

Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2):

1. Intercomparison and application of model products

Youlong Xia,^{1,2} Kenneth Mitchell,¹ Michael Ek,¹ Justin Sheffield,³ Brian Cosgrove,⁴ Eric Wood,³ Lifeng Luo,⁵ Charles Alonge,⁶ Helin Wei,^{1,2} Jesse Meng,^{1,2} Ben Livneh,⁷ Dennis Lettenmaier,⁷ Victor Koren,⁴ Qingyun Duan,⁸ Kingtse Mo,⁹ Yun Fan,¹⁰ and David Mocko¹¹

Received 31 March 2011; revised 15 December 2011; accepted 15 December 2011; published 3 February 2012.

[1] Results are presented from the second phase of the multiinstitution North American Land Data Assimilation System (NLDAS-2) research partnership. In NLDAS, the Noah, Variable Infiltration Capacity, Sacramento Soil Moisture Accounting, and Mosaic land surface models (LSMs) are executed over the conterminous U.S. (CONUS) in realtime and retrospective modes. These runs support the drought analysis, monitoring and forecasting activities of the National Integrated Drought Information System, as well as efforts to monitor large-scale floods. NLDAS-2 builds upon the framework of the first phase of NLDAS (NLDAS-1) by increasing the accuracy and consistency of the surface forcing data, upgrading the land surface model code and parameters, and extending the study from a 3-year (1997–1999) to a 30-year (1979–2008) time window. As the first of two parts, this paper details the configuration of NLDAS-2, describes the upgrades to the forcing, parameters, and code of the four LSMs, and explores overall model-to-model comparisons of land surface water and energy flux and state variables over the CONUS. Focusing on model output rather than on observations, this study seeks to highlight the similarities and differences between models, and to assess changes in output from that seen in NLDAS-1. The second part of the two-part article focuses on the validation of model-simulated streamflow and evaporation against observations. The results depict a higher level of agreement among the four models over much of the CONUS than was found in the first phase of NLDAS. This is due, in part, to recent improvements in the parameters, code, and forcing of the NLDAS-2 LSMs that were initiated following NLDAS-1. However, large inter-model differences still exist in the northeast, Lake Superior, and western mountainous regions of the CONUS, which are associated with cold season processes. In addition, variations in the representation of sub-surface hydrology in the four LSMs lead to large differences in modeled evaporation and subsurface runoff. These issues are important targets for future research by the land surface modeling community. Finally, improvement from NLDAS-1 to NLDAS-2 is summarized by comparing the streamflow measured from U.S. Geological Survey stream gauges with that simulated by four NLDAS models over 961 small basins.

¹Environmental Modeling Center, National Centers for Environment and Prediction, NOAA, Camp Springs, Maryland, USA.

²IMSG at NCEP/EMC, Camp Springs, Maryland, USA.

³Department of Environmental and Civil Engineering, Princeton University, Princeton, New Jersey, USA.

⁴Office of Hydrologic Development, National Weather Service, NOAA, Silver Spring, Maryland, USA.

⁵Department of Geography, Michigan State University, East Lansing, Michigan, USA.

Copyright 2012 by the American Geophysical Union.
0148-0277/12/2011JD016048

⁶AWS Truewind, LLC, Albany, New York, USA.

⁷Department of Environmental and Civil Engineering, University of Washington, Seattle, Washington, USA.

⁸College of Global Change and Earth System Science, Beijing Normal University, Beijing, China.

⁹Climate Prediction Center, National Centers for Environment and Prediction, NOAA, Camp Springs, Maryland, USA.

¹⁰Office of Science and Technology, National Weather Service, Silver Spring, Maryland, USA.

¹¹SAIC at the Hydrological Sciences Laboratory and Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.

Citation: Xia, Y., et al. (2012), Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products, *J. Geophys. Res.*, 117, D03109, doi:10.1029/2011JD016048.

