Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project

Brian A. Cosgrove,[1,2] Dag Lohmann,[3] Kenneth E. Mitchell,[3] Paul R. Houser,[1] Eric F. Wood,[4] John C. Schaake,[5] Alan Robock,[6] Curtis Marshall,[7] Justin Sheffield,[4] Qingyun Duan,[5] Lifeng Luo,[6,8] R. Wayne Higgins,[9] Rachel T. Pinker,[10] J. Dan Tarpley,[11] and Jesse Meng[3]

Received 31 October 2002; revised 27 February 2003; accepted 31 March 2003; published 31 October 2003.

[1] The accuracy of forcing data greatly impacts the ability of land surface models (LSMs) to produce realistic simulations of land surface processes. With this in mind, the multi-institutional North American Land Data Assimilation System (NLDAS) project has produced retrospective (1996–2002) and real-time (1999–present) data sets to support its LSM modeling activities. Featuring 0.125° spatial resolution, hourly temporal resolution, nine primary forcing fields, and six secondary validation/model development fields, each data set is based on a backbone of Eta Data Assimilation System/Eta data and is supplemented with observation-based precipitation and radiation data. Hourly observation-based precipitation data are derived from a combination of daily National Center for Environmental Prediction Climate Prediction Center (CPC) gauge-based precipitation analyses and hourly National Weather Service Doppler radar-based (WSR-88D) precipitation analyses, wherein the hourly radar-based analyses are used to temporally disaggregate the daily CPC analyses. NLDAS observation-based shortwave values are derived from Geostationary Operational Environmental Satellite radiation data processed at the University of Maryland and at the National Environmental Satellite Data and Information Service. Extensive quality control and validation efforts have been conducted on the NLDAS forcing data sets, and favorable comparisons have taken place with Oklahoma Mesonet, Atmospheric Radiation Measurement Program/cloud and radiation test bed, and Surface Radiation observation data. The real-time forcing data set is constantly evolving to make use of the latest advances in forcing-related data sets, and all of the real-time and retrospective data are available online at http://ldas.gsfc.nasa.gov for visualization and downloading in both full and subset forms. *INDEX TERMS:* 1866 Hydrology: Soil moisture; 1899 Hydrology: General or miscellaneous; 3322 Meteorology and Atmospheric Dynamics: Land/atmosphere interactions; 3337 Meteorology and Atmospheric Dynamics: Numerical modeling and data assimilation; *KEYWORDS:* NLDAS, North American Land Data Assimilation System, forcing data, LSM, land surface modeling, LDAS

Citation: Cosgrove, B. A., et al., Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project, *J. Geophys. Res.*, *108*(D22), 8842, doi:10.1029/2002JD003118, 2003.

1. Introduction

[2] An essential component of land surface modeling studies is the forcing data used to drive the participating land surface models (LSMs). No matter how sophisticated

[1]Hydrological Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.

[2]Also at General Sciences Operation, Science Applications International Corporation, Beltsville, Maryland, USA.

[3]Environmental Modeling Center, National Centers for Environmental Prediction, National Oceanic and Atmospheric Administration, Camp Springs, Maryland, USA.

[4]Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey, USA.

[5]Office of Hydrologic Development, National Weather Service, National Oceanic and Atmospheric Administration, Silver Spring, Maryland, USA.

[6]Department of Environmental Sciences, Rutgers University, New Brunswick, New Jersey, USA.

[7]Department of Atmospheric Sciences, Colorado State University, Fort Collins, Colorado, USA.

[8]Now at Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey, USA.

[9]Environmental Modeling Center, National Centers for Environmental Prediction, National Oceanic and Atmospheric Administration, Camp Springs, Maryland, USA.

[10]Department of Meteorology, University of Maryland, College Park, Maryland, USA.

[11]Office of Research, National Environmental Satellite Data and Information Service, National Oceanic and Atmospheric Administration, Camp Springs, Maryland, USA.

0148-0227/03/2002JD003118$09.00

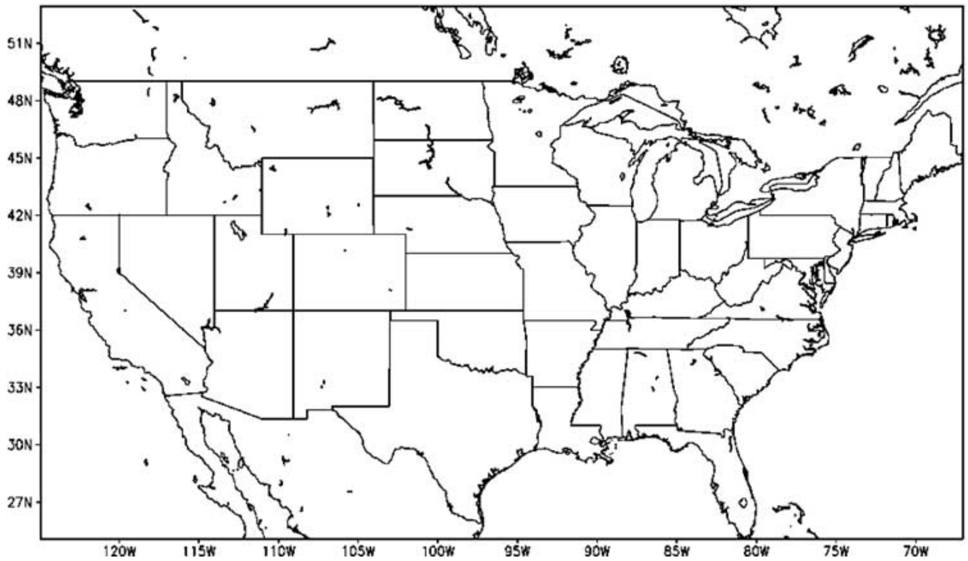

Figure 1. NLDAS 0.125° domain, extending from northern Mexico to southern Canada.

their depiction of land surface processes, or how accurate their boundary and initial conditions are, such models will not produce realistic results if the forcing data is not accurate. LSMs depend upon such externally supplied quantities as precipitation, radiation, temperature, wind, humidity and pressure to forecast land surface states, and errors in any of these quantities can greatly impact simulations of soil moisture, runoff, snow pack and latent and sensible heat fluxes. Each of these forcing quantities can be supplied by atmospheric Numerical Weather Prediction (NWP) models; however, such models are subject to internal model biases and errors in parameterizations that may negatively impact the quality of their output. As such, a more robust approach is to make use of as much observation-based forcing data as possible. This approach is especially important for offline Land Data Assimilation Systems (LDAS). Such systems seek to produce accurate simulations of land surface states by making use of observational data and isolating land surface modeling systems from the biases inherent in internally cycled NWP modeling systems.

[3] With this in mind, the North American Land Data Assimilation System (NLDAS) project [*Mitchell et al.*, 1999; K. E. Mitchell et al., The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system, submitted to *Journal of Geophysical Research*, 2003] has sought to construct quality controlled, spatially and temporally consistent, real-time and retrospective forcing data sets from the best available observations and model output to support its multi-LSM modeling activities. NLDAS is a multi-institution partnership, which involves participants from the National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Prediction Environmental Modeling Center (NCEP/EMC), National Aeronautics and Space Administration Goddard Space Flight Center (NASA GSFC), NOAA National Weather Service Office of Hydrologic Development (NWS/OHD), NOAA National Environmental Satellite Data and Information Service Office of Research and Applications (NESDIS/ORA), Princeton University, Rutgers University, the University of Washington and the University of Maryland. NLDAS utilizes the real-time and retrospective forcing data sets mentioned above to execute the Noah (M. B. Ek et al., Implementation of Noah land surface model advances in the NCEP operational mesoscale Eta model, submitted to *Journal of Geophysical Research*, 2003, hereinafter referred to as Ek et al., submitted manuscript, 2003), Mosaic [*Koster and Suarez*, 1996], VIC [*Liang et al.*, 1996] and Sacramento [*Burnash et al.*, 1973] LSMs. These forcing data sets feature hourly temporal resolution and 0.125° spatial resolution, and have been extensively quality controlled and validated [*Luo et al.*, 2003; *Pinker et al.*, 2003]. While these forcing data sets are based on a backbone of operational NCEP data assimilation fields (derived from merging observations with model fields) they are supplemented with extensive observation-based precipitation and shortwave radiation data—two forcing quantities that characteristically suffer significant biases in NWP assimilation and prediction systems and which greatly influence land surface simulations.

