

## Analysis of Multiple Precipitation Products and Preliminary Assessment of Their Impact on Global Land Data Assimilation System Land Surface States

JON GOTTSCHALCK AND JESSE MENG

*University of Maryland, Baltimore County, Goddard Earth Sciences and Technology Center (GEST), Baltimore, and Hydrological Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, Maryland*

MATT RODELL AND PAUL HOUSER

*Hydrological Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, Maryland*

(Manuscript received 21 June 2004, in final form 31 January 2005)

### ABSTRACT

Precipitation is arguably the most important meteorological forcing variable in land surface modeling. Many types of precipitation datasets exist (with various pros and cons) and include those from atmospheric data assimilation systems, satellites, rain gauges, ground radar, and merged products. These datasets are being evaluated in order to choose the most suitable precipitation forcing for real-time and retrospective simulations of the Global Land Data Assimilation System (GLDAS). This paper first presents results of a comparison for the period from March 2002 to February 2003. Later, GLDAS simulations 14 months in duration are analyzed to diagnose impacts on GLDAS land surface states when using the Mosaic land surface model (LSM).

A comparison of seasonal total precipitation for the continental United States (CONUS) illustrates that the Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP) has the closest agreement with a CPC rain gauge dataset for all seasons except winter. The European Centre for Medium-Range Weather Forecasts (ECMWF) model performs the best of the modeling systems. The satellite-only products [the Tropical Rainfall Measuring Mission (TRMM) Real-time Multi-satellite Precipitation Analysis and the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN)] suffer from a few deficiencies—most notably an overestimation of summertime precipitation in the central United States (200–400 mm). CMAP is the most closely correlated with daily rain gauge data for the spring, fall, and winter seasons, while the satellite-only estimates perform best in summer. GLDAS land surface states are sensitive to different precipitation forcing where percent differences in volumetric soil water content (SWC) between simulations ranged from  $-75\%$  to  $+100\%$ . The percent differences in SWC are generally  $25\%$ – $75\%$  less than the percent precipitation differences, indicating that GLDAS and specifically the Mosaic LSM act to generally “damp” precipitation differences. Areas where the percent changes are equivalent to the percent precipitation changes, however, are evident. Soil temperature spread between GLDAS runs was considerable and ranged up to  $\pm 3.0$  K with the largest impact in the western United States.

### 1. Introduction

Over the past three decades, it has become clear that the land surface exerts a significant impact on the atmospheric boundary layer via fluxes of momentum, energy, and water and therefore impacts weather and cli-

mate (Charney 1975; Charney et al. 1977; Shukla and Mintz 1982; Sud and Smith 1985; Meehl and Washington 1988). Current weather and climate forecast models can benefit from more accurate land surface states via initialization of land surface boundary conditions (Koster and Suarez 2003). The evolution of key land surface states such as skin temperature, soil moisture, and snow that determine the critical fluxes above are largely dictated by precipitation. The amount of water and/or ice in the soil impacts the energy cycle by modulating the partitioning of energy at the land surface between

---

*Corresponding author address:* Jon Gottschalck, Climate Prediction Center, NOAA/National Centers for Environmental Prediction, 5200 Auth Road, Camp Springs, MD 20746.  
E-mail: Jon.Gottschalck@noaa.gov

sensible and latent heat and also can affect the carbon cycle through control of transpiration.

There have been many efforts to produce accurate precipitation estimates for use as part of land surface process studies and for weather and climate model predictions. These include rain-gauge-only estimates, mainly model-based estimates as part of atmospheric data assimilation systems, ground-based radar estimates, satellite-only-based estimates using data from both infrared and microwave retrievals, and merged rain gauge and satellite estimates.

All of these estimates have pros and cons (Barrett and Martin 1981; Arkin and Ardanuy 1989). Gauge estimates are typically the most accurate available and are generally taken to be the "truth" as they provide a direct measure of precipitation at the surface and do not rely on model parameterizations or remote sensing methods. Rain gauges, however, can feature underestimates due to wind and evaporation as well as instrument and human errors (Neff 1977; Sevruk 1982; Legates and Willmott 1990). Model-based precipitation estimates represent synoptic-scale precipitation processes well but often err under convective regimes at the mesoscale (Sperber and Palmer 1995). In addition, many of the model-based products provide estimates at a coarser resolution than observation-based estimates as a result of computational constraints. Ground-based radar estimates provide excellent spatial and temporal resolution but suffer from error associated with elevation angle, ground clutter, virga (precipitation evaporating before reaching the ground), and anomalous propagation (Fulton et al. 1998). Satellite estimates based on infrared retrievals provide excellent temporal coverage but are dependent on cloud-top temperatures and can very often mistake high-level (cold) cloud tops for precipitation (Griffith et al. 1981; Wylie 1979; Arkin and Meisner 1987; Arkin and Xie 1994; Arkin et al. 1994). Microwave precipitation estimates are more physically based and can provide more accurate instantaneous measures of precipitation but are hampered by poor temporal sampling (mounted on polar-orbiting satellites) and complexities in the retrieval due to cloud microphysics and land surface characteristics (Wilheit et al. 1991; Spencer 1993).

Xie and Arkin (1995) compared infrared and microwave satellite estimates with gauge observations and found good spatial agreement during the warm season over the tropical Pacific Ocean but poor continental results during the cold season, especially those based on IR retrievals. Kondragunta and Gruber (1997) showed that merged satellite/gauge products were more realistic in the depiction of the annual and interannual cycles as compared to model-based assimilation systems. Jan-

owiak et al. (1998) compared the Global Precipitation Climatology Project (GPCP; Huffman et al. 1997) merged gauge and satellite precipitation estimates with the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis precipitation product and found comparable large-scale features but substantial regional differences. Moreover, Gruber et al. (2000) compared the GPCP and the Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP; Xie and Arkin 1997; Xie et al. 2003) merged gauge and satellite precipitation datasets and found that although the spatial and temporal correlations were high, significant differences did exist as a result of slightly different input data. The Third Precipitation Intercomparison Project (PIP-3; Adler et al. 2001) analyzed many products and found that the model-based precipitation estimates are poor in the Tropics but comparable over midlatitude continental areas.

The Global Land Data Assimilation System (GLDAS) 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 (Rodell et al. 2004). More details about the GLDAS system are provided in section 2. Because precipitation is one, if not the most critical, atmospheric forcing variable, a comprehensive comparison is being conducted to determine the best precipitation forcing for GLDAS real-time simulations.

Many of the precipitation-related studies cited above have evaluated datasets at coarse spatial and temporal resolution and within coupled global modeling systems. Since GLDAS runs in near-real time and at high global resolution ( $1/4^\circ$ ), it is important to analyze products of higher temporal and spatial resolution. This paper is unique in that it aims to evaluate multiple higher-resolution precipitation datasets (in both time and space) and conduct land surface model simulations offline in order to evaluate the impact on land surface states. The land modeling section of this paper focuses on the impacts and range of sensitivity for GLDAS land surface states when using different precipitation forcing as opposed to validation of the states themselves with ground truth measurements.

This paper presents the current state of work in the analysis of precipitation datasets available to GLDAS. Section 2 describes the precipitation datasets, outlines the GLDAS framework including descriptions of the Mosaic land surface model (LSM), and identifies the types of experiments that are used to diagnose the impacts on GLDAS land surface states. Section 3 provides results in two forms: 1) an offline comparison of pre-

precipitation datasets for the continental United States (CONUS) and 2) analysis of the sensitivity of land surface states to different precipitation datasets. Section 4 discusses the results and implications from this study and outlines future plans.

## 2. Methods

### a. Offline precipitation analysis

The precipitation datasets are evaluated over the CONUS domain only. Table 1 illustrates the datasets used, their relevant specifications, and important product specific details. These include rain gauge estimates from the Climate Prediction Center (CPC) (Higgins et al. 2000), model-based estimates from data assimilation systems, namely the National Aeronautics and Space Administration's (NASA) Goddard Modeling and Assimilation Office (GMAO) Goddard Earth Observing System (GEOS; Pfaendtner et al. 1995), NCEP's Global Data Assimilation System (GDAS; Derber et al. 1991), and that from the European Centre for Medium-Range Weather Forecasts (ECMWF; Persson 2001), a merged ground-based radar/rain gauge estimate [Next-Generation Weather Radar (NEXRAD); Baldwin and Mitchell 1997], satellite-derived products such as the Tropical Rainfall Measuring Mission (TRMM) Real-time Multi-satellite Precipitation Analysis from NASA's Goddard Space Flight Center (GSFC) (HUFFMAN; Huffman et al. 2003), Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) from the University of California, Irvine (Hsu et al. 1997; Hsu et al. 1999), and the CMAP rain gauge/satellite product.

The CPC rain gauge dataset is considered "truth" in the analysis described in section 3a and includes approximately 5500 stations per day from the River Forecast Center (RFC) and Climate Anomaly Database. The Higgins product utilizes a modified Cressman scheme and makes use of a number of quality control measures to produce a  $\frac{1}{4}^\circ$  CONUS product. Figure 1 illustrates the density of gauges used in developing the Higgins gauge product. It is important to note, however, that the Higgins gauge dataset does not correct for systematic bias resulting from some of the deficiencies mentioned earlier (e.g., underestimates from wind and evaporation or instrument errors). Impacts of this are mentioned later in the discussion. The satellite-derived products listed above are chosen to not only evaluate estimates using varying methods (see Table 1) but also and more importantly these products are available in real time and at high resolution—two main requirements of GLDAS.

The time period of analysis ranges from March 2002 through February 2003. Although most of the products above are available before spring 2002, the authors wished to compare time periods only when all the datasets used in this study were available. The TRMM Real-time Multi-satellite precipitation analysis (HUFFMAN) only came online in late January 2002. The purpose of this study was to evaluate most of the real-time precipitation datasets available to GLDAS on a "level playing field"—during equivalent time periods. Seasonal sums are calculated only for time periods when data files for all precipitation estimates are available. All precipitation products are continuous over the CONUS domain (e.g., no missing data). Table 2 shows the data availability for all precipitation products for the four seasons. The sums are calculated for each precipitation estimate on the native grid of the dataset then interpolated (bilinear method) to the Higgins rain gauge grid. All datasets are taken in their original resolution except CMAP, which originally is a pentad  $2.5^\circ \times 2.5^\circ$  product. To fully utilize this data in real-time GLDAS runs, the CMAP dataset is disaggregated in time and interpolated in space using GDAS data, which results in a product with specifications that are shown in Table 1. The CMAP disaggregation is accomplished through the following steps. First, pentad sums are calculated for the GDAS data. Next, using the individual 6-hourly and the pentad totals of the GDAS estimates, the ratio of the total precipitation at 6-hourly intervals is calculated. Finally, these 6-hourly ratios are applied to the CMAP pentad totals to produce 6-hourly CMAP precipitation estimates.

### b. Land surface modeling

#### 1) GLDAS

The GLDAS project is an extension of the existing and more mature North American LDAS (NLDAS) project (Mitchell et al. 2004). Quality land surface states from GLDAS are important for accurate land surface model initialization and can lead to improvements in weather and climate seasonal forecasts (Koster and Suarez 2003). Table 3 shows the available basic options of the GLDAS interface. GLDAS runs at varying spatial and temporal resolutions, drives four LSMs, and can be "forced" by both model- and observation-based atmospheric forcing (adjusted to a consistent elevation definition). GLDAS utilizes a tiling approach to represent subgrid variability based on the 1-km vegetation classification from University of Maryland (UMD) (Hansen et al. 2000) [derived from data from the Moderate Resolution Imaging Spectroradiometer

TABLE 1. Specifications and important details of the precipitation products used in this study. All the listed datasets are currently available.

| Dataset name | Dataset type | Resolution (spatial/temporal)       | Domain (lat/lon) | Source                                   | Start date                                  | Important product details   |
|--------------|--------------|-------------------------------------|------------------|--|---|---|
| Higgins      | Gauge        | 0.25° × 0.25°, daily                | CONUS            | NOAA/CPC                                 | Jan 1996                                    | Real-time version, utilizes a Cressman analysis, ~5500 stations per day, quality assurance checks.  |
| GEOS         | Model        | 1.0° × 1.25°, <sup>a</sup> 3 hourly | 90°S–90°N, 180°  | NASA GSFC                                | Jan 2000                                    | Utilizes conventional (rawinsondes, dropsondes, buoys, ship reports, surface weather obs, aircraft) and satellite data (cloud-track winds, TOVS <sup>b</sup> atmospheric soundings, QuikSCAT sea surface winds, SSM/I tropical precipitable water, solar backscatter ultraviolet instrument (SBUV) ozone, Reynolds SST).  |
| GDAS         | Model        | ~0.4°, <sup>c</sup> 6 hourly        | 90°S–90°N, 180°  | NOAA/Environmental Modeling Center (EMC) | Long history, Jan 2000 for equivalent files | Utilizes conventional (rawinsondes, dropsondes, buoys, ship reports, Coastal-Marine Automated Network (C-MAN) platforms, surface weather obs, aircraft), satellite (cloud-track winds, GOES radiances, SSM/I, and TRMM rain rate, SSM/I and QuikSCAT sea surface winds, SBUV ozone, TOVS atmospheric soundings), and ground radar data [NOAA Profiler Network wind profilers, Weather Surveillance Radar-1988 Doppler (WSR-88D) NEXRAD velocity–azimuth display (VAD) winds]. |
| ECMWF        | Model        | ~0.25°, 3 hourly                    | 90°S–90°N, 180°  | ECMWF                                    | Jan 1979                                    | Utilizes conventional (rawinsondes, dropsondes, ship reports, buoys, aircraft, surface weather obs) and satellite data (TOVS and SSM/I atmospheric soundings, Meteosat and GOES radiances, QuikSCAT sea surface winds, SBUV ozone).   |
| HUFFMAN      | Satellite    | 0.25° × 0.25°, 3 hourly             | 60°S–60°N, 180°  | NASA GSFC                                | Feb 2002                                    | Utilizes 1) the Goddard Profiling Algorithm (GPROF) (Kummerow et al. 2001) applied to SSM/I and TRMM Microwave Imager (TMI) passive microwave measurements and 2) the merged 4-km IR brightness temperature dataset from CPC (Janowiak et al. 2000) converted to IR precipitation estimates. The IR precipitation estimates are calibrated to the microwave data locally in time and space. The final product is a merging of these two estimates (3B42RT).                   |

TABLE 1. (Continued)

| Dataset name | Dataset type | Resolution (spatial/temporal)            | Domain (lat/lon)                                      | Source                           | Start date | Important product details  |
|--------------|--------------|--|---|----------------------------------|------------|--|
| PERSIANN     | Satellite    | $0.25^\circ \times 0.25^\circ$ , hourly  | $60^\circ\text{S}$ – $60^\circ\text{N}$ , $180^\circ$ | University of California, Irvine | Mar 2000   | Utilizes neural network function classification/approximation procedures. Uses the merged 4-km IR brightness temperature dataset from CPC and converts to IR precipitation estimates as the basis for the product. Passive microwave measurements from TRMM TMI are used to update neural network parameters (Sorooshian et al. 2000).   |
| CMAP         | Merged       | $\sim 0.4^\circ$ , <sup>c</sup> 6 hourly | $90^\circ\text{S}$ – $90^\circ\text{N}$ , $180^\circ$ | NOAA/CPC                         | Jan 1979   | Merged rain gauge, satellite (infrared and microwave), and model-based product. Estimate produced in two-step process: 1) The nongauge estimates are combined linearly through the maximum likelihood estimation method (linear combination coefficients are inversely proportional to the squares of the local random error of the individual sources). Gauges are used to calculate the random error: 2) the gauge data is blended with output from the first step based on Reynolds (1988). |
| NEXRAD       | Merged       | 4 km, hourly                             | CONUS   | NOAA/NCEP                        | May 1996   | Merged ground radar, rain gauge product using optimal estimation theory (Seo 1998a,b).   |

<sup>a</sup> GEOS4 resolution (spring/summer comparisons use GEOS3,  $1.0^\circ \times 1.0^\circ$  spatial resolution).

