with global pedon databases and continuous

- pedotransfer functions. *Water Resour. Res.*, **36**, 3653–3662.
- 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. Amer. Meteor. Soc.*, **81**, 1281–1299.
- Rodell, M., and J. S. Famiglietti, 1999: Detectability of variations in continental water storage from satellite observations of the time dependent gravity field. *Water Resour. Res.*, **35**, 2705–2723.
- , P. R. Houser, U. Jambor, J. Gottschalck, C.-J. Meng, K. Arsenault, N. DiGirolamo, and D. Hall, 2002: Use of MODIS-derived snow fields in the Global Land Data Assimilation System. *GAPP Mississippi River Climate and Hydrology Conf.*, New Orleans, LA, GAPP and AMS, 118.
- Sellers, P. J., Y. Mintz, and A. Dalcher, 1986: A simple biosphere model (SiB) for use within general circulation models. *J. Atmos. Sci.*, **43**, 505–531.
- Shapiro, R., 1987: A simple model for the calculation of the flux of direct and diffuse solar radiation through the atmosphere. Air Force Geophysics Lab, AFGL-TR-87-0200, Hanscom AFB, MA, 40 pp.
- Toll, D. L., J. Entin, and P. Houser, 2001: Land surface heterogeneity on surface energy and water fluxes. *SPIE 2001, Eighth Int. Symp. on Remote Sensing*, Vol. 452, Toulouse, France, SPIE, 267–270.
- Turk, F. J., G. Rohaly, J. D. Hawkins, E. A. Smith, A. Grose, F. S. Marzano, A. Mugnai, and V. Levizzani, 2000: Analysis and assimilation of rainfall from blended SSM/I, TRMM and geostationary satellite data. Preprints, *10th Conf. on Satellite Meteorology and Oceanography*, Long Beach, CA, Amer. Meteor. Soc., 66–69.
- Verdin, K. L., and S. K. Greenlee, 1996: Development of continental scale digital elevation models and extraction of hydrographic features. *Proc. Third Int. Conf./Workshop on Integrating GIS and Environmental Modeling*, Santa Fe, NM, National Center for Geographic Information and Analysis, CD-ROM, 8.2.
- Vukovich, F. M., R. Wayland, and D. L. Toll, 1997: The surface heat flux as a function of ground cover for climate models. *Mon. Wea. Rev.*, **125**, 572–586.
- 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** (D11), 11 761–11 774.
- Zhan, X., J. K. Entin, P. R. Houser, J. P. Walker, and R. H. Reichle, 2002: Application of Kalman filtering for soil moisture data validation with NASA's Land Data Assimilation System. *Amer. Geophys. Union, Eos, Trans.* **83** (Suppl.), S194.