2. Retrospective Forcing

[4] The hourly NLDAS retrospective forcing data set features a 0.125° spatial resolution, is valid over the central North American NLDAS domain illustrated in Figure 1, and was produced in collaboration with NLDAS project members at NASA GSFC. The data set extends from 1996 to 2002, and a 1 October 1996 to 30 September 1999 subset serves as the basis for several companion NLDAS papers in this JGR issue. The retrospective forcing data set was constructed especially for the purposes of (1) executing the participating NLDAS LSMs for periods that overlap with special validation data sets, such as soil moisture [*Robock et al.*, 2003], (2) taking advantage of input forcing data that is not available in real time, especially additional gauge observations of precipitation (W. Shi et al., A unified rain gauge data set and multiyear daily precipitation reanalysis for the United States, submitted to *Journal of Geophysical Research*, 2003, hereinafter referred to as Shi et al., submit-

Table 1. Contents of NLDAS Forcing Files[a]

Model Based	Observation Based
Primary Fields	
2 m temperature	GOES-based downward shortwave radiation
2 m specific humidity	stage 2/gauge-based precipitation
Surface pressure	
10 m U wind component	
10 m V wind component	
Downward longwave radiation	
Convective precipitation	
Secondary Fields	
Downward shortwave radiation	GOES-based skin temperature
Total precipitation	GOES-based photosynthetically active radiation
Convective available potential energy	stage 2 precipitation

[a]Each hourly file contains 15 meteorological fields and includes both observation and model-based data. Nine of these 15 fields are primary forcing fields, while 6 are secondary fields that may be used for additional modeling and validation efforts. The primary observation-based NLDAS precipitation field is based on hourly disaggregation of daily CPC gauge-only totals using hourly stage 2 (Doppler radar with gauge bias correction) precipitation data.

ted manuscript, 2003), and (3) making use of observation-based precipitation and radiation data sets that pass through additional quality control checks which are not available to their real-time counterparts. Generation of the retrospective forcing data was made possible by the extensive Global Energy and Water Cycle Experiment Continental-Scale International Project (GCIP) sponsored archives of gauge-based, radar-based, and model-based data produced by EMC and CPC of NCEP at NCAR, and of GOES-based fields produced by the University of Maryland.

[5] Each hourly NLDAS forcing file is stored in the Gridded Binary (GRIB) format [*Stackpole*, 1994] to maintain NCEP operational protocol and to conserve storage space (each hourly file is about 2 megabytes in size). These files feature 15 instantaneous and accumulated meteorological fields, and include both observation and model-based data. Nine of these 15 fields are primary forcing fields, while six are secondary fields that are included to facilitate additional modeling and validation efforts (see Table 1).

[6] Model-based fields are derived from 3 hourly NCEP Eta Data Assimilation System (EDAS) output fields [*Rogers et al.*, 1996], and from 3 hourly and 6 hourly Eta mesoscale model forecast fields when EDAS data is unavailable from the archives, which occurs approximately 8% of the time. On the basis of the Eta model, the EDAS system is NCEP's main regional coupled 4D data assimilation system, is continuously cycled, and utilizes a series of 3 hourly forecast and analysis cycles to depict the current state of the atmosphere using observational data sources. The Noah LSM serves as the fully coupled land surface component of the Eta model, and validation studies have examined the sensitivity of the Eta model to upgrades in the Noah LSM (Ek et al., submitted manuscript, 2003), and to the choice of coupled LSM itself [*Betts et al.*, 1997]. Since the EDAS system is constrained through data assimilation, it should be noted that the system would be less sensitive to the choice of coupled LSM than would other nondata assimilating systems.

[7] In order to make use of EDAS/Eta data, values are first spatially interpolated from a standard 40 km (nominal) Lambert-conformal EDAS/Eta output grid, to the 0.125° NLDAS grid. The 40 km grid is used by the NWS Advanced Weather Interactive Processing System (AWIPS) and is commonly referred to as the AWIPS 212 grid. A bilinear interpolation procedure is used for all variables except for precipitation, which is interpolated using a budget bilinear method. The budget bilinear method views the input grid as bilinear overlapping hat functions, and the output grid as zero-order nonoverlapping step functions. This bilinear assumption allows for interpolation to higher-resolution grids, while at the same time the overall interpolation process allows for conservation of area average values. Once spatially interpolated, the data are then temporally interpolated from their native 3 hourly (EDAS, Eta) or 6 hourly (Eta) temporal resolutions, to the hourly time step required by NLDAS. Linear temporal interpolation algorithms are applied to all variables except those in the precipitation or shortwave radiation categories. In the case of precipitation, a block method is used wherein the precipitation rate is assumed to be constant between EDAS/Eta data points. Shortwave data are interpolated using a solar zenith angle-based process, which serves to more accurately depict the diurnal solar cycle than would a linear interpolation method.

[8] As NLDAS 0.125° topography differs significantly from the topography of the 40 km EDAS/Eta AWIPS 212 output grid (a product of spatial interpolation from the native Eta/EDAS computational grid), the next processing step involves adjusting the surface pressure, incident longwave radiation, 2 m temperature and 2 m humidity EDAS/Eta-based fields to account for such differences (Figure 2). Temperature is adjusted at each NLDAS gridpoint using the following equation:

$$T_{\mathrm{NLDAS}} = T_{\mathrm{EDAS}} + \gamma \Delta Z, \tag{1}$$

where T_{NLDAS} is the 2 m temperature (K) in the NLDAS system, T_{EDAS} is the EDAS/Eta 2 m temperature (K), γ is the lapse rate (assumed to be -6.5 K km^{-1}), and ΔZ is the difference in elevation (m) between the NLDAS and EDAS/Eta topography. This adjusted 2 m temperature is then used to adjust the 2 m pressure. Starting with the Hydrostatic Approximation and Ideal Gas law,

$$\frac{\delta p}{\delta z} = -\rho g \tag{2a}$$

and

$$p = \rho R T, \tag{2b}$$

where p is pressure (Pa), z is height (m), ρ is density (kg m^{-3}), g is gravity (m s^{-2}), and R is the Gas Constant (J kg^{-1} K^{-1}), we arrive at

$$\partial z = -\frac{RT}{g}\frac{\partial p}{p}. \tag{3}$$

Assuming that

$$\int T \partial Z \approx \frac{T_{\mathrm{EDAS}} + T_{\mathrm{NLDAS}}}{2} = \bar{T} = T_{\mathrm{mean}}, \tag{4}$$

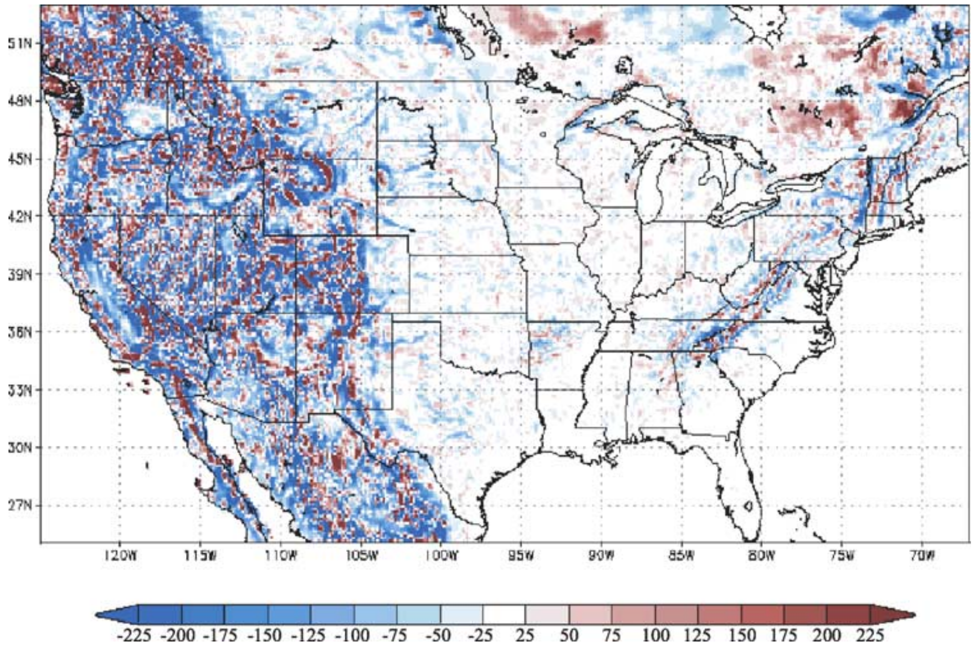

Figure 2. Difference in height (m) between 0.125° NLDAS topography and EDAS/Eta topography. EDAS/Eta temperature, longwave radiation, humidity, and pressure variables are each adjusted according to these differences before use in the LDAS project.