<sup>b</sup> TOVS: Television Infrared Observation Satellite (TIROS) Operational Vertical Sounder.

<sup>c</sup> GDAS2 resolution (spring/summer comparisons use GDAS1,  $\sim 0.7^\circ$  spatial resolution).

(MODIS)]. Soil information is based on a 5-min resolution soil database described by Reynolds et al. (2000) and includes porosity and the percentage of sand, silt, and clay. Leaf area index (LAI) data from MODIS

[made available from the Department of Geography at Boston University (BU)], is also incorporated in this study (Myneni et al. 2002). This paper illustrates results using the Mosaic LSM (Koster and Suarez 1996).

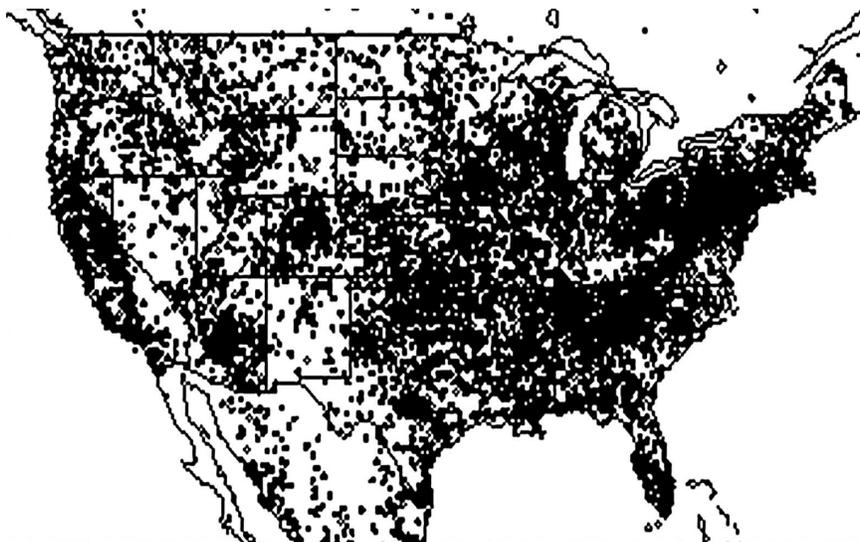


FIG. 1. An illustration of the typical density of the rain gauge network used in creating the Higgins daily precipitation product.

TABLE 2. The percentage of data available for all precipitation estimates for each season.

| Dataset  | Mar–Apr–<br>May | Jun–Jul–<br>Aug | Sep–Oct–<br>Nov | Dec–Jan–<br>Feb |
|----------|-----------------|-----------------|-----------------|-----------------|
| Higgins  | 98.9            | 100.0           | 100.0           | 100.0           |
| GEOS     | 100.0           | 100.0           | 100.0           | 100.0           |
| GDAS     | 100.0           | 100.0           | 100.0           | 100.0           |
| ECMWF    | 98.9            | 100.0           | 100.0           | 100.0           |
| HUFFMAN  | 98.9            | 97.0            | 93.0            | 97.9            |
| PERSIANN | 100.0           | 99.3            | 100.0           | 99.5            |
| CMAP     | 100.0           | 100.0           | 100.0           | 100.0           |
| NEXRAD   | 100.0           | 100.0           | 100.0           | 100.0           |

## 2) LAND SURFACE MODEL

Mosaic is a well-established and theoretically sound LSM with roots in the Simple Biosphere model (SiB) of Sellers et al. (1986) (Koster and Suarez 1996). The model allows explicit vegetation control over the computed surface energy and water balances with environmental stresses acting to increase canopy resistance and thus decrease transpiration. The scheme includes a canopy interception reservoir and three soil reservoirs: a thin layer near the surface, a middle layer that encompasses the remainder of the root zone and a lower recharge layer for long term storage. Bare soil evaporation, transpiration, and interception loss occur in parallel, and runoff occurs both as overland flow during precipitation events and as groundwater drainage. A complete snow budget is also included.

### c. Simulation strategy

All GLDAS runs are global in domain, feature a  $1/4^\circ$  spatial resolution, use a 15-min time step, cover a time period from January 2002 through February 2003, and

TABLE 3. Basic options available in the GLDAS user interface.

|                      |   |
|----------------------|---|
| Spatial resolution   | 0.25°; 0.5°; 1.0°; $2.0^\circ \times 2.5^\circ$   |
| Temporal resolution  | Adjustable model time step and output interval  |
| Land surface model   | Mosaic; Community Land Model version 2 (CLM2); Noah; Variable Infiltration Capacity (VIC) |
| Forcing              | Various model and satellite-derived 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. (2000)  |
| Leaf area index      | Lookup table; Advanced Very High Resolution Radiometer (AVHRR)-derived; MODIS-derived     |

use GEOS baseline atmospheric forcing. GEOS operational forecast model output uses the Physical-Space Statistical Analysis System (Cohn et al. 1998) to assimilate data and utilizes observations from several platforms (see Table 1). GEOS 3-hourly fields are produced on a  $1^\circ \times 1.25^\circ$  global grid, and GLDAS uses incoming shortwave and longwave radiation, 2-m temperature and specific humidity, 10-m zonal and meridional wind, total and convective precipitation, and surface pressure from GEOS as atmospheric forcing. To initialize the LSM, GLDAS is integrated for 10 yr at a  $2.0^\circ \times 2.5^\circ$  resolution using atmospheric forcing from the year 2001 while running the Mosaic LSM. The land surface state restart files are then interpolated to  $1/4^\circ$  and each simulation is continued beginning on 1 January 2002. Five GLDAS runs are conducted:

- 1) GEOS precipitation replaced with Higgins precipitation (Higgins)
- 2) GEOS precipitation (GEOS)
- 3) GEOS precipitation replaced with PERSIANN precipitation (PERSIANN)
- 4) GEOS precipitation replaced with HUFFMAN precipitation (HUFFMAN)
- 5) GEOS precipitation replaced with CMAP precipitation (CMAP)

The rationale for these simulations is to demonstrate the impact and *range* of sensitivity on land surface states for distinct types of precipitation estimates—a model-based dataset, two satellite-only datasets, and a merged gauge/satellite dataset. The GEOS precipitation product was chosen as the model-based dataset because it showed some of the largest errors for CONUS when compared to the Higgins gauge dataset. Consequently, it was decided that it would be best in order to study the range of impacts on land surface states. For clarity and in order to directly link to the offline CONUS comparisons, results are only shown for the CONUS. The precipitation forcing is interpolated to the GLDAS grid (budget-bilinear method) and replaces the GEOS baseline precipitation.

## 3. Results

### a. CONUS precipitation analysis

#### 1) MARCH–MAY 2002

Figure 2 illustrates seasonal total precipitation (shading) for spring for all precipitation estimates. The absolute differences between each product and Higgins are shown by the contours. Figures 2i and 2j illustrate the 1996–2003 average seasonal total precipitation and standard deviation for reference in order to understand

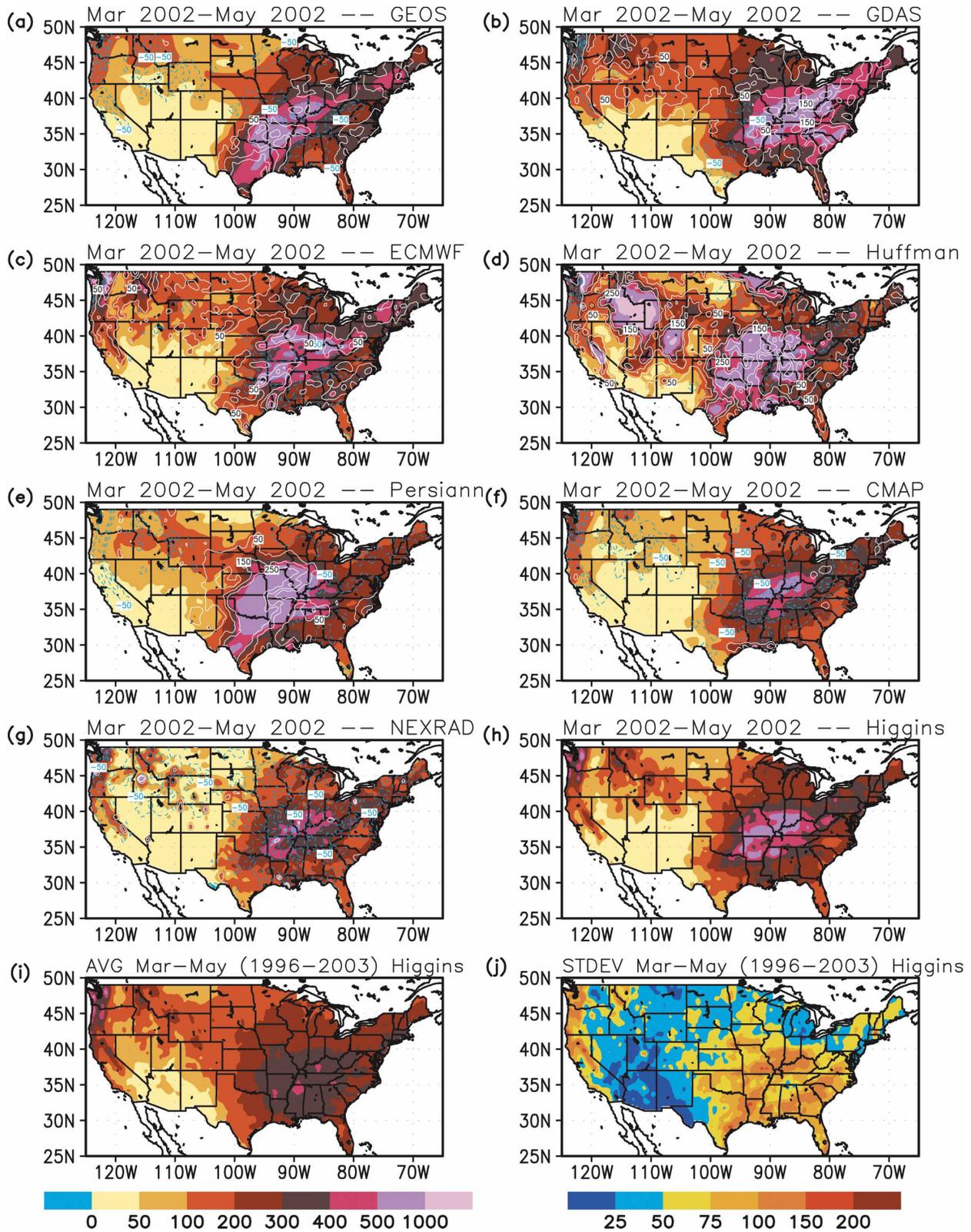


FIG. 2. CONUS total precipitation (mm) for Mar, Apr, and May 2002 for (a) GEOS, (b) GDAS, (c) ECMWF, (d) HUFFMAN, (e) PERSIANN, (f) CMAP, (g) NEXRAD, and (h) Higgins gauge. Contours [white (+), blue (-)] depict the differences with Higgins (i.e., GEOS-Higgins). The Higgins 8-yr seasonal average and standard deviation are shown in (i) and (j).

1) where this season ranks compared to a longer-term average (wet or dry season) and 2) how the differences compare to normal interannual variability (over the 8-yr period).

The Higgins gauge data (Fig. 2h) shows a few main features. These include a maximum area of precipitation (500–1000 mm) in the central Mississippi and Ohio River valleys, a dry region in the southwest (0–50 mm), and another maximum in precipitation along the northwest coast (up to 500–1000 mm). All products, in one form or another, show this maximum area. The CMAP, ECMWF, and NEXRAD estimates capture the location, spatial variability, and magnitude the best while the GEOS and PERSIANN datasets show a more widespread maximum farther west and southwest. The GDAS estimate elongates this area farther to the northeast. Across the southwest, the model products perform best with estimates generally ranging from 0 to 50 mm. There are large differences in the products across the Rockies and interior west. Of the models, ECMWF captures the location and spatial variability of the overall pattern the best. The PERSIANN product represents the area reasonably well, while the HUFFMAN accumulation is very high (more than 500 mm greater than Higgins) across parts of the Rockies. It is important to note that the CMAP product does not capture the detail in the mountainous west because of the low original resolution of the CMAP data, relatively coarse resolution of GDAS data used in the temporal disaggregation, and also due to the low number of gauges in this area. Along the West Coast, the ECMWF, NEXRAD, and HUFFMAN estimates perform the best while the other products underestimate the precipitation, especially PERSIANN and GEOS where differences from Higgins range from 2–5–300 mm and 25–200 mm, respectively.