integration yields

$$\Delta Z = \frac{R\bar{T}}{g} \ln\left(\frac{p_{\mathrm{EDAS}}}{p_{\mathrm{NLDAS}}}\right) \tag{5}$$

and therefore

$$p_{\mathrm{NLDAS}} = \frac{p_{\mathrm{EDAS}}}{\exp\left(\frac{g\Delta Z}{RT_{\mathrm{mean}}}\right)}. \tag{6}$$

With the goal of maintaining an identical atmospheric demand for water vapor at the EDAS/Eta and NLDAS grid heights, the EDAS/Eta specific humidity q_{EDAS} (g g^{-1}) is elevation adjusted to arrive at q_{NLDAS} (g g^{-1}) by first assuming a constant relative humidity (%) throughout the ΔZ. The q_{NLDAS} required for this assumption to remain true based on the existing EDAS/Eta and NLDAS temperature and pressure values is then calculated. Combining the equation of state for water vapor and dry air with the definition of specific humidity, we can write

$$q_{\mathrm{sat_{EDAS}}} = \frac{0.622 * e_{\mathrm{sat_{EDAS}}}}{p_{\mathrm{EDAS}} - (0.378 * e_{\mathrm{sat_{EDAS}}})} \tag{7}$$

and

$$q_{\mathrm{sat_{NLDAS}}} = \frac{0.622 * e_{\mathrm{sat_{NLDAS}}}}{p_{\mathrm{EDAS}} - (0.378 * e_{\mathrm{sat_{EDAS}}})}, \tag{8}$$

where q_{sat} is the saturated specific humidity (g g^{-1}), e_{sat} is the saturated vapor pressure (hPa), p is the pressure (hPa), and, according to Wexler's saturated water vapor pressure equation,

$$e_{\mathrm{sat_{NLDAS}}} = 6.112 \exp\left[\frac{17.67(T_{\mathrm{NLDAS}} - 273.15)}{(T_{\mathrm{LDAS}} - 273.15) + 243.5}\right] \tag{9}$$

and

$$e_{\mathrm{sat_{EDAS}}} = 6.112 \exp\left[\frac{17.67(T_{\mathrm{EDAS}} - 273.15)}{(T_{\mathrm{LDAS}} - 273.15) + 243.5}\right]. \tag{10}$$

Given that

$$\mathrm{RH_{EDAS}} = \left(\frac{q_{\mathrm{EDAS}}}{q_{\mathrm{sat_{EDAS}}}}\right) * 100 \tag{11}$$

and that the RH (relative humidity) is being held constant over the course of the adjustment, the definition of specific humidity can be rewritten to yield

$$q_{\mathrm{NLDAS}} = \left(\frac{\mathrm{RH_{EDAS}} * q_{\mathrm{sat_{NLDAS}}}}{100}\right), \tag{12}$$

which represents the specific humidity value adjusted for elevation differences between NLDAS and EDAS/Eta topography.

[9] Downward longwave radiation is adjusted by starting with the Stefan-Boltzmann law:

$$L = \varepsilon\sigma T^4, \tag{13}$$

where L is the incident longwave radiation (W m^{-2}), ε is emissivity, and σ is the Stefan-Boltzmann constant (W m^{-2} K^{-4}). This leads to

$$L_{\mathrm{NLDAS}} = \frac{\varepsilon_{\mathrm{NLDAS}}\sigma}{\varepsilon_{\mathrm{EDAS}}\sigma} \left(\frac{T_{\mathrm{NLDAS}}}{T_{\mathrm{EDAS}}}\right)^4 * L_{\mathrm{EDAS}}, \tag{14}$$

where

$$\varepsilon_{\mathrm{NLDAS}} = 1.08 * \left\{1 - \exp\left[-e_{\mathrm{NLDAS}}^{\left(\frac{T_{\mathrm{NLDAS}}}{2016}\right)}\right]\right\} \tag{15}$$

Table 2. Observation-Based Precipitation Data Used in Derivation of NLDAS Precipitation Product[a]

Gridded Analysis	Advantages	Disadvantages
NWS stage 2 Doppler radar/gauge	hourly, 4 km	errors in radar-based magnitude; extremely limited coverage over Canada, Mexico; gaps from equipment downtime and topography
CPC real-time daily rain gauge data	less bias than radar estimates; PRISM adjustment; 0.125° resolution; least squares distance analysis	coarse temporal resolution; limited coverage over Canada, Mexico
CPC reprocessed daily rain gauge data	less bias than radar estimates; improved station density; improved QC checks	coarse temporal resolution; light coverage over Canada, Mexico; 0.25° resolution; overly smooth spatial analysis scheme; currently only available through 1998

[a]Hourly stage 2 product is used to temporally disaggregate daily CPC rain gauge data to produce hourly radar/gauge-based NLDAS precipitation data.

and

$$\varepsilon_{\mathrm{EDAS}} = 1.08^{*}\left\{1 - \exp\left[-e_{\mathrm{EDAS}}^{\left(\frac{T_{\mathrm{EDAS}}}{2016}\right)}\right]\right\} \tag{16}$$

and e is the vapor pressure (hPa) computed from

$$e_{\mathrm{NLDAS}} = \left[\frac{(q_{\mathrm{NLDAS}}{}^{*}p_{\mathrm{NLDAS}})}{0.622}\right] \tag{17}$$

and

$$e_{\mathrm{EDAS}} = \left[\frac{(q_{\mathrm{EDAS}}{}^{*}p_{\mathrm{EDAS}})}{0.622}\right]. \tag{18}$$

The calculations described above are a vital step in the production of forcing data, and yield significant temperature adjustments of up to 6 K, pressure adjustments of up to 120 hPa, longwave adjustments of up to 40 W m^{-2}, and specific humidity adjustments of up to 2 g kg^{-1}.

[10] Once these adjustments are completed, observation-based precipitation and shortwave radiation data are added to the forcing files. These two variables greatly impact land surface processes, but unfortunately, NWP models have an especially hard time accurately simulating these fields. As such, it is vital that NLDAS LSMs have access to precipitation and shortwave radiation data that is free from the NWP model biases and related problems. Toward this end, daily gauge-based precipitation data, hourly Doppler radar data, and Geostationary Operational Environmental Satellite (GOES)-based shortwave data are used to produce NLDAS forcing fields in addition to the EDAS/Eta-based fields described above.

[11] Each of the observation-based precipitation data sources mentioned above has characteristic strengths and weaknesses that are described in Table 2. Reflecting the strengths of each data set, NLDAS precipitation is derived by using the hourly Doppler radar product to temporally disaggregate the daily gauge product. This process, described in detail below, capitalizes on the accuracy of the daily gauge product, and on the temporal and spatial resolution of the Doppler radar product.

[12] CPC daily gauge analyses serve as the backbone of the NLDAS hourly precipitation forcing. These analyses are produced in both a 0.25° reanalysis mode [*Higgins et al.*, 2000], which is currently available through 1998 and is used in the retrospective NLDAS forcing described here, and in a 0.125° real-time mode (Shi et al., submitted manuscript, 2003), which is used by the real-time NLDAS forcing described in section 3. The real-time data set typically draws on 6,500 daily gauge reports, while the reprocessed data set is based on approximately 13,000 daily gauge reports. As such, the advantage of the reanalysis, also known as the "unified" analysis, is the approximate doubling of the number of applied gauge reports. Such a doubling occurs because CPC is able to combine or unify real-time reports with non-real-time reports (such as those from the NWS Cooperative Observer Network) obtained later. In NLDAS, the unified gauge-only daily precipitation analysis is interpolated to 0.125° resolution, and, along with the real-time gauge-only daily data, is temporally disaggregated into hourly fields. This is accomplished by deriving hourly disaggregation weights from NWS real-time, 4 km stage 2 hourly precipitation analyses. The stage 2 product consists of WSR-88D Doppler radar-based precipitation estimates that have been bias corrected using hourly multiagency gauge data [*Fulton et al.*, 1998], and mosaicked into a national product over the Continental United States (CONUS) by NCEP/EMC [*Baldwin and Mitchell*, 1997]. This CONUS mosaic of the stage 2 product is interpolated to 0.125° and any gaps in radar coverage (which total on average 13% of the area of the CONUS and are due to lack of radar coverage or equipment maintenance) are filled in with nearest neighbor stage 2 data from within a 2° radius. If no stage 2 data are available, then EDAS/Eta data are used instead. The patched, hourly stage 2 fields are then divided by fields of patched stage 2 daily precipitation totals to create hourly temporal disaggregation weights representing the proportion of the 24 hour total precipitation which fell in each hour. If the daily stage 2 total is zero in an area of nonzero CPC precipitation, hourly weights are set to 1/24 to spread the precipitation evenly over the entire day. These hourly weights are then multiplied by the daily gauge-only CPC precipitation analysis to arrive at temporally disaggregated, hourly NLDAS fields. Since the stage 2 data is only used to derive the hourly disaggregation weights, a daily summation of these NLDAS precipitation fields will exactly reproduce the original CPC daily precipitation analysis. This multistep temporal disaggregation process is illustrated in Figure 3 for a region of the southwestern United States.