The NEXRAD data shows the highest correlation (Fig. 3) with the Higgins data especially across the eastern 2/3 of the United States and along the West Coast where the correlation is consistently above 0.9. The correlation of the CMAP product is high, and in the absence of the NEXRAD product it performs the best. Also, it is evident that the model products offer greater temporal agreement than the satellite-only estimates. GDAS correlation values are above 0.7 for much of the eastern United States and along the West Coast. The PERSIANN and HUFFMAN satellite-only products demonstrate lower values across the interior west and southwest (0.0–0.5).

## 2) JUNE–AUGUST 2002

During the summer there are large differences between datasets when compared to the gauge data.

Figure 4 shows that the Higgins estimates are generally 200–400 mm across the eastern 2/3 of the United States with maximums in the upper Midwest and along the immediate Gulf Coast and Florida ranging from 500 to 1000 mm. The NEXRAD, CMAP, and ECMWF estimates offer the best agreement in these areas with the ECMWF showing the greatest accumulation. The GEOS and GDAS model products do show similar patterns of precipitation but produce accumulations significantly larger (500–1000 mm and greater) and more widespread, especially in the southeast United States, than the Higgins gauge data. This finding is consistent with errors typical of model-based precipitation under convective regimes (e.g., Sperber and Palmer 1995) such as that found during this time of year across the southeast. PERSIANN and HUFFMAN show much greater precipitation across the plains and upper Midwest (greater by 300 mm). A plausible explanation for this positive bias is mentioned in section 4. This finding is at odds with Xie and Arkin (1995), which documented good agreement of these types of products with gauge observations during the warm season.

Correlation of daily precipitation (Fig. 5) during summer is not as high as spring. The more convective (and so variable) nature of precipitation during the summer months does not make this a surprise. Once again the NEXRAD estimate offers the best correlation, with values ranging greater than 0.9 over most of the eastern 2/3 of the United States and upper West Coast. The satellite estimates show better correlation (0.6–0.8) than the model products especially across the plains and Midwest—the same region for which these products showed the large precipitation accumulation described above. Xie and Arkin (1995) also reported high warm season correlations for midlatitude land areas (0.6–0.8). The correlation values for model estimates are much lower ranging from 0.1 to 0.5 across large areas of this region, with GDAS illustrating the best correlation of the three products. This difference between the satellite and model-based products is most likely a result of the more convective nature of precipitation during this time of the year. It is clear from this figure that the satellite products seem to better diagnose the timing of organized precipitation events across the central United States but tend to overestimate their magnitude.

## 3) SEPTEMBER–NOVEMBER 2002

The difference in seasonal precipitation is the least during the fall season (Fig. 6) and generally range below  $\pm 100$  mm. The Higgins gauge data shows generally 300–400 mm of precipitation across the eastern 1/3 of the United States with a maximum along the Gulf

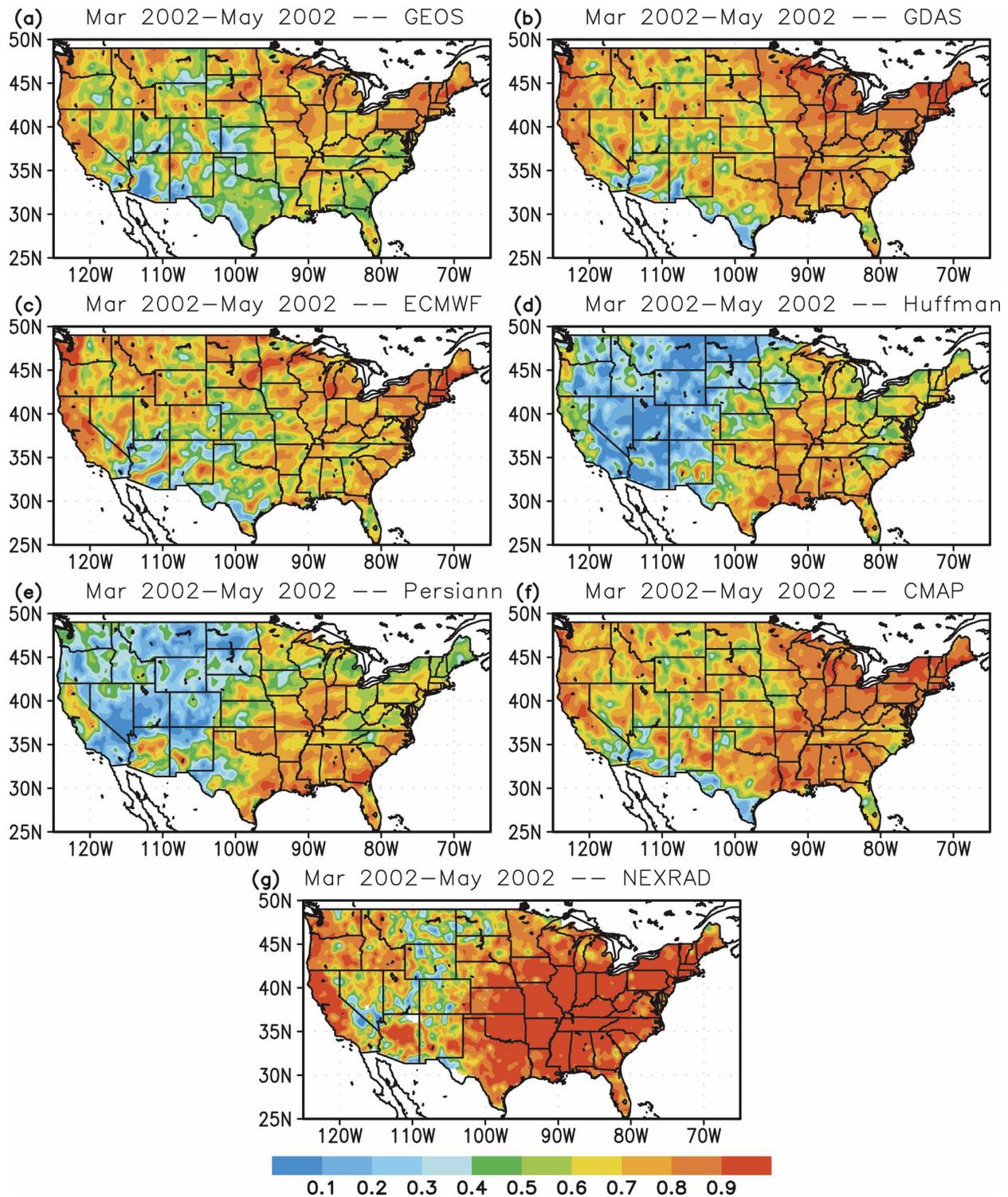


FIG. 3. CONUS correlation of daily precipitation with Higgins gauge data for Mar, Apr, and May 2002 for (a) GEOS, (b) GDAS, (c) ECMWF, (d) HUFFMAN, (e) PERSIANN, (f) CMAP, and (g) NEXRAD.

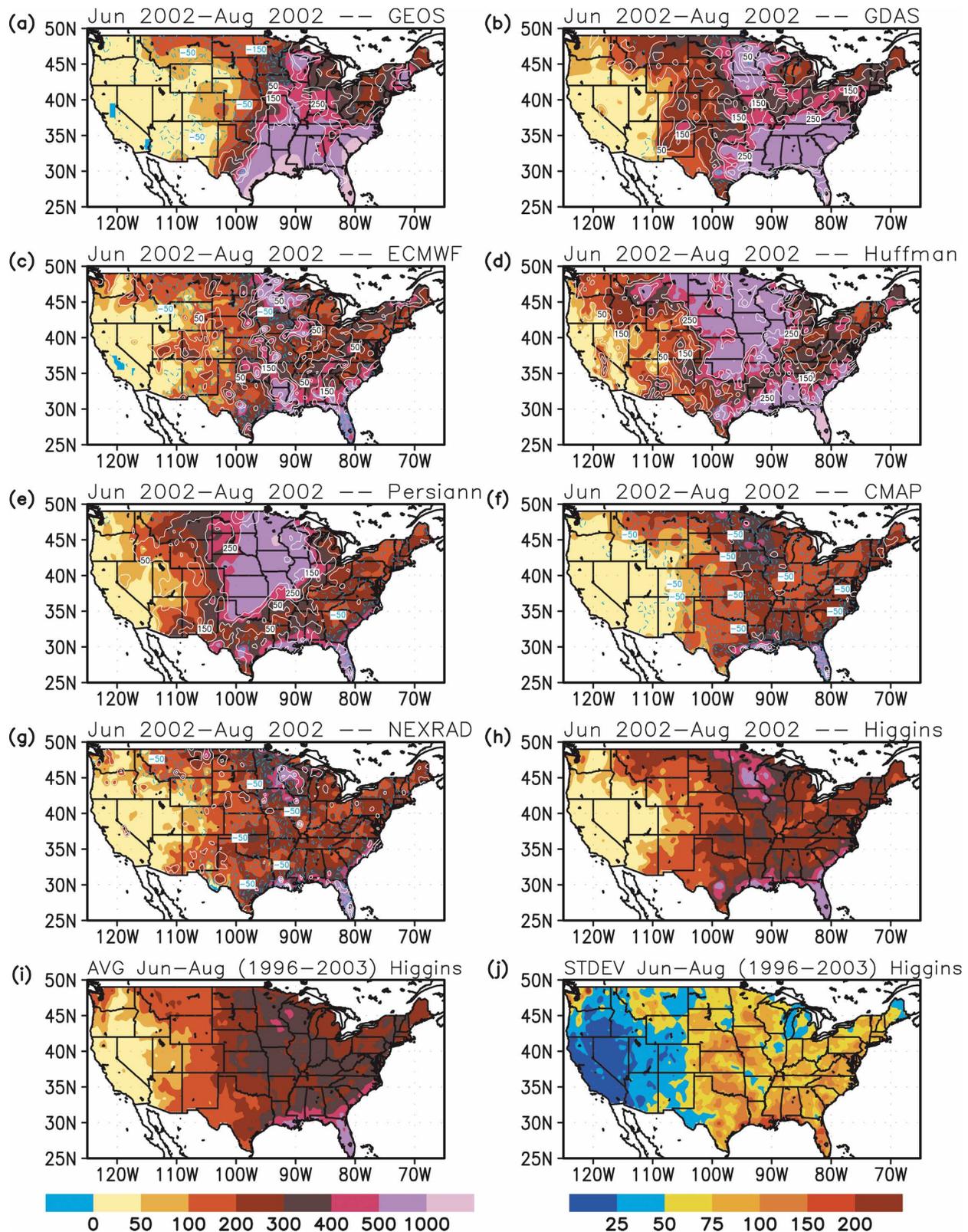


FIG. 4. CONUS total precipitation (mm) for Jun, Jul, and Aug 2002 for (a) GEOS, (b) GDAS, (c) ECMWF, (d) HUFFMAN, (e) PERSIANN, (f) CMAP, (g) NEXRAD, and (h) Higgins gauge. Contours [white (+), blue (-)] depict the differences with Higgins (i.e., GEOS-Higgins). The Higgins 8-yr seasonal average and standard deviation are shown in (i) and (j).

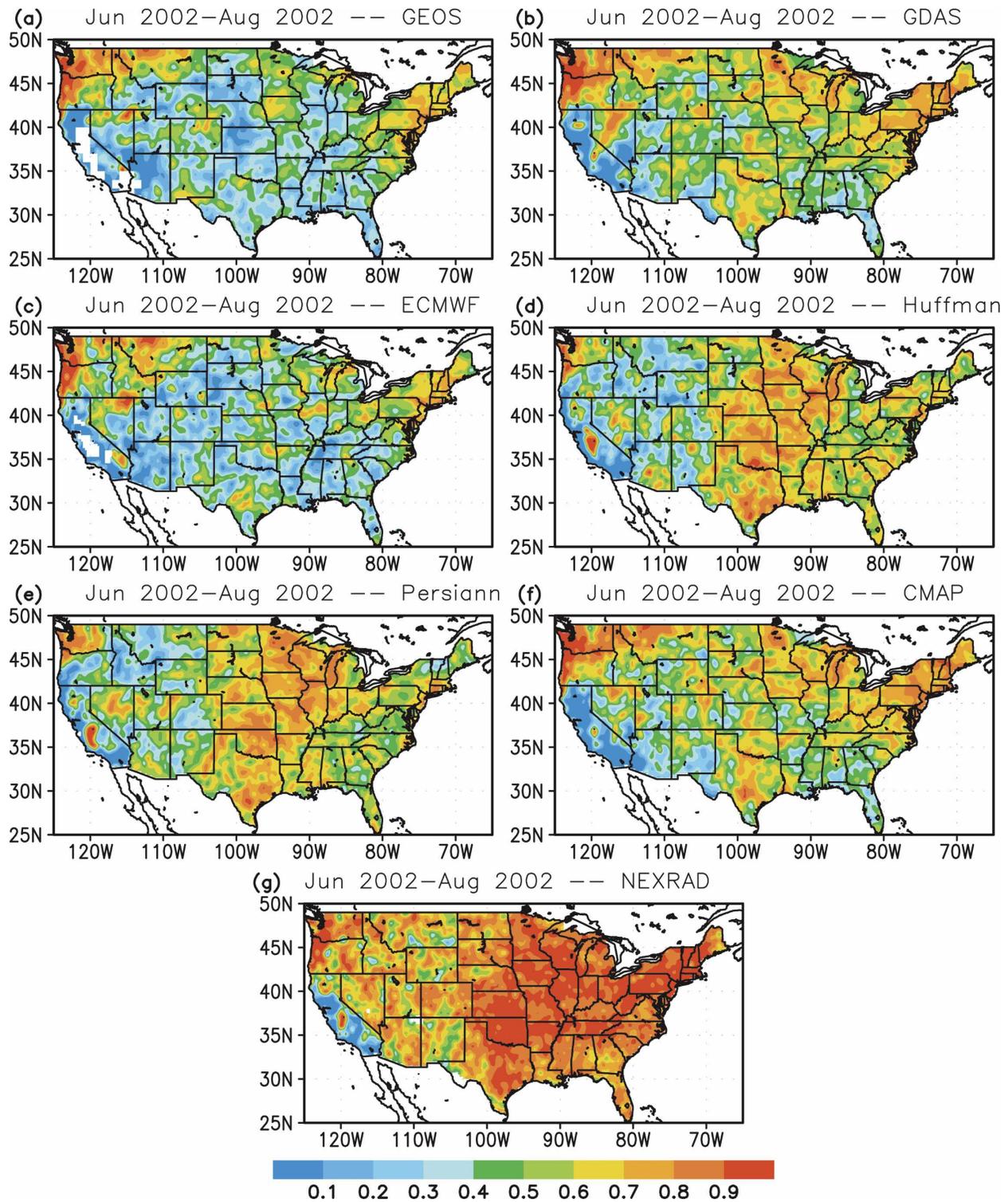


FIG. 5. CONUS correlation of daily precipitation with daily Higgins gauge data for Jun, Jul, and Aug 2002 for (a) GEOS, (b) GDAS, (c) ECMWF, (d) HUFFMAN, (e) PERSIANN, (f) CMAP, and (g) NEXRAD.