[13] Examples of daily total precipitation fields obtained from each of the three observation-based and one model-based data sources used to generate NLDAS precipitation

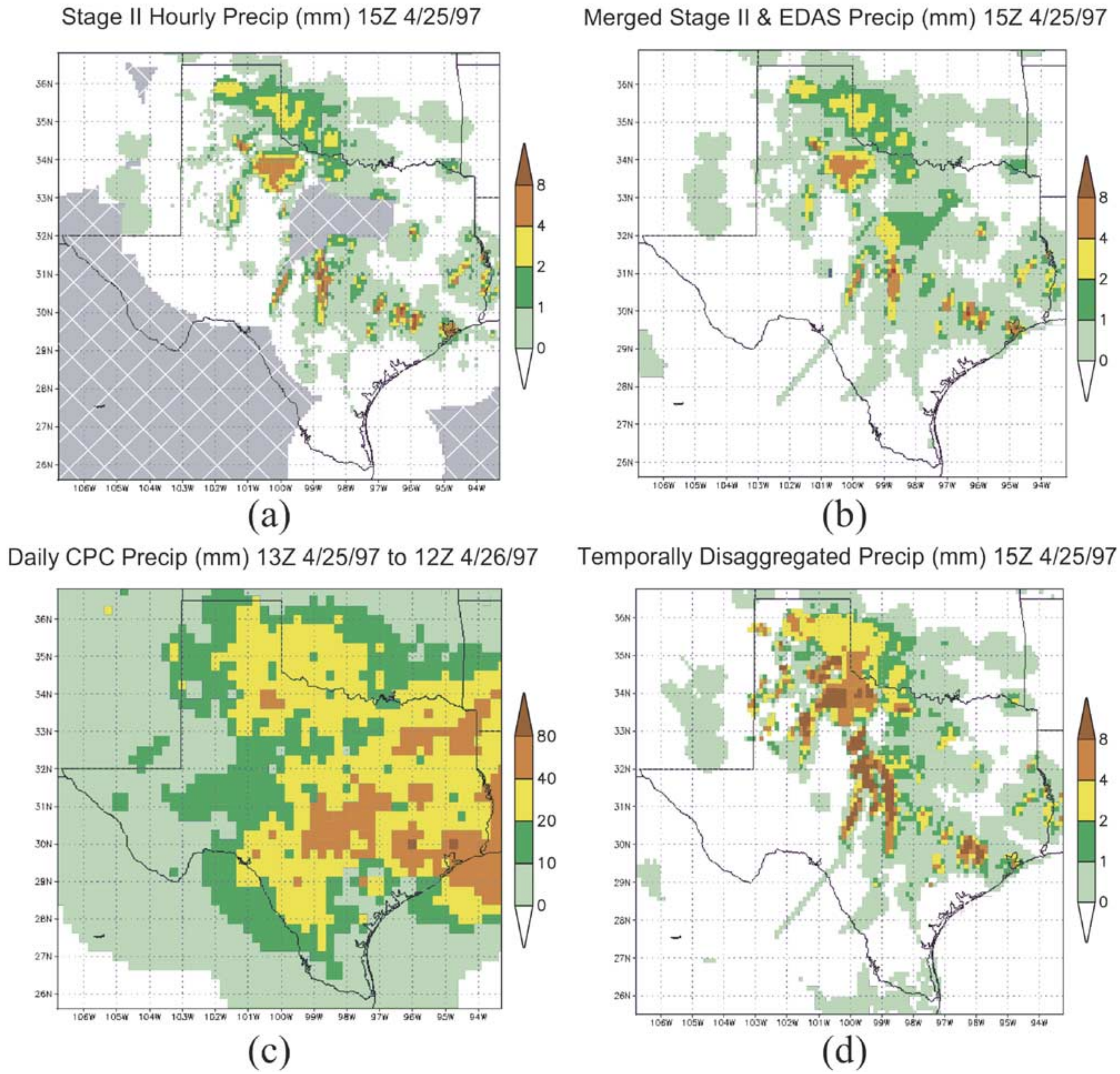

Figure 3. Sample derivation of NLDAS precipitation field (mm hr^{-1}) for 15Z on 25 April 1997. Stage 2 precipitation data is interpolated to (a) 0.125° and (b) any missing data filled with nearest neighbor stage 2 or EDAS precipitation data. (c) Daily CPC gauge-based data then interpolated to 0.125°. (d) Finally, temporal disaggregation weights derived from stage 2 data and applied to daily CPC gauge data to produce hourly gauge-based NLDAS precipitation fields. The crosshatch pattern indicates areas with no data.

data are given in Figure 4, and highlight the strengths and weaknesses of each data source. Comparison of Figure 4a with Figure 4b illustrates the increased gauge density that characterizes unified precipitation analyses. By contrast, comparison of Figure 4d with Figure 4a highlights the substantial errors that can occur in EDAS precipitation forcing for individual events. Although still present, such errors decreased after 21 July 2001 when NCEP began the operational assimilation of observed precipitation in the EDAS suite (Ek et al., submitted manuscript, 2003).

[14] The observation-based shortwave radiation fields featured in the NLDAS forcing data set are derived through application of algorithms to data gathered by NOAA's GOES satellites, via a close collaboration between the University of Maryland (UMD) and NESDIS/ORA [*Pinker et al.*, 2003]. UMD processes this raw data to create hourly, 0.5° instantaneous downward surface shortwave fields, which are then bilinearly interpolated to 0.125° spatial resolution for use in the NLDAS project (Figure 5). As these remotely sensed data are valid at 15 minutes after the hour, they must also be temporally interpolated to be valid on the hour so as to match the NLDAS forcing file conventions. This is accomplished with a zenith angle-based temporal interpolation algorithm that maintains the

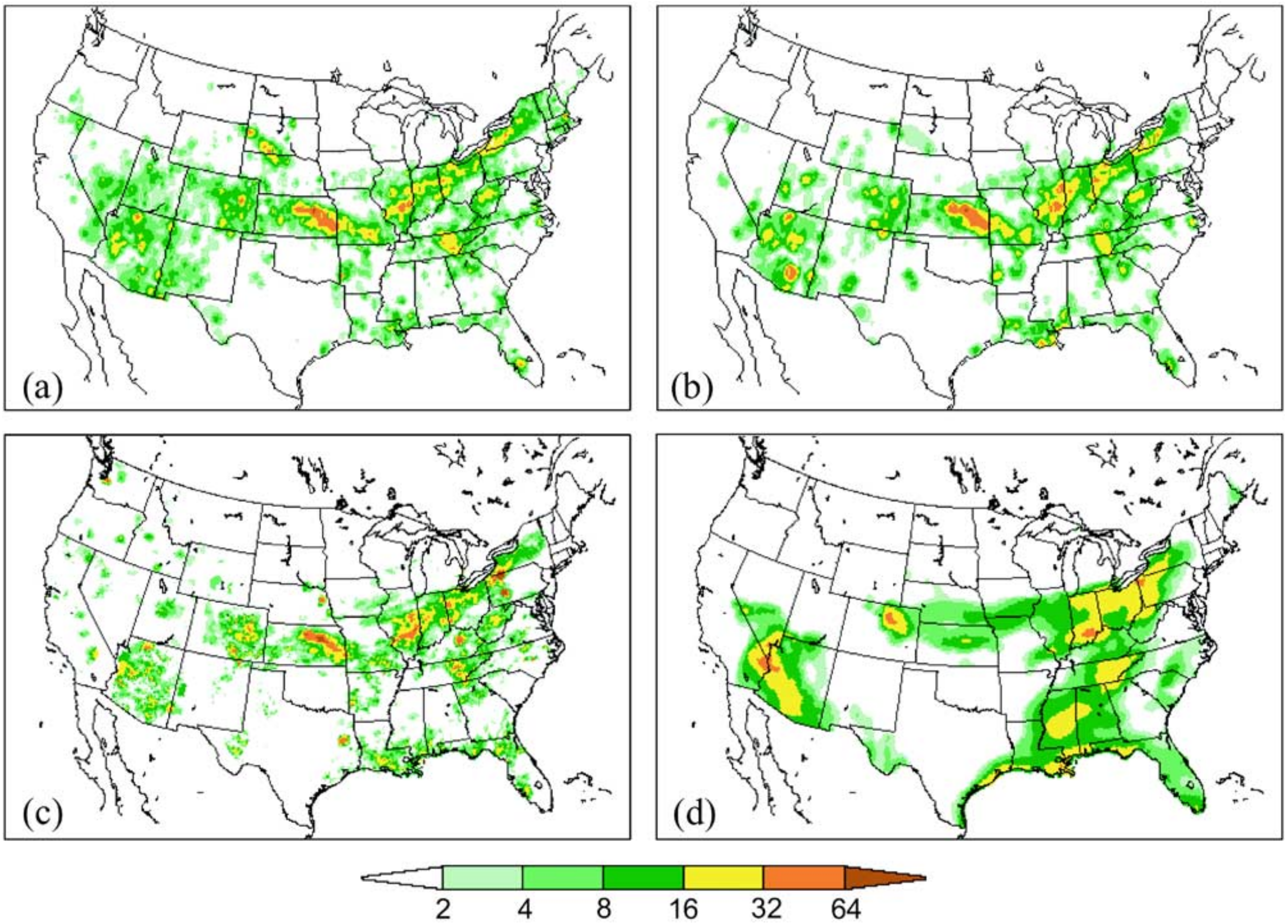

Figure 4. Sample data (mm precipitation per day) from each of the four precipitation data sets used in construction of NLDAS merged precipitation product. Data includes (a) CPC reprocessed or "unified" gauge data, (b) CPC real-time gauge data, (c) stage 2 radar/gauge data, and (d) EDAS precipitation data. Example data are for the 24 hour period ending 12Z 23 July 1998.

proper solar diurnal cycle. The UMD product is only valid over locations featuring solar zenith angles of less than 75°, and so must be supplemented with EDAS/Eta data near the day/night terminator when used by NLDAS LSMs. While there is some discontinuity between the EDAS/Eta and GOES fields along this line, it is not severe, and allows the NLDAS data set to make maximum use of available constituent data sources.

[15] Once all of the observation and model-based forcing fields are produced, they pass through quality control algorithms that ensure the integrity of the data. These checks are based on the Assistance for Land surface Modeling Activities (ALMA) forcing data conventions (http://www.lmd.jussieu.fr/ALMA), and examine the data for spurious, unrealistic values. Such checks are reinforced by in-depth NLDAS forcing validation efforts over the Southern Great Plains (SGP) region [*Luo et al.*, 2003]. These extensive studies have compared NLDAS data to Oklahoma Mesonet, Atmospheric Radiation Measurement Program (ARM)/cloud and radiation test bed (CART), and Surface Radiation (SURFRAD) values. As exemplified by Figure 6, these studies have demonstrated the high level of realism attained in the 0.125° NLDAS forcing data set (one exception is a low bias in the NLDAS precipitation forcing over high mountain elevations in the western CONUS, see section 3). Figure 7 highlights the improvements in accuracy of NLDAS shortwave radiation data attained when using GOES versus EDAS-based radiation data sets. While, in general, EDAS shortwave radiation is characterized by a 10% high bias, the GOES product is characterized by a smaller 5% low bias. A notable exception occurs in snow covered regions, where GOES data is often characterized by a high bias (albeit smaller than that of EDAS) due to the snow's impact on the cloud detection scheme of the GOES retrieval algorithm [*Pinker et al.*, 2003].

3. Real-Time NLDAS Forcing

[16] The NLDAS research effort is focused on both near real-time as well as retrospective high-quality land surface simulations. As such, it requires a near-real-time land surface forcing data set in addition to the retrospective data set described above. Both data sets are hourly and 0.125°, and feature identical sets of primary and secondary model and observation-based variables. They are produced using the same spatial and temporal interpolation methods and are quality controlled in similar fashions.

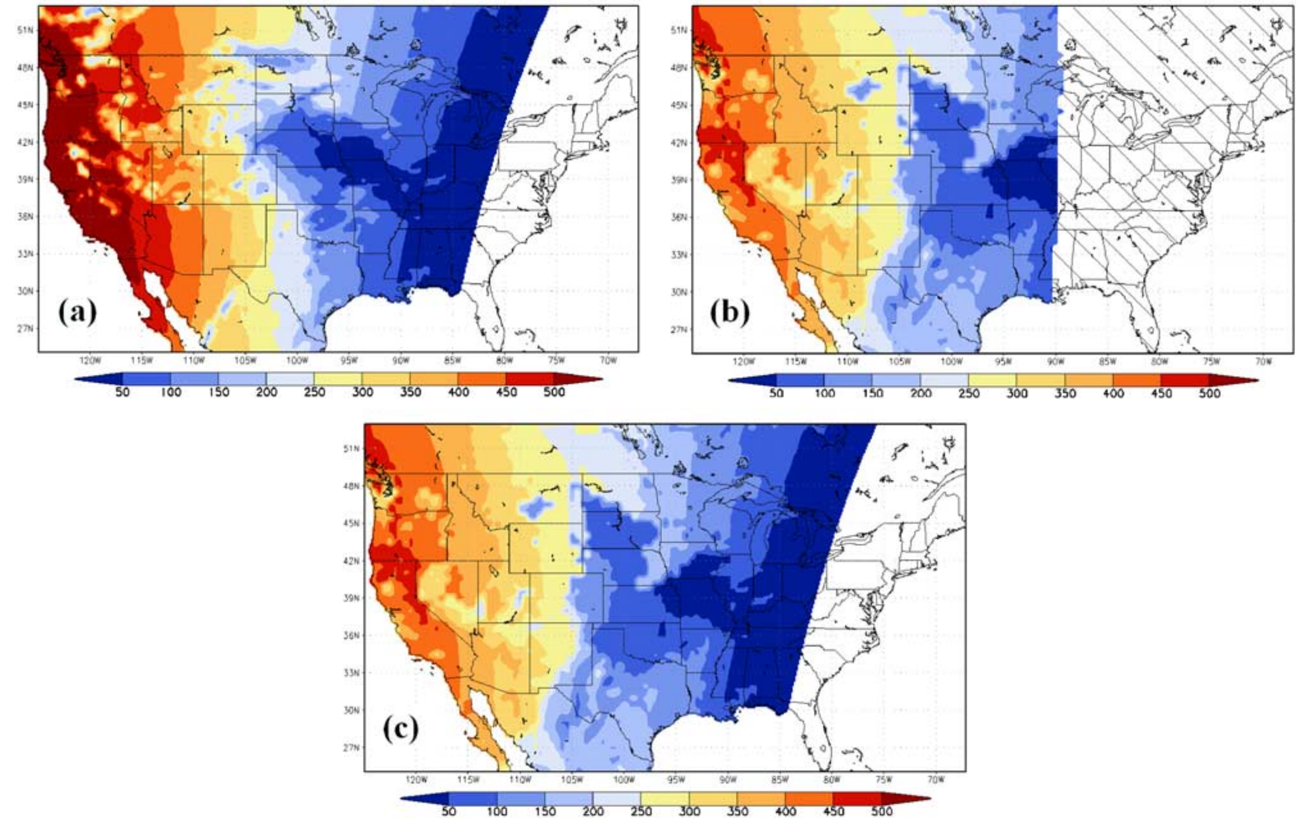

Figure 5. Downward solar radiation fields (W m^{-2}) at 00Z on 21 April 2002 from (a) Eta, (b) GOES, and (c) NLDAS merged data sets. Eta and GOES fields are merged to form NLDAS merged product. The diagonal pattern indicates areas with no data, while white represents zero values.

[17] However, differences in the philosophies and realities behind each forcing data set lead to notable differences between them. While the retrospective data set was produced over the course of one computer run, the real-time production effort depends upon an extensive and interrelated series of dynamic real-time data streams, C Shell scripts and Fortran programs to produce new forcing data once per day. In addition, the retrospective data set covers the period from 1996 to 2002, in contrast to the real-time data set, produced by NCEP, which covers the years from 1999 to present. So