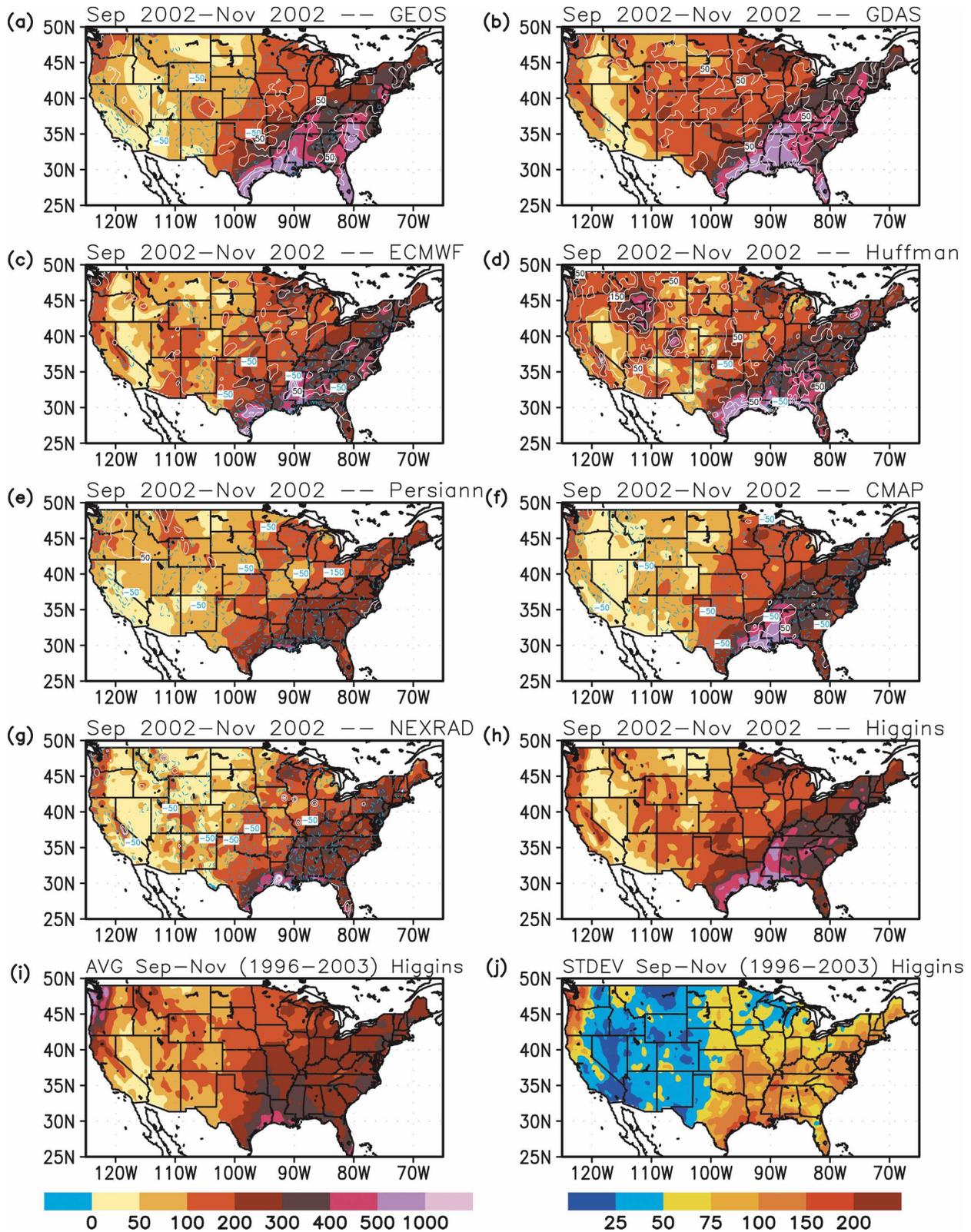


FIG. 6. CONUS total precipitation (mm) for Sep, Oct, and Nov 2002 for (a) GEOS, (b) GDAS, (c) ECMWF, (d) HUFFMAN, (e) PERSIANN, (f) CMAP, (g) NEXRAD, and (h) Higgins gauge. Contours [white (+), blue (-)] depict the differences with Higgins (i.e., GEOS-Higgins). The Higgins 8-yr seasonal average and standard deviation are shown in (i) and (j).

Coast ranging from 500 to 1000 mm. There is a secondary maximum (300–400 mm) across the upper Mississippi Valley. All of the products show these main features in a similar way with the CMAP, ECMWF, and HUFFMAN estimates (other than the interior west) the closest to the gauge data. GEOS and GDAS indicate greater accumulations along the Gulf Coast with values ranging from 500 to 1000 mm across some of the region. GDAS also produces greater and more uniform precipitation across the interior west while the ECMWF data represent the West Coast the best of the model products. The PERSIANN satellite estimate does not represent the West Coast maximum well where differences range from  $-25$  to 100 mm.

The correlation of daily precipitation (Fig. 7) during the fall is mixed. NEXRAD and CMAP show the best correlation with the Higgins daily precipitation product with values consistently ranging above 0.9 mainly across the eastern United States and along the West Coast. Of the model products, GDAS and ECMWF show a greater correlation especially across the northeast and West Coast. The two satellite products are very similar in their values, which generally range from 0.5 to 0.9 in the east but are lower in the west.

#### 4) DECEMBER 2002–FEBRUARY 2003

Figure 8 illustrates the wintertime total precipitation for the Higgins gauge dataset and shows three interesting features. These include a southwest-to-northeast pattern of precipitation from the Gulf Coast into the northeast (300–500 mm), a dry area in the upper Midwest (less than 50 mm), and a high accumulation along the northwest coast (500–1000 mm). The ECMWF, CMAP, NEXRAD, and GDAS datasets do very well with these three features although the CMAP accumulation is lower across the northwest coast. Both the ECMWF and GDAS products accurately represent the West Coast and eastern areas but overestimate the precipitation across the central United States (25–50 mm). The PERSIANN and GEOS datasets do not depict the northwest coast maximum well (underestimates of up to  $-200$ –300 mm) but are comparable to the gauge data in the other areas of the United States. On the other hand, the other satellite-only product, the HUFFMAN dataset, accurately captures the West Coast maximum. A substantial positive bias is evident in the HUFFMAN product across areas of the interior west and northern plains and is discussed later in section 4.

The correlation of daily precipitation (Fig. 9) illustrates clear differences between the satellite-only precipitation estimates and the other products. The NEXRAD, CMAP, ECMWF, GDAS, and GEOS correlations values are comparable, with the highest in the

east and along the West Coast (0.7–0.9). The GEOS values are the lowest of the five products. The PERSIANN and HUFFMAN correlation ranges from 0.5 to 0.8 across the southeast United States, which is similar to the other products although slightly less. The correlation, however, is substantially lower (0.0–0.5) for both of these estimates across the upper Midwest, the Rockies, and the interior west. These results are consistent with the findings of Xie and Arkin (1995) which showed correlation values generally ranging from 0.0 to 0.5 for the cold season when comparing the Geostationary Operational Environmental Satellite (GOES) precipitation index (GPI) infrared product (Arkin and Meisner 1987) and a microwave product based on Special Sensor Microwave Imager (SSM/I) data (Grody 1991) with the Global Precipitation Climatology Center (Schneider et al. 1993) gauge dataset.

#### 5) CONUS SUMMARY

Table 4 summarizes the results for the entire CONUS domain and lists both the spatial root-mean-square error (rmse) of seasonal totals and mean correlation of daily precipitation between each product and Higgins. The table shows that the CMAP product indicates the least error (excluding NEXRAD) over all times of the year except during the winter months. The ECMWF precipitation demonstrates the least error of the model products. The satellite products indicate the largest errors during all seasons. The seasonal precipitation totals and subsequent error statistics summarized above match well the findings of PIP-3, which compared 31 different model, satellite, merged, and climatological precipitation products over a 12-month period for 1992. Midlatitude land results showed the merged satellite/gauge products had the lowest errors (10–20 mm), followed by the model-based products (25–50 mm), with the ECMWF model indicating the lowest error (31 mm). The satellite product errors were greater and generally ranged from 25 to 75 mm.

The CMAP product (excluding NEXRAD) shows the highest correlation throughout the year except during the summer months where it is comparable to the satellite-only products. During the spring and winter, respectively, the model products showed high correlation while the satellite indicated the lowest correlation. The satellite products, however, indicate the highest correlation during the summer months. The encouraging results shown in this study for the CMAP product are consistent with improvement of precipitation estimates when merged with gauge data as part of GPCP (Krajewski et al. 2000). Moreover, correlation for merged precipitation estimates were shown to be extremely high (0.9–1.0) as part of PIP-3. Model-based

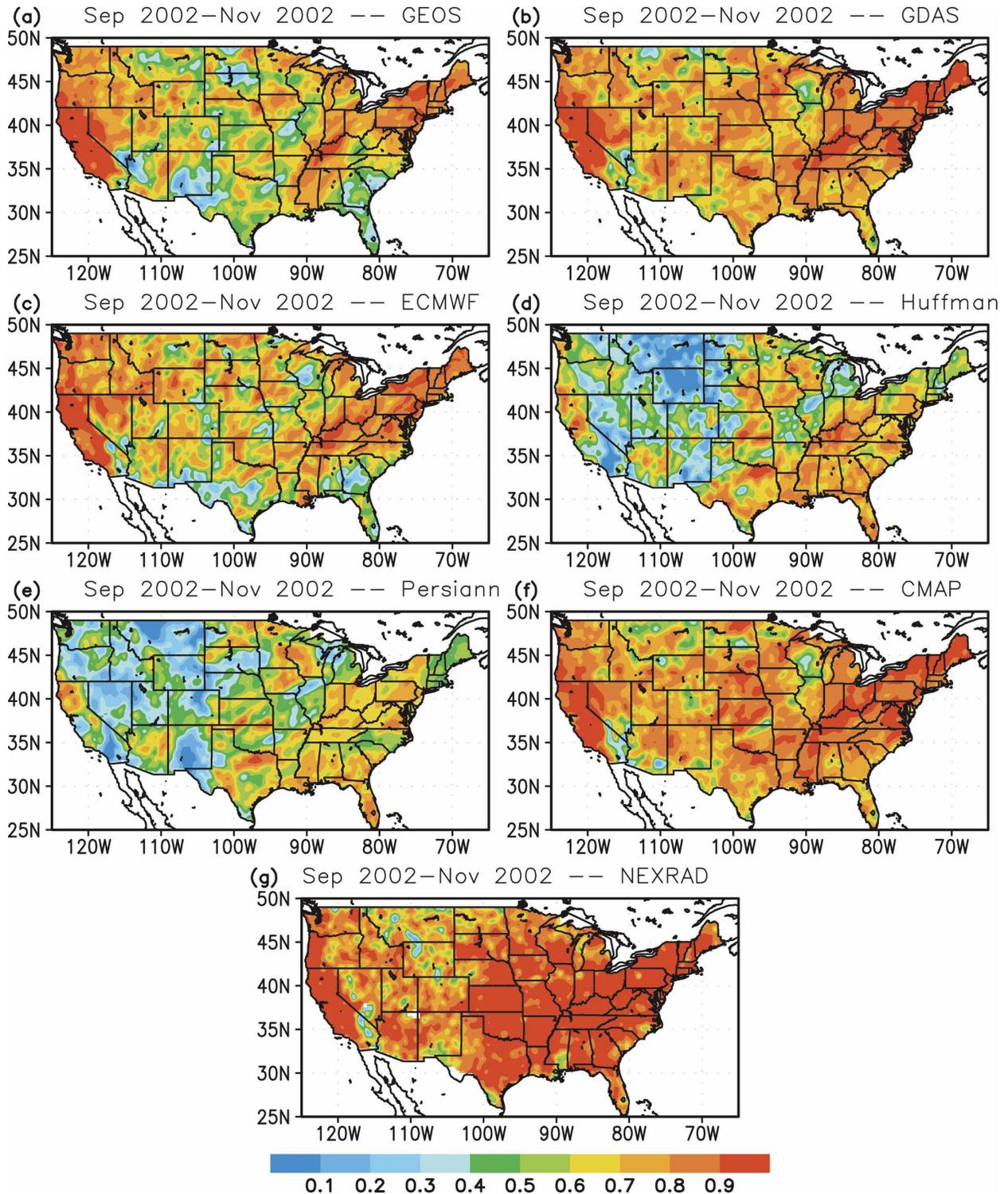


FIG. 7. CONUS correlation of daily precipitation with daily Higgins gauge data for Sep, Oct, and Nov 2002 for (a) GEOS, (b) GDAS, (c) ECMWF, (d) HUFFMAN, (e) PERSIANN, (f) CMAP, and (g) NEXRAD.