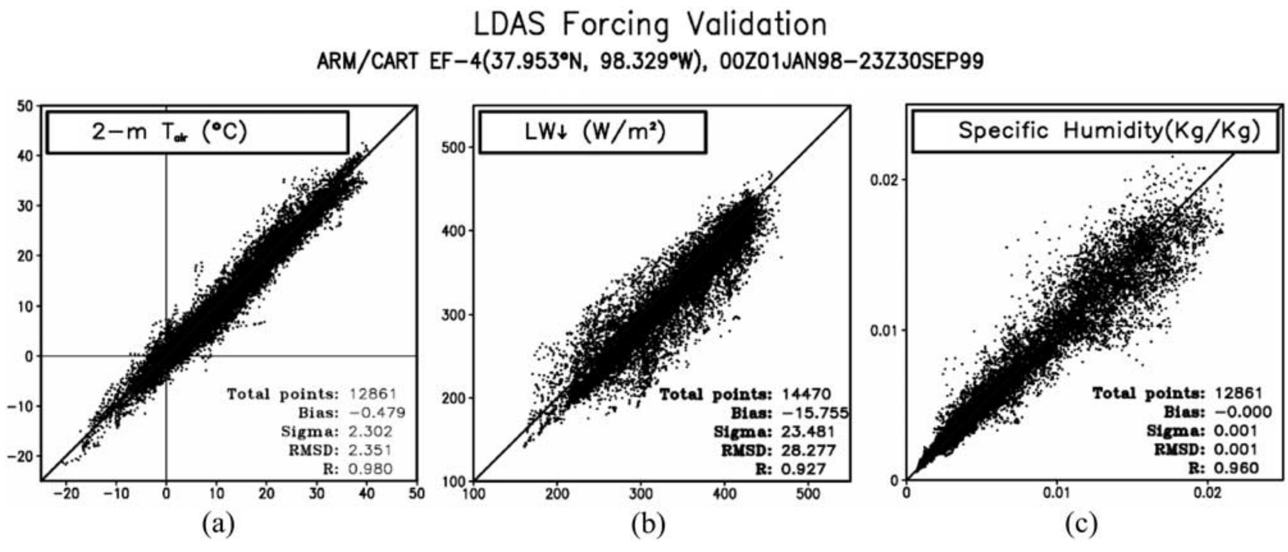

Figure 6. Validation of (a) NLDAS 2 m temperature (°C), (b) longwave radiation (W m^{-2}), and (c) specific humidity (Kg Kg^{-1}) data against ARM/CART observations from 1 January 1998 to 30 September 1999.

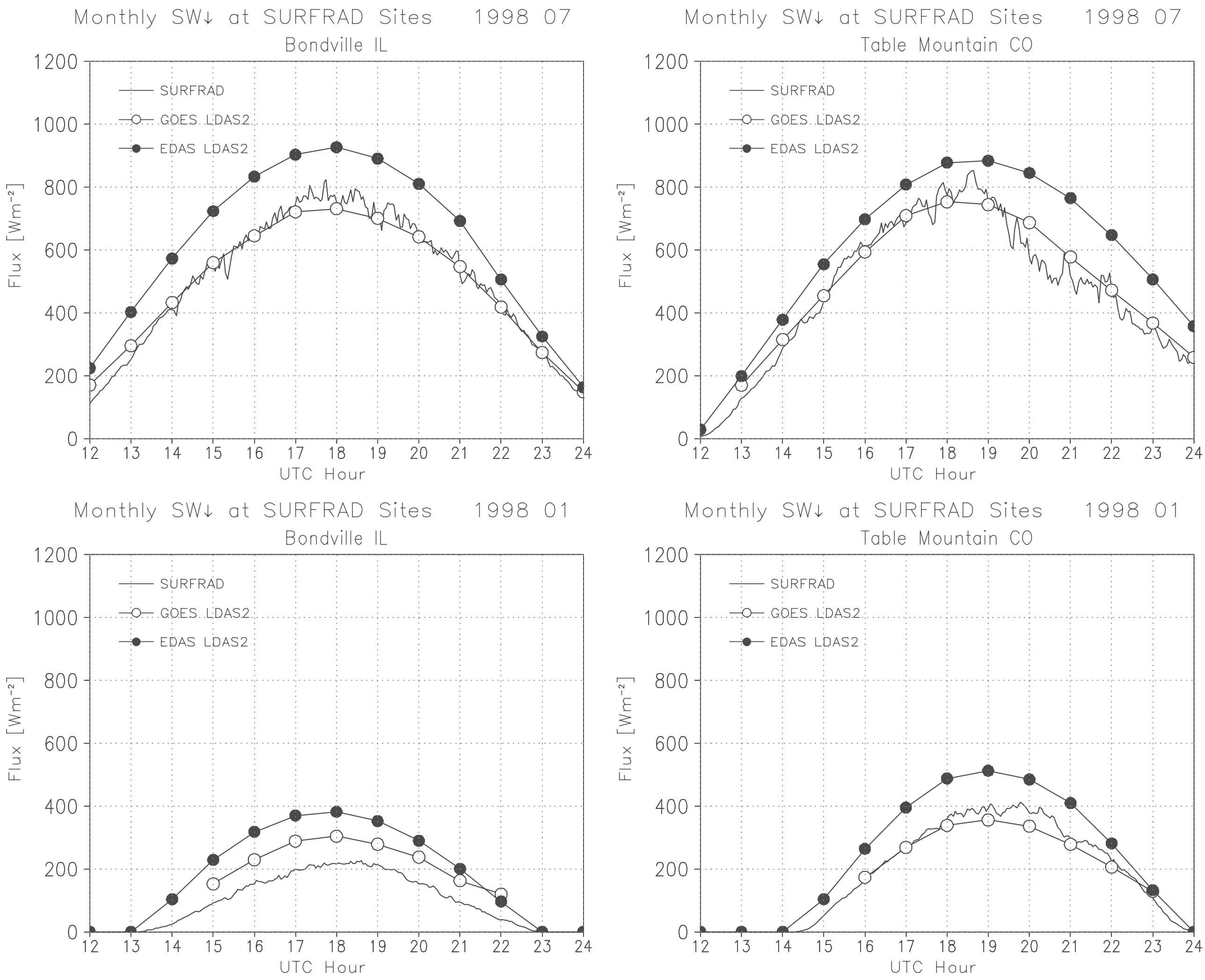

Figure 7. Validation of NLDAS GOES-based (open circles) and EDAS-based (closed circles) shortwave radiation (W m^{-2}) against SURFRAD (solid line) ground-based measurements at Bondville, Illinois, and Table Mountain, Colorado, SURFRAD sites from January and July of 1998. (bottom left) Snow cover leads to a wintertime high bias in GOES radiation values at Bondville.

while the retrospective data set can make use of reprocessed and enhanced sources of data (i.e., reprocessed CPC daily precipitation analyses), the real-time data set must rely upon only those data sources that are available in the production time window (i.e., real-time CPC precipitation analyses). In addition, the production of near real-time data can be impacted by computing and network conditions, as well as by the day-to-day availability of critical data sets. Errors can also surface in the data sets or code used to construct the NLDAS forcing data. While real-time production, in a process of continual improvement, can quickly adapt to ensure that future forcing files are not affected, any files already produced in the past with the erroneous data sets or code remain tainted. By contrast, when such problems emerge in the production of the retrospective data set, the forcing production can simply be restarted from the beginning using error free code or data.

[18] As a case in point, a calibration error was discovered in the NESDIS GOES satellite data used in the real-time production of NLDAS shortwave radiation forcing data. This error persisted from April to December of 1999, and manifested itself through excessively high solar radiation values. Upon discovery of the problem, real-time production of NLDAS forcing data was immediately adapted to make use of EDAS/Eta data in place of NESDIS data. As the problem was not discovered until July, 4 months of real-time forcing data had already been produced using the erroneous GOES data. Compounding this problem, errors in zenith angle interpolation led to a 3 hour phase shift in EDAS/Eta-based NLDAS radiation data from May 1999 to February 2000. This phase shift was fixed as soon as it was detected, but as was the case with the calibration problem above, the real-time nature of the production process meant that several months of problematic shortwave radiation data were already permanently stored in the real-time forcing data archive. Of note, since February 2000 to the present, no significant additional processing errors have been discovered in the NLDAS real-time forcing fields.