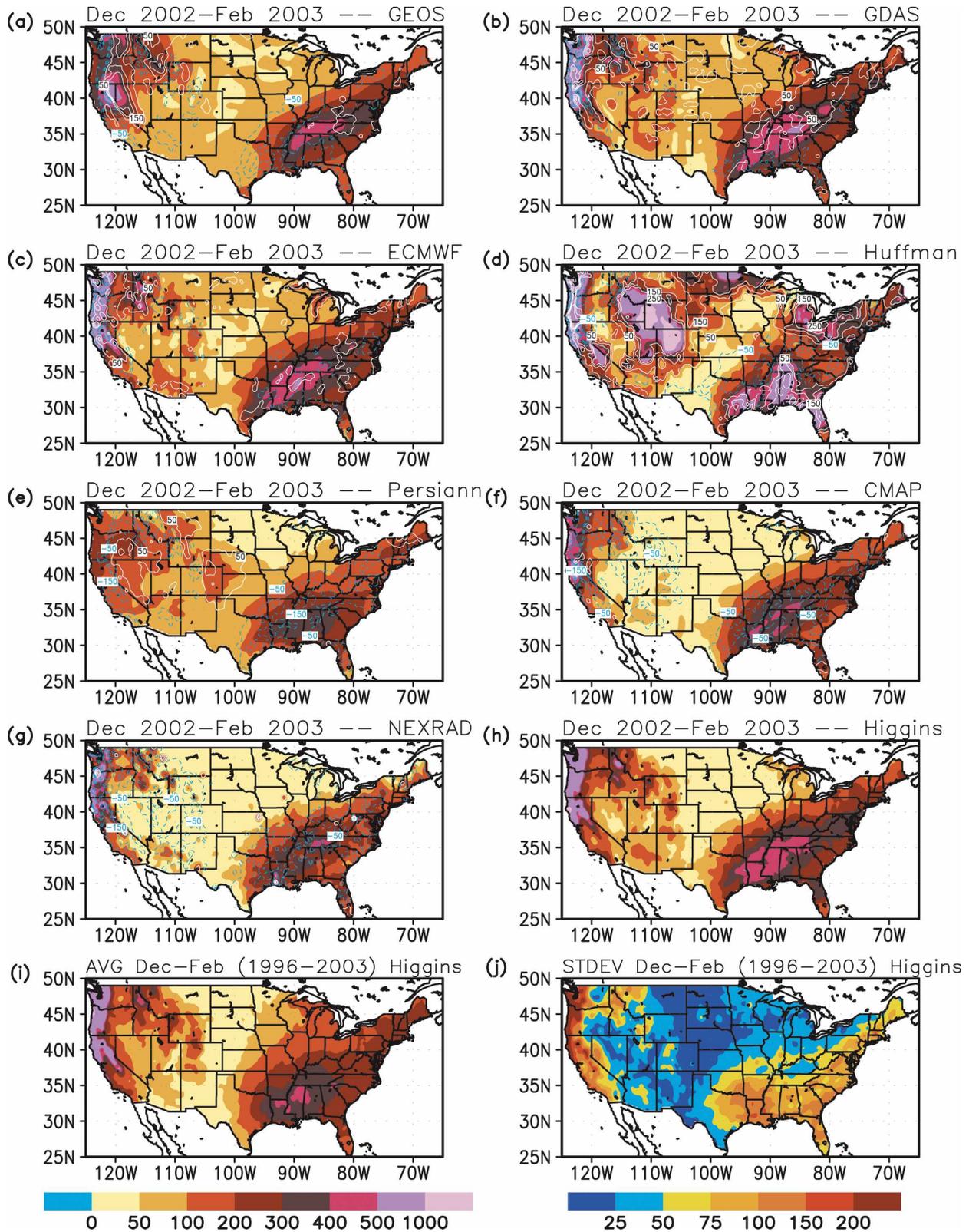


FIG. 8. CONUS total precipitation (mm) for Dec 2002, Jan 2003, and Feb 2003 over CONUS for (a) GEOS, (b) GDAS, (c) ECMWF, (d) HUFFMAN, (e) PERSIANN, (f) CMAP, (g) NEXRAD, and (h) Higgins gauge. Contours [white (+), blue (-)] depict the differences with Higgins (i.e., GEOS-Higgins). The Higgins 8-yr seasonal average and standard deviation are shown in (i) and (j).

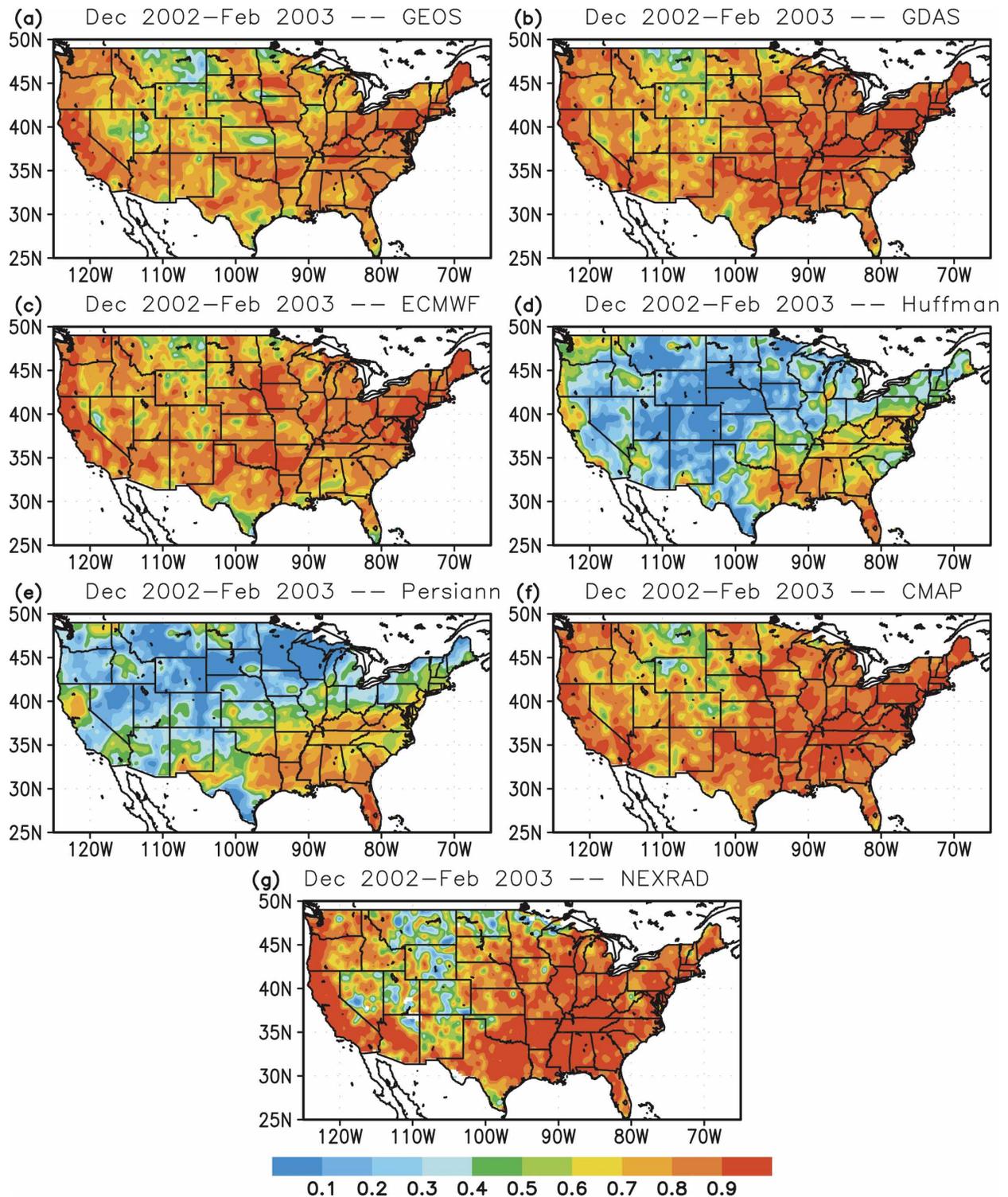


FIG. 9. CONUS correlation of daily precipitation with daily Higgins gauge data for Dec 2002, Jan 2003, and Feb 2003 for (a) GEOS, (b) GDAS, (c) ECMWF, (d) HUFFMAN, (e) PERSIANN, (f) CMAP, and (g) NEXRAD.

TABLE 4. Spatial root-mean-square error of seasonal total precipitation (mm) and mean correlation of daily precipitation between each product and Higgins for the entire CONUS domain. The lowest error and highest correlation for each season is highlighted in bold (not including the NEXRAD product). MAM: Mar–Apr–May; JJA: Jun–Jul–Aug; SON: Sep–Oct–Nov; DJF: Dec–Jan–Feb.

| Dataset  | CONUS rms error (mm) |           |           |           | CONUS correlation |             |             |             |
|----------|----------------------|-----------|-----------|-----------|-------------------|-------------|-------------|-------------|
|          | MAM                  | JJA       | SON       | DJF       | MAM               | JJA         | SON         | DJF         |
| GEOS     | 94                   | 245       | 88        | 115       | 0.62              | 0.41        | 0.65        | 0.76        |
| GDAS     | 95                   | 168       | 78        | 66        | <b>0.72</b>       | 0.53        | 0.77        | <b>0.82</b> |
| ECMWF    | 69                   | 108       | 66        | <b>56</b> | 0.66              | 0.41        | 0.68        | 0.80        |
| PERSIANN | 132                  | 189       | 90        | 131       | 0.51              | <b>0.56</b> | 0.51        | 0.40        |
| HUFFMAN  | 294                  | 222       | 94        | 286       | 0.47              | 0.54        | 0.53        | 0.36        |
| CMAP     | <b>69</b>            | <b>75</b> | <b>60</b> | 90        | <b>0.72</b>       | 0.55        | <b>0.78</b> | <b>0.82</b> |
| NEXRAD   | 74                   | 67        | 73        | 101       | 0.79              | 0.77        | 0.85        | 0.78        |

products indicated slightly lower correlation (0.7–0.8) with the ECMWF, indicating the best correlation while the satellite products illustrated a wider range of values from 0.4 to 0.6.

#### b. GLDAS simulations

The results in this section focus on some important GLDAS land surface states such as top 1-m volumetric soil water content (SWC), soil temperature ( $T_{\text{soil}}$ ) (a temperature in the deep soil unaffected by diurnal variations), and snow water equivalent (SWE). These are integrated quantities and represent important variables for land surface model process and initialization studies. Monthly averaged outputs from the five simulations outlined in section 2c are presented for August 2002 and February 2003 for CONUS.

Figure 10a illustrates average SWC (expressed as a percentage) from the Higgins simulation (top row) and the percent differences (shading) between Higgins and the GEOS, PERSIANN, HUFFMAN, and CMAP runs for August 2002 (next four rows). The contours show the percent differences in seasonal precipitation between each precipitation product and Higgins taken from results shown in section 3a. The Higgins simulation shows that the greatest absolute SWC at the end of the summer is in Florida and along sections of the Gulf Coast (near 40%). Widespread moderate SWC is evident across much of the central United States and generally ranges from 16% to 32% while low SWC (<12%) exists across most of the western United States.

The GEOS, PERSIANN, and HUFFMAN simulations all show substantial differences (shaded areas) from the observed precipitation (Higgins run). The pattern of differences, however, varies considerably. For instance, the GEOS simulation depicts a drier land surface across the west (generally 10%–50% less) but a much more wet southern and eastern United States (greater than 100% in some areas). The PERSIANN and HUFFMAN runs indicate generally a much more

moist land surface across the central United States and most of the west with the exception of the immediate northwest coast. This pattern is consistent with the overestimation of summertime precipitation by the satellite products over the central United States (Fig. 4). Also, these runs show greater SWC across the west (especially HUFFMAN) than does the Higgins simulation. The HUFFMAN SWC is high across the Rockies and sections of the northern plains because of leftover soil moisture as a result of the large overestimation of precipitation during the spring (Fig. 2). The precipitation during this time, although frozen initially, later melts during the late spring and summer. The CMAP simulation, on the other hand, illustrates substantially less error from the Higgins simulation where percent errors do not exceed 5%–20% except along the immediate northwest coast, and this is consistent with the offline precipitation comparison findings in sections 3a(1) and 3a(2).

Soil moisture plays a major role in evapotranspiration and in turn substantially impacts both surface and deeper layer soil temperatures through modifications in the surface energy budget. Consequently, it is important to diagnose the magnitude of changes in  $T_{\text{soil}}$  when using different precipitation forcing. Figure 10b illustrates  $T_{\text{soil}}$  output from the five GLDAS runs in an equivalent format as Fig. 10a (except differences are in absolute terms) and shows that the greatest differences between the simulations are located in the western United States. After the summer season, the GEOS and CMAP runs indicate a warmer land surface, while the PERSIANN and HUFFMAN runs indicate a cooler land surface. In both cases the differences are large and approach  $\pm 3.0$  K in some areas. In the eastern United States, the differences from the Higgins land surface are smaller for all simulations—generally less than  $\pm 0.5$  K.

Figure 11a illustrates how SWC evolves by the end of February 2003. At this time, Higgins shows a more moist eastern CONUS (up to 36%–40%) and West Coast (>40%) while the northern plains was generally

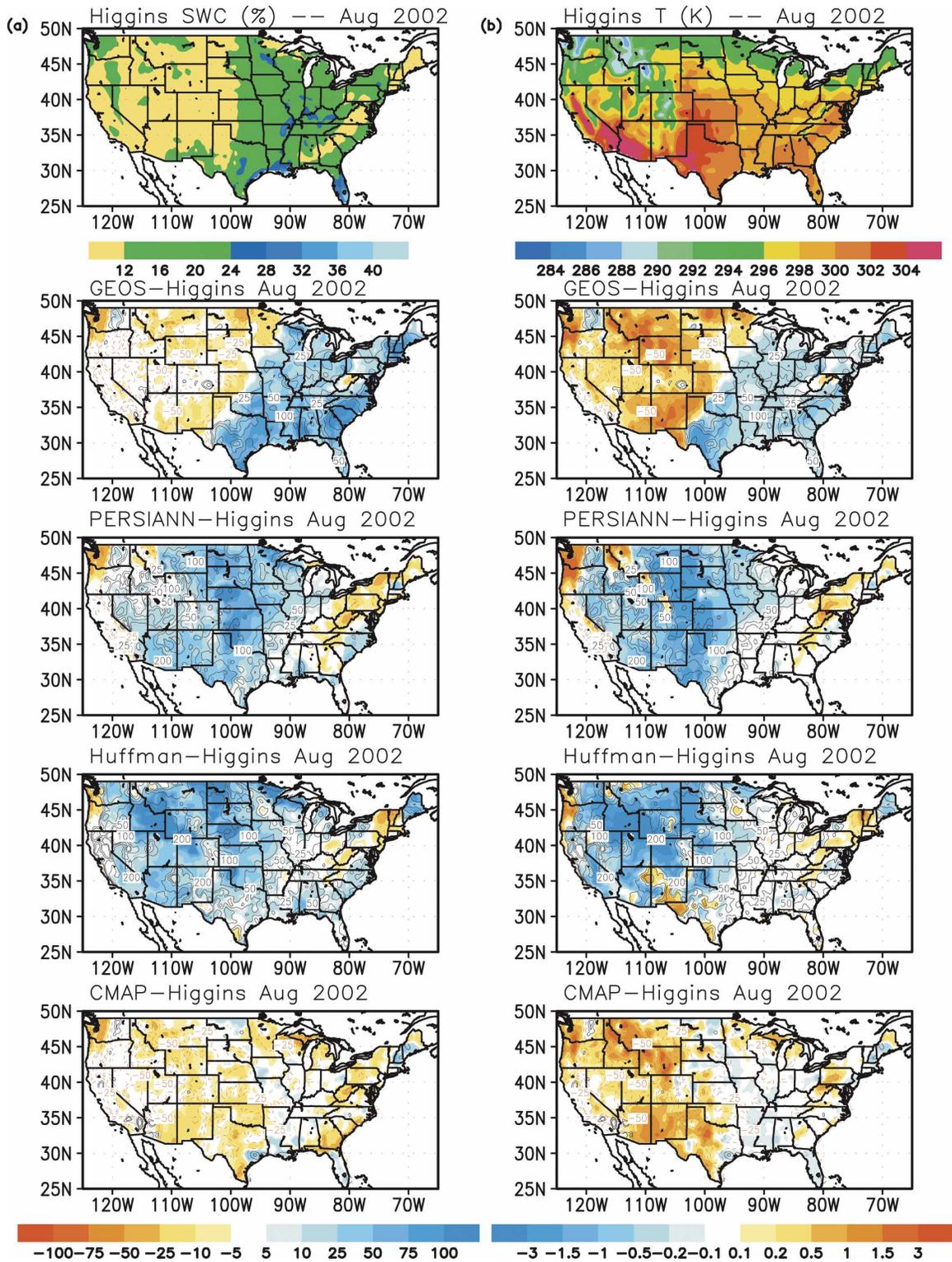


FIG. 10. Monthly average (a) volumetric soil water content (%) and (b) soil temperature for Aug 2002. (top row) Higgins GLDAS runs and (bottom four rows) differences between Higgins and GEOS, PERSIANN, HUFFMAN, and CMAP GLDAS runs. Contours [black (+), red (-)] are the percent precipitation differences taken from sections 3a(1) and 3a(2).

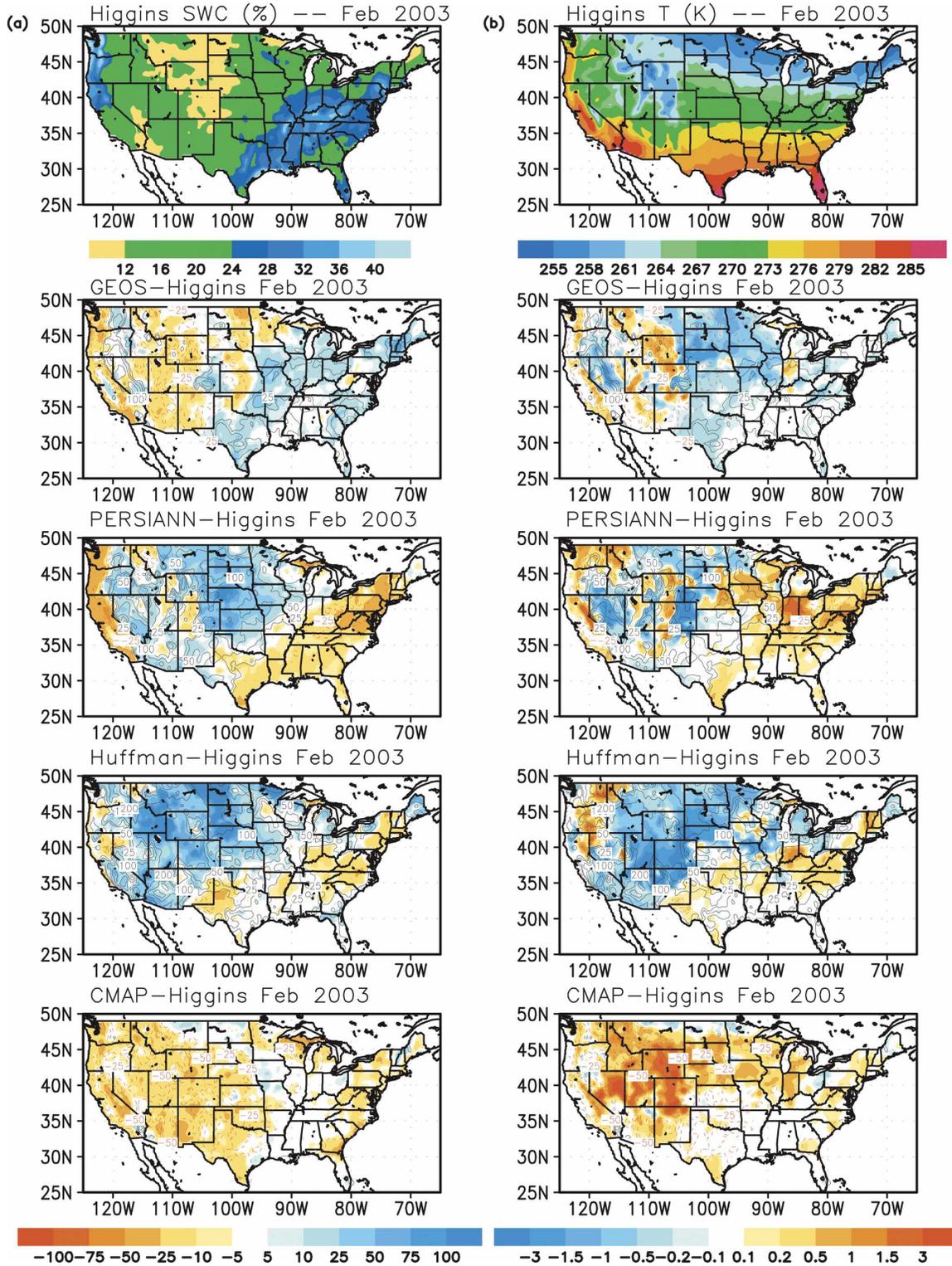


FIG. 11. Monthly average (a) volumetric soil water content (%) and (b) soil temperature for Feb 2003. (top row) Higgins GLDAS runs and (bottom four rows) differences between Higgins and GEOS, PERSIANN, HUFFMAN, and CMAP GLDAS runs. Contours [black (+), red (-)] are the percent precipitation differences taken from sections 3a(1)–(4).

unchanged. The lowest SWC is evident across the northern Rockies ( $<12\%$ ). Substantial differences remain between the other simulations and the states produced by Higgins. The GEOS run becomes more in line with that shown by Higgins in the eastern CONUS where differences now range from  $-5\%$  to  $50\%$ . The PERSIANN and HUFFMAN runs also indicate drier conditions in the eastern United States from August 2002 and now indicate small negative differences ( $5\%$ – $50\%$ ). Both simulations remain more wet across most of the central United States and interior west. In addition, the PERSIANN run indicates a substantial area of drier conditions along the West Coast, which is consistent with the precipitation comparison during the winter season (Fig. 8e). The CMAP simulation continues to show the least error when compared to the Higgins run but greater negative differences are evident across the western United States as compared to August.

Substantial differences remain between the simulations at the end of February for  $T_{\text{soil}}$ . The GEOS run illustrates several changes including a weakening of the warmer temperatures in the interior west, a cooler land surface across the northern plains, and generally comparable temperatures across the eastern United States. The PERSIANN and HUFFMAN runs both show a slightly warmer land surface in the eastern United States with a cooler land surface in the west, especially for HUFFMAN. The CMAP simulation indicates a substantially warmer land surface in the western United States ( $>3$  K in some areas)—a signal enhanced from August 2002.

Another important quantity predicted by land surface models is SWE. Figure 12 illustrates SWE for the Higgins simulation and the differences from Higgins of the GEOS, PERSIANN, HUFFMAN, and CMAP simulations during February 2003. The Higgins simulation indicates higher SWE in the interior west ( $100$ – $400$   $\text{kg m}^{-2}$ ) and also in the northeast United States ( $100$ – $200$   $\text{kg m}^{-2}$ ). The PERSIANN and CMAP simulations both underestimate the SWE totals in the northeast United States and Great Lakes by generally  $25$ – $100$   $\text{kg m}^{-2}$  while the HUFFMAN and GEOS runs generally overestimate this feature. A common feature of the SWE values across much of the western United States is that the GEOS, PERSIANN, and CMAP runs generally underestimate SWE up to and greater than  $100$   $\text{kg m}^{-2}$  in some areas. The HUFFMAN run substantially overestimates SWE in this area and is a result of the overestimation of precipitation during the late fall and winter—initiating by complications due to microwave precipitation estimates in frozen surface areas (this idea is discussed in section 4). Figure 12 illustrates SWE for the month of February. The snowpack, however, generally

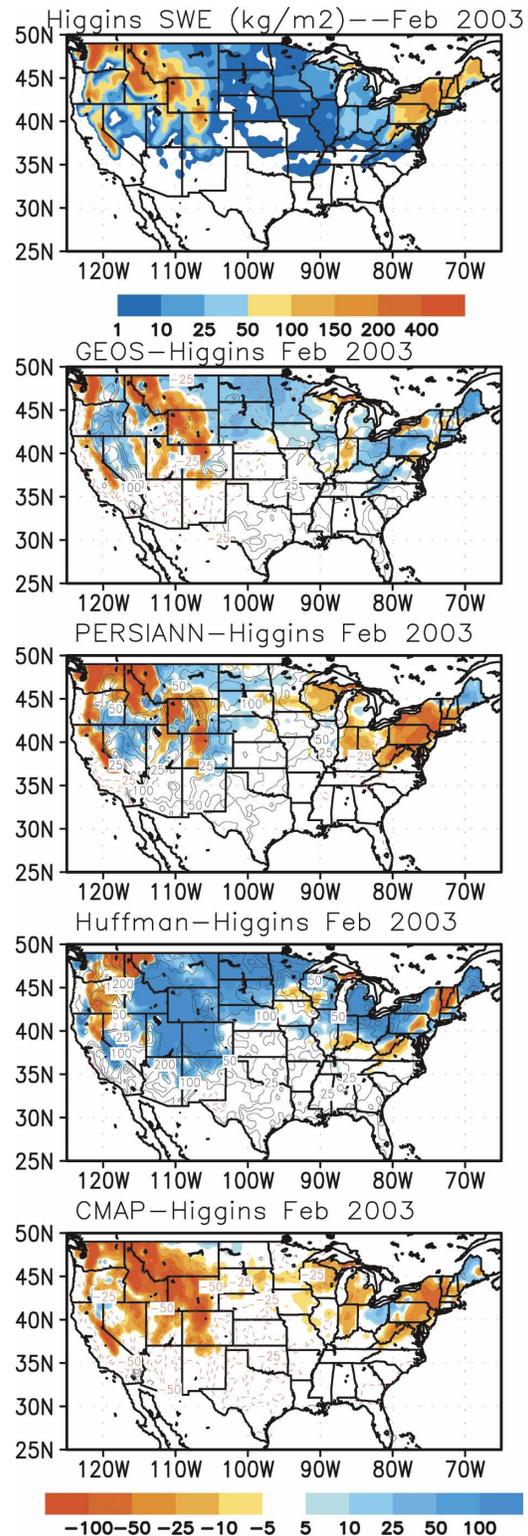


FIG. 12. Monthly average snow water equivalent ( $\text{kg m}^{-2}$ ) for Feb 2003. (top row) The Higgins GLDAS run and (bottom four rows) differences between Higgins and GEOS, PERSIANN, HUFFMAN, and CMAP GLDAS runs. Contours [black (+), red (-)] are the percent precipitation differences taken from sections 3a(1)–(4).

shows the greatest variability during the later spring-time melt period. Therefore, it is important to note that the evolution and sensitivity of the snowpack for different precipitation forcing may be greater during the springtime melt period than that illustrated here.