[19] Given its freedom from real-time production constraints, the NLDAS retrospective forcing data set was not

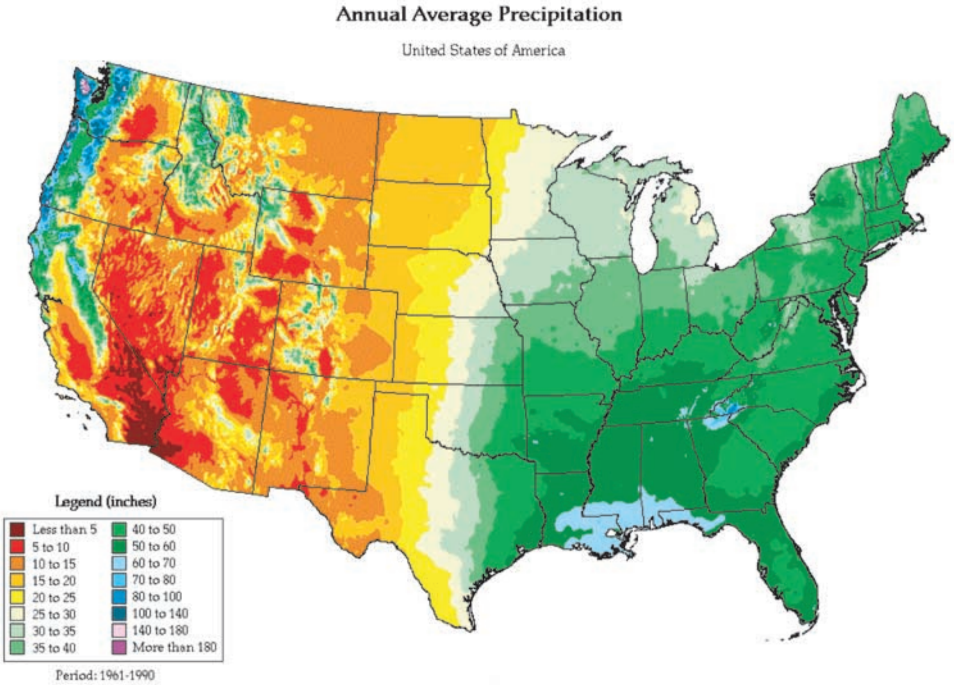

Figure 8. Annual total of monthly PRISM precipitation climatology (in) used in real-time NLDAS forcing data generation (Oregon Climate Service). Modeling was performed using the PRISM model, based on 1961–1990 normals from NOAA Cooperative stations and NRCS SNOTEL sites.

permanently affected by these same issues, and was able to take advantage of the knowledge gleaned from the real-time production to avoid the same problems. In particular, with the information that NESDIS GOES data contained a calibration error, production of retrospective data was altered to allow for the use of UMD-processed GOES data that did not feature the same calibration issue.

[20] Real-time production of data is a challenging process of continual improvement and monitoring of input real-time data streams and operations. However, compensating for the added challenges in real-time production compared to retrospective production is the ability of the real-time forcing algorithms to take advantage of recent advances in forcing-related input data. While the NLDAS retrospective data set is a static product covering 1996–2002, the dynamic NLDAS real-time data set can leverage improvements in forcing-related input as they occur in the present day. A prime example of this is the use of a 2.5 minute CONUS-wide precipitation climatology (Figure 8), known as the Parameter-Elevation Regressions on Independent Slopes Model (PRISM) data set [*Daly et al.*, 1994], in the derivation of the real-time NLDAS precipitation forcing fields. Since 1 February 2002, the PRISM climatology has been applied to daily precipitation gauge data to account for topographical influences on precipitation. It raises or lowers grid point daily precipitation values based on the ratio of the PRISM climatology at the given grid point versus that at the location of the input precipitation gauge observation. The primary net effect is to notably increase the analyzed precipitation amounts at high elevations. This addresses the concerns raised by *Pan et al.* [2003], who show that the retrospective NLDAS precipitation forcing has a 50 percent low bias over the high elevations of the predominant mountain ranges of the western CONUS.

[21] Along with the application of the PRISM-based precipitation adjustment, the daily gauge data used in the production of real-time NLDAS forcing is produced at a resolution of 0.125° with a least squares distance weighting scheme. This stands in contrast to the Cressman analysis scheme used to produce the reprocessed CPC daily gauge product, and reduces the tendency toward overly smooth precipitation fields. Validation of this product against precipitation observations withheld from the analysis scheme has shown that its representation of daily precipitation is superior to that seen in the non-PRISM Cressman-based CPC product.

[22] Near-term improvements in the accuracy and resolution of other forcing-related fields can also be expected, and the flexible nature of the real-time data set ensures that it is well situated to take advantage of such improvements as they occur. Among such changes is an expected increase in resolution of GOES-based shortwave radiation fields from 0.5° to 0.125°. Additionally, within the next year, data from the native Eta computational grid (currently 12 km) will be directly interpolated to the 1/8th NLDAS grid, bypassing the intermediate 40 km AWIPS 212 grid. The real-time data set will incorporate each of these advances as they occur.

4. Data Archiving and Dissemination

[23] Integral aspects of NLDAS real-time and retrospective data set activities include data archiving and dissemination. In order to generate the multiyear retrospective forcing data set, NASA GSFC assembled an archive of

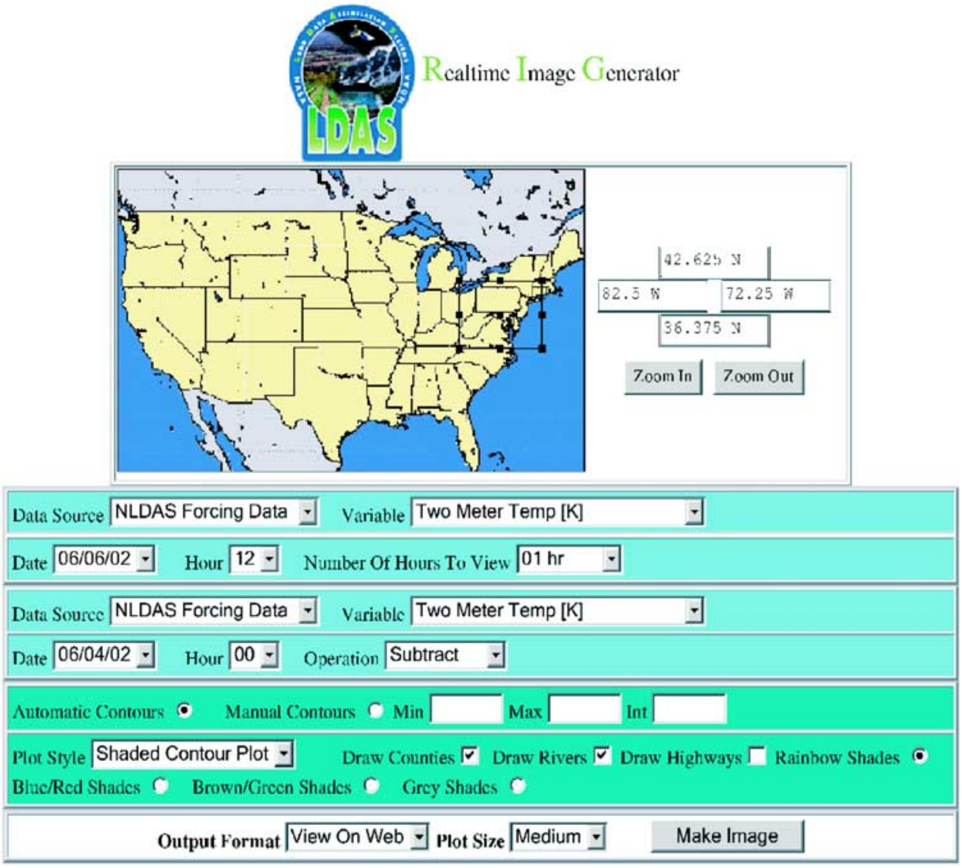

Figure 9. Real-Time Image Generator featured on LDAS web site (http://ldas.gsfc.nasa.gov). The interface allows users to visualize and download LDAS forcing data and model output.

the raw data sets needed to create the NLDAS forcing product. Stretching from 1996 to the present time, these data sets include 3 hourly EDAS and ETA data, 6 hourly ETA data, hourly GOES radiation and Doppler precipitation data, and daily CPC rainfall data. Updated automatically once per day with C shell download scripts, the data archive contains unaltered raw data files, allowing for the possible creation of new forcing files in the future utilizing improved interpolation procedures or different combinations of input data. Similar collections of C shell scripts allow NOAA NCEP to collect the data it needs to produce the NLDAS real-time forcing once per day. Retrospective forcing is produced and archived at NASA GSFC, while real-time forcing is produced and archived at NOAA NCEP.

[24] Taken together, the real-time and retrospective products are unparalleled collections of data which are extremely useful to the scientific community, and which can find application in areas ranging from flood management simulations to flux boundary simulations. In order to effectively distribute the forcing data sets to the user community, the NLDAS project has adopted a three-pronged approach: (1) FTP of complete data sets, (2) FTP of data subsets, and (3) Online visualization of data subsets. Users who require access to the full NLDAS forcing data sets can download such data through the LDAS web site (http://ldas.gsfc.nasa.gov). Here, they can download tarred monthly forcing files that average 1.5 gigabytes in size. Those users who only require a few forcing variables, or who wish to obtain temporal or spatial subsets of the NLDAS forcing data, can make use of the data subset tool on the LDAS web site. With this, they may subset data as desired, and ftp the requested information as needed. And finally, for those who wish only to visualize and analyze the data, and not download the raw files, the LDAS web site features a Real-Time Image Generator (Figure 9). This flexible, web-based tool allows users to select any variable or time period of interest in the NLDAS data archives, and enables them to visualize data, compare fields and download images.