Figures 10–12 can be used to link the differences in precipitation described in section 3a quantitatively with the subsequent impacts on land surface states evident by the GLDAS simulations. Several main points can be made by evaluating these figures. Focusing on SWC (Figs. 10, 11a), it can be seen in both seasons that the percent differences in GLDAS SWC typically are less than the percent differences in precipitation (contours) taken from section 3a. The implication of this is that the GLDAS system, specifically the Mosaic LSM, generally acts to “dampen” the imposed precipitation differences. There is, however, substantial variation in the degree of damping over the CONUS, and in some areas the percent differences in SWC are or nearly are equivalent to the percent forcing differences. For example, this can be seen with the GEOS simulation in New England, with the PERSIANN simulation in the mid-Atlantic and Northeast, and with the CMAP simulation in the southeast and the northern Great Lakes. The extent of damping is determined by both land surface characteristics (vegetation type and soil type) and surface meteorology (amount of precipitation, incoming radiation, and low-level wind speed, temperature, and humidity) as these variables impact the nonlinear processes in the soil/canopy system.

For soil temperature during the spring and summer months, negative differences in precipitation tend to result in larger absolute increases in temperature than do equivalent percent positive differences in precipitation that result in absolute decreases in temperature (Fig. 10b). By the end of the simulations in February 2003, however, there are some interesting changes, brought on by cold-season processes. The precipitation–soil temperature relationship seen in Fig. 10b during the summer (greater precipitation, cooler temperatures or less precipitation, warmer temperatures) breaks down during the winter season in some areas, and  $T_{\text{soil}}$  is not as closely linked to precipitation as SWC. An example of this is illustrated by looking at the GEOS simulation in Fig. 11b where in the northern plains, most of the region indicates a cooler  $T_{\text{soil}}$  despite areas of negative percent differences in precipitation. During the winter, GEOS produces greater precipitation than Higgins (contours in Fig. 8a) and since temperatures are cold, most of this precipitation falls in the form of snow (see Fig. 12b) and therefore results in a cooler soil system despite less precipitation.

## 4. Discussion

### a. CONUS precipitation comparison

The CONUS precipitation comparison outlined in section 3a indicates a wide range of accuracies when comparing several types of precipitation estimates with the Higgins gauge dataset. These differences are dependent not only on the types of products but also with season. Excluding the NEXRAD ground radar, the CMAP precipitation estimates agree most closely with the Higgins gauge data. Average CONUS rmse shows that the CMAP product performs the best for the spring, summer, and fall periods. Moreover, comparisons of the correlation of daily precipitation illustrate that the CMAP estimates are superior during all seasons.

It is important to note that the CMAP and Higgins gauge networks are independent. The Higgins dataset includes approximately 5500 gauges from the River Forecast Center and Climate Anomaly Database, while the CMAP dataset utilizes far fewer (~1200) gauges from the GTS network. Consequently, although it is possible that these products may overlap to some degree, they can be considered for all intents and purposes independent. In this study, the comparison of CMAP with gauge data is only conducted for CONUS so that there may be a question as to whether the validity of the results can be extrapolated globally. As seen in Fig. 13, the number of rain gauges in most areas around the world is similar to that available in the CONUS and sufficient for the authors to make the assumption that the good performance demonstrated over the CONUS domain applies in most areas globally.

There are important issues highlighted by these results that require further discussion. Although the Higgins gauge data is used for verification in this study, as mentioned in the introduction, there are nontrivial errors due to “undercatch” of precipitation when using rain gauge datasets. The Higgins gauge data does not correct for these systematic errors as some other datasets (Adams and Lettenmeier 2003). The findings presented here would benefit from additional comparison using a dataset of this type. Since the majority of the precipitation estimates are greater than the Higgins product, it is reasonable to expect the errors presented here to be less. Moreover, the CMAP and NEXRAD estimates would also agree more closely with the other products as they utilize gauge observations as well.

The good performance of the model-based assimilation products, relative to the satellite methods, is an important finding of this study. These precipitation estimates were, in general, more accurate (both in rmse

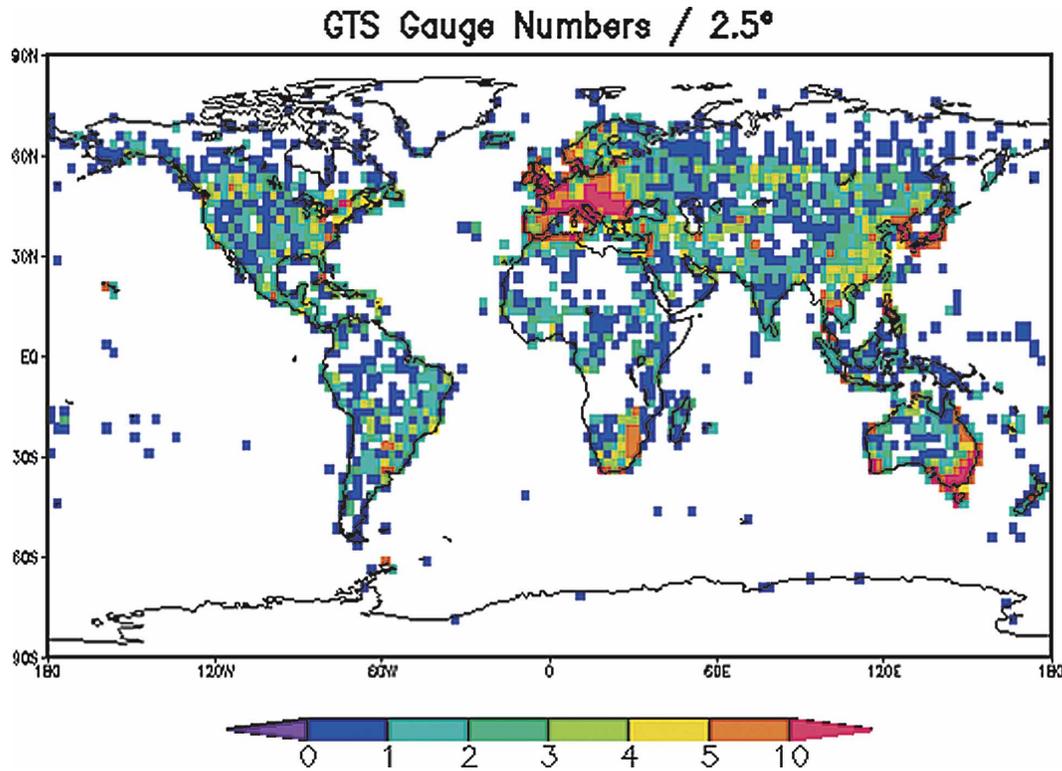


FIG. 13. The number of gauges in each  $2.5 \times 2.5$  grid used in the generation of the CMAP precipitation estimate.

and correlation) than the satellite-derived products, with the ECMWF data showing the greatest accuracy of the three models. The good performance of the ECMWF models highlights an important implication—that assimilating atmospheric observations (those as part of the ECMWF product for CONUS), under certain circumstances, appears to rival the ingesting of ground gauge measurements as part of the CMAP product. The ramifications of this are clear and noteworthy for land surface modeling as limitations as a result of rain gauge density may be lessened.

Although the model products generally perform well relative to the satellite estimates, there are some exceptions, however. First, during the fall season, the satellite product and model estimate errors are comparable. Second, the correlations of daily precipitation during the summer across the central United States are greater for the satellite estimates. The higher degree of temporal agreement during summer is important because it shows that although these products overestimate the total rainfall during the period, the timing of intra- and interdaily areas of precipitation associated with meso-scale convective systems are more accurately represented by the satellite estimates as opposed to the model-based products. Along these lines, these findings

warrant further investigation as they have implications in GLDAS land surface modeling by positively impacting the simulation of surface runoff during the warm season where precipitation often exists over a limited area with greater daily variability.

The satellite-only products suffer from two important biases for CONUS when compared to the other precipitation estimates in section 3a. These are 1) an overestimate of precipitation for the central United States during the summer and 2) an overestimate of precipitation in the northern central plains and mountainous west during the spring and winter. The latter appears limited to the HUFFMAN precipitation estimate. The summertime overestimation of precipitation by the satellite-only products is a result of utilizing geostationary IR data and detecting cold cloud-top temperatures from high cloudiness. High cloud tops from convective systems (e.g., anvils, etc.) during the summer season are most likely responsible for this widespread area of greater precipitation for these two products. Figure 14a depicts cirrus reflectance from MODIS for the June–August 2002 time period and shows a local maximum in the central United States as compared to the rest of North American land areas. Although the extent and magnitude of this cirrus area is smaller than that for

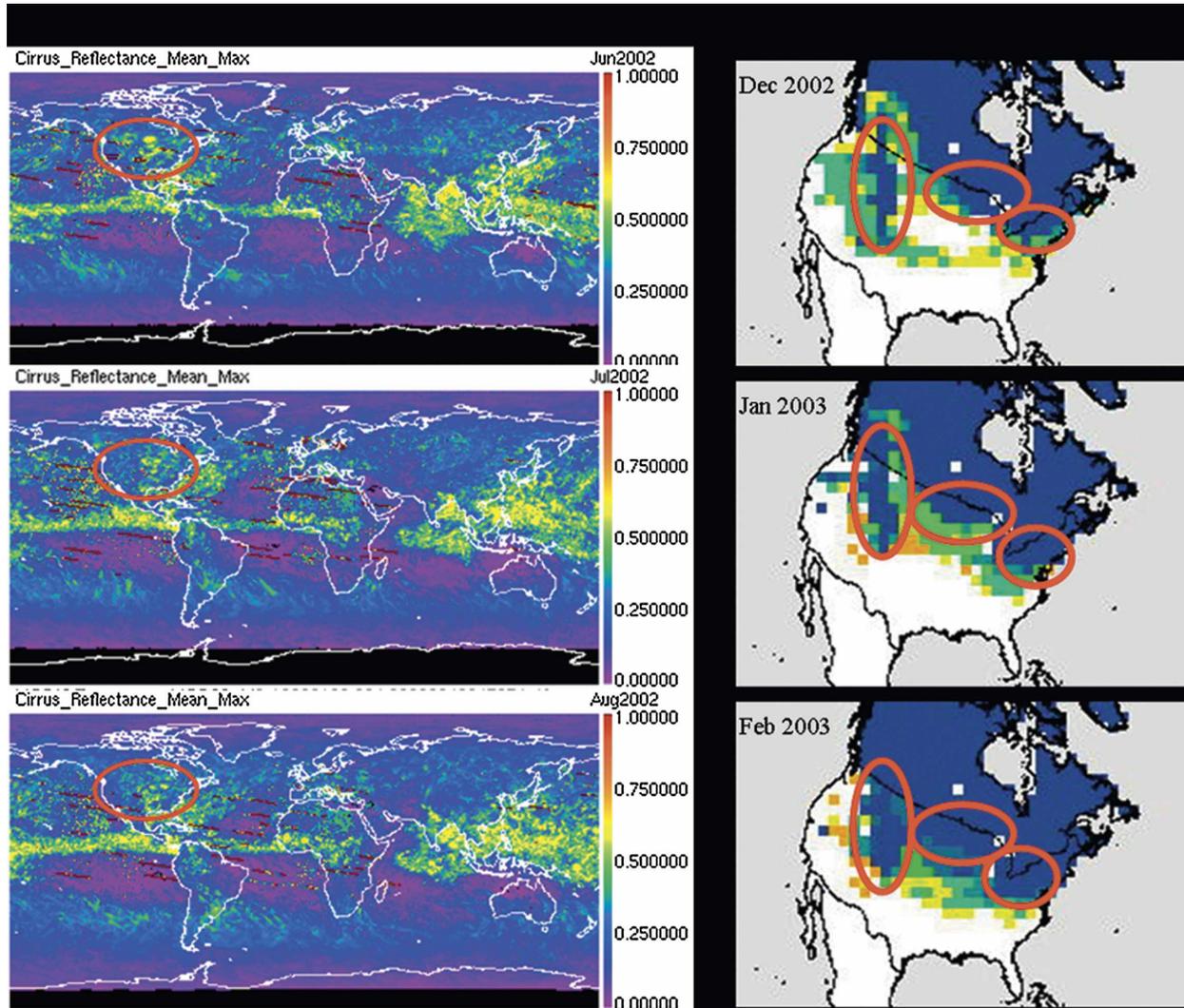


FIG. 14. (a) Cirrus reflectance for Jun, Jul, and Aug 2002 from MODIS (red circles highlight the North American land area) and (b) percentage of snow cover for Dec 2002 and Jan and Feb 2003. (From Rutgers University Global Snow Laboratory; red circles indicate the highest frequency of snow cover coincident with the high HUFFMAN precipitation estimates.)

deep convection in the Tropics, it nonetheless indicates the presence of mid-high, cold cloud tops in the central United States. The CMAP estimate does not suffer from this bias since its estimates are constrained using available rain gauge data.

The second bias, mainly affecting the HUFFMAN estimate, is a result of the application of passive remote sensing techniques (Table 1) in the presence of frozen surfaces. Passive microwave precipitation retrieval schemes falter and then fail near regions of frozen, snowy, or icy surfaces. Under these circumstances, for the HUFFMAN product, the microwave estimate is not used, but in order to provide a continuous precipitation field, the IR estimate at the same location must be calibrated by a microwave estimate from a nearby lo-

cation through interpolation (G. J. Huffman, NASA GSFC, 2004, personal communication). Unfortunately, it is often the case that the clouds and land surface are much colder than the surrounding areas that were used as part of the interpolation process, and this produces large precipitation estimates (G. J. Huffman, NASA GSFC, 2004, personal communication). Figure 14b illustrates the percentage of snow cover for December 2002 and January and February 2003 for North America and illustrates greater than 90% snow cover across areas of the northern plains, the Rockies, and near the Great Lakes. These areas match very well, especially for January 2003, with coincident areas of high wintertime precipitation totals indicated by the HUFFMAN dataset in section 3a (Fig. 8d). This bias is

expected to be removed with the upcoming reprocessing of the TRMM Real-time Multi-satellite Precipitation Analysis (G. J. Huffman, NASA GSFC, 2004, personal communication).