5. Conclusion

[25] Land surface models depend heavily upon accurate forcing data in order to produce realistic simulations of land surface processes. With this in mind, the NLDAS project has produced real-time and retrospective 0.125°, hourly forcing data sets. Supporting the real-time and retrospective modeling efforts of the project, each data set uses EDAS/Eta model output as a data backbone, and incorporates observation-based radiation and precipitation data when available. The incorporation of these two observation-based variables is of great importance, for though they exert a large influence on surface processes, NWP models, such as the Eta model, find them particularly difficult to represent accurately. Each data set has been quality controlled based on ALMA forcing data standards, and each is available

from the LDAS web site in either full or subset form through FTP and a Real-Time Image Generator. Along with its constituent, raw data sets, the retrospective data set extends from 1996 to 2002, and will be lengthened to cover 2003 in upcoming months. Overlapping a portion of this coverage, the real-time forcing data set extends from 1999 to the present. Updated once per day, it is dynamic in nature, and will incorporate advances in forcing-related data sets as they occur. Improvements may include increases in the backbone EDAS/Eta data resolution, advances in the treatment of precipitation interpolation, and increases in the spatial extent and resolution of GOES-based data, and will maximize the accuracy and utility of the NLDAS forcing data.

[26] **Acknowledgments.** The work on this project by NASA/GSFC/HSB was supported by NASA's Terrestrial Hydrology Program (P. Houser, PI). The work by NCEP/EMC, NWS/OHD, and NESDIS/ORA was supported by the NOAA OGP grant for the NOAA Core Project for GCIP/GAPP (co-PIs K. Mitchell, J. Schaake, D. Tarpley). The work by Rutgers University was supported by NOAA OGP GAPP grant GC99-443b (A. Robock, PI), the Cook College Center for Environmental Prediction, and the New Jersey Agricultural Experiment Station, and additionally, Figures were drawn with GrADS, created by Brian Doty. The work by Princeton was supported by NOAA OGP GAPP grant NA86GP0258 (E. Wood, PI). The work by NCEP/CPC was supported by NOAA/NASA GAPP Project 8R1DA114 (R. W. Higgins, PI). The work by University of Maryland was supported by grants NA56GPO233, NA86GPO202 and NA06GPO404 from NOAA/OGP and by NOAA grant NA57WC0340 to University of Maryland's Cooperative Institute for Climate Studies (R. Pinker, PI). We thank SNOTEL for the snow station data, USGS for the streamflow data, DoE for the ARM/CART meteorological and heat flux data, and NOAA for the SURFRAD data that were provided to the project at no cost, and the NOAA Office of Global Programs and NASA Land surface Hydrology Program for their purchase of the Oklahoma Mesonet meteorological and soil moisture and temperature data for their funded investigators.

References

Baldwin, M., and K. E. Mitchell, The NCEP hourly multi-sensor U.S. precipitation analysis for operations and GCIP research, in *Preprints, 13th AMS Conference on Hydrology*, pp. 54–55, Am. Meteorol. Soc., Boston, Mass., 1997.

Betts, A. K., F. Chen, K. E. Mitchell, and Z. I. Janjic, Assessment of the land surface and boundary layer models in two operational versions of the NCEP Eta model using FIFE data, *Mon. Weather Rev.*, *125*, 2896–2916, 1997.

Burnash, R. J. C., R. L. Ferral, and R. A. McGuire, A generalized streamflow simulation system—Conceptual modeling for digital computers, report, Joint Fed. and State River Forecast Cent. U.S. Natl. Weather Serv./Calif. State Dept. of Water Resour., Sacramento, Calif., 1973.

Daly, C., R. P. Neilson, and D. L. Phillips, A statistical-topographic model for mapping climatological precipitation over mountainous terrain, *J. Appl. Meteorol.*, *33*, 140–158, 1994.

Fulton, R. A., J. P. Breidenbach, D. J. Seo, D. A. Miller, and T. O'Bannon, The WSR-88D rainfall algorithm, *Weather Forecasting*, *13*, 377–395, 1998.

Higgins, R. W., W. Shi, E. Yarosh, and R. Joyce, Improved United States precipitation quality control system and analysis, *NCEP/Climate Prediction Center Atlas*, vol. 7, 40 pp., U.S. Dept. of Commerce/Natl. Weather Serv. Natl. Oceanic and Atmos. Admin., Camp Springs, Md., 2000.

Koster, R. D., and M. J. Suarez, Energy and water balance calculations in the Mosaic LSM, *Tech. Mem. 104606*, *9*, NASA, Greenbelt, Md., 1996.

Liang, X., E. F. Wood, and D. Lettenmaier, Surface and soil moisture parameterization of the VIC-2L model: Evaluation and modifications, *Global Planet. Change*, *13*, 195–206, 1996.

Luo, L., et al., Validation of the North American Land Data Assimilation System (NLDAS) retrospective forcing over the southern Great Plains, *J. Geophys. Res.*, *108*(D22), 8843, doi:10.1029/2002JD003246, in press, 2003.

Mitchell, K. E., et al., GCIP Land Data Assimilation System (LDAS) project now underway, *Gewex News*, *9*(4), 3–6, 1999.

Pan, M., et al., Snow process modeling in the North American Land Data Assimilation System (NLDAS): 2. Evaluation of model-simulated snow water equivalent, *J. Geophys. Res.*, *108*(D22), 8850, doi:10.1029/2003JD003994, in press, 2003.

Pinker, R. T., et al., Surface radiation budgets in support of the GEWEX Continental Scale International Project (GCIP) and the GEWEX Americas Prediction Project (GAPP), including the North American Land Data Assimilation System (NLDAS) project, *J. Geophys. Res.*, *108*(D22), 8844, doi:10.1029/2002JD003301, in press, 2003.

Robock, A., et al., Evaluation of the North American Land Data Assimilation System over the southern Great Plains during the warm season, *J. Geophys. Res.*, *108*(D22), 8846, doi:10.1029/2002JD003245, in press, 2003.

Rogers, E., T. L. Black, D. G. Deaven, G. J. DiMego, Q. Zhao, M. Baldwin, N. W. Junker, and Y. Lin, Changes to the operational "Early" Eta analysis/forecast system at the National Centers for Environmental Prediction, *Weather Forecasting*, *11*, 391–413, 1996.

Stackpole, J. D., A guide to GRIB: The World Meteorological Organization form for the storage of weather product information and the exchange of weather product messages in gridded binary form, 1st ed., *Off. Note 388*, Natl. Meteorol. Cent. Natl. Weather Serv. Natl. Oceanic and Atmos. Admin., Camp Springs, Md., Feb. 1994.

B. A. Cosgrove, Hydrological Sciences Branch, NASA Goddard Space Flight Center, Mail Code 974.1, Greenbelt, MD 20771, USA. (brian.cosgrove@gsfc.nasa.gov)

Q. Duan and J. C. Schaake, Office of Hydrologic Development, NOAA National Weather Service, SSMC2, W/OHD12, 1325 East-West Highway, Silver Spring, MD 20910, USA. (qingyun.duan@noaa.gov; john.schaake@noaa.gov)

R. W. Higgins, Environmental Modeling Center (W/NP2), National Centers for Environmental Prediction, NOAA Science Center, 5200 Auth Road, Camp Springs, MD 20746-4304, USA. (wayne.higgins@noaa.gov)

P. R. Houser, Hydrological Sciences Branch, NASA Goddard Space Flight Center, Mail Code 974, Greenbelt, MD 20771, USA. (paul.r.houser@nasa.gov)

D. Lohmann, J. Meng, and K. E. Mitchell, Environmental Modeling Center (W/NP2, Room 207), National Centers for Environmental Prediction, NOAA Science Center, 5200 Auth Road, Camp Springs, MD 20746-4304, USA. (dlohmann@ncep.noaa.gov; jesse.meng@noaa.gov; kenneth.mitchell@noaa.gov)

L. Luo, J. Sheffield, and E. F. Wood, Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ 08544, USA. (lluo@rutgers.edu; justin@princeton.edu; efwood@princeton.edu)

C. Marshall, Department of Atmospheric Sciences, Colorado State University, Fort Collins, CO 80523, USA. (curtis@atmos.colostate.edu)

R. T. Pinker, Department of Meteorology, Space Sciences Building, University of Maryland, College Park, MD 20742, USA. (pinker@atmos.umd.edu)

A. Robock, Department of Environmental Sciences, Rutgers University, 14 College Farm Road, New Brunswick, NJ 08901-8551, USA. (robock@envsci.rutgers.edu)

J. D. Tarpley, Office of Research, NOAA/NESDIS, World Weather Building, Room 711, 5200 Auth Road, Camp Springs, MD 20746-4304, USA. (dan.tarpley@noaa.gov)