The differences between precipitation estimates in many cases throughout this study are greater than that shown by normal interannual variability. In general, the standard deviation of seasonal precipitation does not exceed 200 mm except during the winter along the West Coast. The highest values are across the southeast United States and along the West Coast except during the summer when higher values are also prevalent across the central United States. For instance, the differences of the satellite-only estimates during the summer in the central plains are 2–5 times greater and clearly indicate the negative potential impact on land surface states over longer-term GLDAS simulations.

#### *b. GLDAS land surface states*

The GLDAS simulations indicate that there is a considerable impact on land surface states when using different precipitation forcing. The percent differences in SWC ranged from  $-75\%$  to  $+100\%$  for both summer and winter seasons and changes in  $T_{\text{soil}}$  ranged up to  $\pm 3.0$  K. The greatest differences generally are located over the western CONUS, especially for  $T_{\text{soil}}$ . Generally speaking, the percent differences in SWC are  $25\%$ – $75\%$  less than the percent differences in precipitation forcing, indicating a considerable “damping” of initial precipitation differences. There are instances, however, where the land surface state differences are of equivalent magnitude to that of the precipitation forcing.

The range of differences cited above are large and most likely will produce considerable impacts on seasonal weather and climate forecasts when used for land surface model initialization. Rind (1982) found that a 75% decrease in soil moisture led to substantially higher surface temperature and decreased precipitation. Oglesby and Erickson (1989) showed that a decrease in soil wetness from 0.11 to 0.01 (90% decrease) produced an increase in surface temperature, increased ridging aloft, and a northward shift of the jet stream. A better simulation of anomalous rainfall in 1993 was described by Beljaars et al. (1996) when moving from soil moisture initialized with 25%–100% field capacity (75% range). Schär et al. (1999) reported that precipitation was heavily dependent on soil moisture when conducting experiments of anomalies 50% below and 100% above the control. Koster and Suarez (2003) found that when using “realistic” soil moisture (characteristic ensemble anomalies of 10% saturation) a substantial and statistically significant impact on summer-

time precipitation for some continental regions, including the CONUS domain, was evident.

One can interpret the percent differences between the different GLDAS simulations as “pseudoanomalies” where the Higgins GLDAS simulation is taken to be the standard or climatological land surface state. In this case, the “anomalies” shown here range from  $-75\%$  to  $+100\%$  and are well within or above the range of cited values above that resulted in substantial and statistically significant impacts on monthly and seasonal weather forecasts. The findings of the above studies in combination with the results presented here indicate the importance of using the most accurate precipitation forcing, as large differences often are directly translated into equally large percent differences in land surface states such as SWC,  $T_{\text{soil}}$ , and SWE, which in turn result in extensive changes in seasonal weather prediction through modification of precipitation and surface temperature—among other variables.

Based on the findings from this study, the GLDAS group plans to use the disaggregated CMAP precipitation data described here for its precipitation forcing. A comprehensive evaluation of the representation of the diurnal cycle and distribution of precipitation rate for the products described in this study is ongoing. If necessary, the authors plan to adjust the temporal disaggregation methods when using the pentad CMAP precipitation data based on these findings.

The GLDAS team plans to address a number of the shortcomings of this study in the near-term. First, we plan to extend the CONUS analysis of precipitation totals and correlation to additional years of data. We plan to also utilize the recently released CPC morphing precipitation dataset (CMORPH) made available by the National Oceanic and Atmospheric Administration (NOAA)/CPC in future comparisons. Second, the problems with the HUFFMAN precipitation over land areas during the cold season makes comparison of its estimates difficult for some areas and we plan to adjust our analysis when the reprocessed product is available. Third, in this paper the use of observations to validate the precipitation estimates are limited to the CONUS. We plan to utilize other rain gauge networks across the globe in order to more fully measure the reliability of all of these products.

In the long term, a research goal that the GLDAS team plans to proceed toward is to objectively classify precipitation statistics (i.e., totals, daily correlation, diurnal cycle, etc.) based on the meteorological environment in order to understand in a quantified manner which product(s) perform the best under a given meteorological environment. Armed with this information, simulations can be performed wherein different

precipitation estimates are “switched on and off” depending on the current meteorological situation in a given geographical location during a given season. An evaluation would then be performed as to the impacts of using this combination of precipitation datasets as compared to using each individual dataset alone.

*Acknowledgments.* The authors thank George Huffman at NASA GSFC for his helpful information and the efforts of Pedro Viterbo (ECMWF) in obtaining additional ECMWF precipitation data. The authors would also like to thank three anonymous reviewers for their insightful comments and suggestions to improve this manuscript and all the data providers for the use of their respective datasets. This research was supported by NASA’s Earth Science Enterprise.

#### REFERENCES

- Adams, J., and D. Lettenmeier, 2003: Adjustment of global gridded precipitation for systematic bias. *J. Geophys. Res.*, **108**, 4001, doi:10.1029/2001JD001260.
- 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.
- Arkin, P. A., and B. Meisner, 1987: The relationship between large-scale convective rainfall and cold cloud over the Western Hemisphere during 1982–1984. *Mon. Wea. Rev.*, **115**, 51–74.
- , and P. E. Ardanuy, 1989: Estimating climatic-scale precipitation from space: A review. *J. Climate*, **2**, 1229–1238.
- , and P. P. Xie, 1994: The Global Precipitation Climatology Project: First Algorithm Intercomparison Project. *Bull. Amer. Meteor. Soc.*, **75**, 401–419.
- , R. Joyce, and J. E. Janowiak, 1994: IR techniques: GOES precipitation index. *Remote Sens. Rev.*, **11**, 107–124.
- Baldwin, M., and K. Mitchell, 1997: The NCEP hourly multisensor U.S. precipitation analysis for operations and GCIP research. Preprints, *13th Conf. on Hydrology*, Long Beach, CA, Amer. Meteor. Soc., 54–55.
- Barrett, E. C., and D. W. Martin, 1981: *The Use of Satellite Data in Rainfall Monitoring*. Academic Press, 340 pp.
- Beljaars, A. C., P. Viterbo, M. J. Miller, and A. K. Betts, 1996: The anomalous rainfall over the United States during July 1993: Sensitivity to land surface parameterization and soil moisture anomalies. *Mon. Wea. Rev.*, **124**, 362–383.
- Charney, J., 1975: Dynamics of deserts and drought in Sahel. *Quart. J. Roy. Meteor. Soc.*, **101**, 193–202.
- , W. Quirk, S. Chow, and J. Kornfield, 1977: Comparative study of effects of albedo change on drought in semi-arid regions. *J. Atmos. Sci.*, **34**, 1366–1385.
- 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.
- 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.
- Fulton, R., J. Breidenbach, and D. Seo, 1998: The WSR-88D rainfall algorithm. *Wea. Forecasting*, **13**, 377–395.
- Griffith, C. G., J. A. Augustine, and W. L. Woodley, 1981: Satellite rain estimation in U.S. High Plains. *J. Appl. Meteor.*, **20**, 53–66.
- Grody, N. C., 1991: Classification of snow cover and precipitation using the Special Sensor Microwave Imager. *J. Geophys. Res.*, **96**, 7423–7435.
- Gruber, A., X. Su, M. Kanamitsu, and J. Schemm, 2000: The comparison of two merged rain gauge–satellite precipitation datasets. *Bull. Amer. Meteor. Soc.*, **81**, 2631–2644.
- 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.
- Higgins, R. W., W. Shi, and E. Yarosh, 2000: *Improved United States Precipitation Quality Control System and Analysis*. NCEP/Climate Prediction Center Atlas 7, 40 pp.
- Hsu, K., X. Gao, S. Sorooshian, and H. V. Gupta, 1997: Precipitation estimation from remotely sensed information using artificial neural networks. *J. Appl. Meteor.*, **36**, 1176–1190.
- Hsu, L., H. V. Gupta, X. Gao, and S. Sorooshian, 1999: Estimation of physical variables from multi-channel remotely sensed imagery using a neural network: Application to rainfall estimation. *Water Resour. Res.*, **35**, 1605–1618.
- Huffman, G. J., and Coauthors, 1997: The Global Precipitation Climatology Project (GPCP) combined precipitation datasets. *Bull. Amer. Meteor. Soc.*, **78**, 5–20.
- , 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 real and post-real time. Preprints, *12th Conf. on Satellite Meteorology and Oceanography*, Long Beach, CA, Amer. Meteor. Soc., CD-ROM, P4.11
- Janowiak, J. E., A. R. Livezey, and G. Huffman, 1998: A comparison of the NCEP–NCAR reanalysis precipitation and the GPCP rain-gauge-satellite combined dataset with observational error consideration. *J. Climate*, **11**, 2960–2979.
- , R. J. Joyce, and Y. Yarosh, 2000: A real-time global half-hourly pixel resolution infrared dataset and its applications. *Bull. Amer. Meteor. Soc.*, **82**, 205–217.
- Kondragunta, C., and A. Gruber, 1997: Intercomparison of spatial and temporal variability of various precipitation estimates. *Adv. Space Sci.*, **19**, 457–460.
- Koster, R. D., and M. J. Suarez, 1996: Energy and Water Balance Calculations in the Mosaic LSM. NASA Tech. Memo. 104606, Vol. 9, 76 pp.
- , and —, 2003: Impact of land surface initialization on seasonal precipitation and temperature prediction. *J. Hydrometeorol.*, **4**, 408–423.
- Krajewski, W., G. Ciach, J. McCollum, and C. Bacotiu, 2000: Initial validation of the global precipitation climatology project monthly rainfall over the United States. *J. Appl. Meteor.*, **39**, 1071–1086.
- Kummerow, C. Y., and Coauthors, 2001: The evolution of the Goddard profiling algorithm (GPROF) for rainfall estimation from passive microwave sensors. *J. Appl. Meteor.*, **40**, 1801–1840.
- Legates, D., and C. Willmott, 1990: Mean seasonal and spatial variability in gage-corrected, global precipitation. *Int. J. Climatol.*, **10**, 111–127.
- Meehl, G. A., and W. M. Washington, 1988: A comparison of soil moisture sensitivity in global climate models. *J. Atmos. Sci.*, **45**, 1476–1492.

- Mitchell, K. E., and Coauthors, 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.
- Myneni, R. B., and Coauthors, 2002: Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sens. Environ.*, **83**, 214–231.
- Neff, E., 1977: How much rain does a rain gage? *J. Hydrol.*, **35**, 213–220.
- Oglesby, R. J., and D. J. Erickson, 1989: Soil moisture and the persistence of North American drought. *J. Climate*, **2**, 1362–1380.
- Persson, A., 2001: User guide to ECMWF forecast products. Meteorological Bulletin 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.
- 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 pedo-transfer functions. *Water Resour. Res.*, **36**, 3653–3662.
- Reynolds, R. W., 1988: A real-time global sea surface temperature analysis. *J. Climate*, **1**, 75–87.
- Rind, D., 1982: The influence of ground moisture conditions in North America on summer climate as modeled in the GISS GCM. *Mon. Wea. Rev.*, **110**, 1487–1494.
- Rodell, M., and Coauthors, 2004: The global land data assimilation system. *Bull. Amer. Meteor. Soc.*, **85**, 381–394.
- Schär, C., D. Lüthi, U. Beyerle, and E. Heise, 1999: The soil-precipitation feedback: A process study with a regional climate model. *J. Climate*, **12**, 722–741.
- Schneider, U., B. Rudolf, and W. Ruth, 1993: The spatial sampling error of areal mean monthly precipitation totals analyzed from gauge measurements. *Extended Abstracts, Fourth Int. Conf. on Precipitation*, Iowa City, IA, 80–82.
- 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.
- Seo, D. J., 1998a: Real-time estimation of rainfall fields using rain gauge data under fractional coverage conditions. *J. Hydrol.*, **208**, 25–36.
- , 1998b: Real-time estimation of rainfall fields using radar rainfall and rain gauge data. *J. Hydrol.*, **208**, 37–52.
- Sevruk, B., 1982: Methods of correction for systematic error in point precipitation measurement for operational use. WMO Operation Hydrology Rep. 21, World Meteorological Organization, 26 pp.
- Shukla, J., and Y. Mintz, 1982: Influence of land-surface evapotranspiration on the earth's climate. *Science*, **215**, 1498–1501.
- Sorooshian, S., K.-L. Hsu, X. Gao, H. V. Gupta, B. Imam, and D. Braithwaite, 2000: Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bull. Amer. Meteor. Soc.*, **81**, 2035–2046.
- Spencer, R., 1993: Global oceanic precipitation from the MSU during 1979–91 and comparison to climatologies. *J. Climate*, **6**, 1301–1326.
- Sperber, K. R., and T. N. Palmer, 1995: Interannual tropical rainfall variability in general circulation model simulations associated with the atmospheric model intercomparison project. PCMDI Rep. 28, Lawrence Livermore National Laboratory, Livermore, CA, 79 pp. [Available from Lawrence Livermore National Laboratory, 7000 East Ave., Livermore, CA 94550-9234.]
- Sud, Y. C., and W. E. Smith, 1985: Influence of local land-surface processes on the Indian monsoon—A numerical study. *J. Climate Appl. Meteor.*, **24**, 1015–1036.
- Wilheit, T., A. Chang, and L. Chiu, 1991: Retrieval of monthly rainfall indices from microwave radiometric measurements using probability distribution functions. *J. Atmos. Oceanic Technol.*, **8**, 118–136.
- Wylie, D. D., 1979: An application of a geostationary satellite rain estimation technique to an extratropical area. *J. Appl. Meteor.*, **18**, 1640–1648.
- Xie, P. P., and P. A. Arkin, 1995: An intercomparison of gauge observations and satellite estimates of monthly precipitation. *J. Appl. Meteor.*, **34**, 1143–1160.
- , and —, 1997: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc.*, **78**, 2539–2558.
- , J. E. Janowiak, P. A. Arkin, R. Adler, A. Gruber, R. Ferraro, G. J. Huffman, and S. Curtis, 2003: GPCP Pentad precipitation analyses: An experimental dataset based on gauge observations and satellite estimates. *J. Climate*, **16**, 2197–2214.