1. Introduction

[2] The first phase of the multiinstitution North American Land Data Assimilation System (NLDAS-1) [Mitchell *et al.*, 2004] project was initiated to provide reliable initial land surface states to coupled atmosphere-ocean-land models in an effort to improve weather predictions. The ability of land surface models (LSMs) to accurately reproduce observed water and energy fluxes is an important cornerstone in the development of land data assimilation systems for providing such initial states. Toward this end, the NLDAS-1 team ran the Noah [Ek *et al.*, 2003], Mosaic [Koster and Suarez, 1996], Sacramento Soil Moisture Accounting (SAC-SMA) [Burnash *et al.*, 1973; Burnash, 1995], and Variable Infiltration Capacity (VIC) [Liang *et al.*, 1994] models within a common modeling framework from the period 1996–2005. They generated real-time and retrospective land states along with surface water and energy fluxes over a 1/8th degree conterminous U.S. (CONUS) domain. The simulations and accompanying evaluation demonstrated the potential of such a system for generating reasonable depictions of the hydrologic cycle, and for providing land surface initial conditions in real-time. The research partnership supporting this effort included participants from the National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Prediction (NCEP) Environmental Modeling Center (EMC), the National Aeronautics and Space Administration's (NASA) Goddard Space Flight Center (GSFC), the NOAA National Weather Service (NWS) Office of Hydrologic Development (OHD), the NOAA Environmental Satellite, Data, and Information Service (NESDIS), the NOAA Climate Prediction Center (CPC), and several universities including Princeton University, the University of Washington, the University of Maryland, and Rutgers University.

[3] The NLDAS-1 multimodel analysis also proved valuable by revealing a number of regional biases stemming from errors in the surface forcing data and shortcomings in the model parameterizations and parameter values. Under funding from the NOAA Climate Prediction Program of the Americas (CPPA), the NOAA NCEP EMC land team is spearheading a second phase of NLDAS (NLDAS-2) to leverage the infrastructure and findings of NLDAS-1 to provide improved forcing data and LSM simulations. A major focus of NLDAS-2 is to generate land surface products for decision support applications, with particular emphasis placed on the support of the drought monitoring and forecasting efforts of the National Integrated Drought Information System (NIDIS, <http://www.drought.gov>) and on the analysis of large-scale flood events.

1.1. Summary of NLDAS-1

[4] Phase two of the NLDAS project has its roots in NLDAS-1, and much can be learned from an examination of the results of the original project. The development and evaluation of the NLDAS-1 land surface products were described in a series of papers that formed a special issue of this journal.

An overview of the system was given by Mitchell *et al.* [2004], and Cosgrove *et al.* [2003a] described the development of the real-time and retrospective forcing data. Luo *et al.* [2003] used observed forcing data at 72 Oklahoma Mesonet stations [Brock *et al.*, 1995] and 24 Atmospheric Radiation Measurement/Cloud and Radiation Testbed stations (ARM/CART) [Stokes and Schwartz, 1994] over the Southern Great Plains to evaluate NLDAS-1 downward solar radiation, downward long wave radiation, 10-m wind speed, specific humidity, 2 m air temperature, surface pressure, and precipitation. The results indicated good agreement between NLDAS forcing data and observations for all meteorological variables except for hourly precipitation.

[5] In a complementary study, Robock *et al.* [2003] employed observations from sites across the Southern Great Plains to evaluate the performance of the four LSMs in terms of simulating warm season diurnal cycles and the regional distribution of energy fluxes, soil moisture, and soil temperature. Although the four models captured the regional spatial distribution of surface energy fluxes and states, as well as the phase of their mean diurnal cycles, there were still sizable differences between the models and the observations, especially in the partitioning of net radiation into sensible and latent heat. Focusing on the cold season, Sheffield *et al.* [2003] used satellite-measured snow cover extent to assess the NLDAS-1 simulations, and found that Mosaic and Noah underestimated snow cover extent by over 20%, VIC overestimated snow cover content by over 20%, and SAC-SMA's companion snow model, Snow17 [Anderson, 1973], was unbiased on average. The special issue also contains an article by Pan *et al.* [2003], who used snow water equivalent (SWE) from approximately 600 Snowpack Telemetry (SNOWTELE) stations to show that all four models underestimated observed SWE due to overly low NLDAS-1 snowfall forcing data. A related issue was uncovered by Lohmann *et al.* [2004], who evaluated modeled streamflow against observations from 1145 U.S. Geological Survey (USGS) small basin stream gauges. They found that all four models underestimated observed streamflow over mountainous regions, likely due to the lack of consideration in the NLDAS-1 forcing data of the topographic enhancement of precipitation and the under-catch of snowfall at precipitation gauges. Besides biases in the precipitation data, other errors in the forcing data also negatively impacted NLDAS-1 model simulations. For example, the blending of downward solar radiation from the Eta Data Assimilation System (EDAS) with observation-based values from the Geostationary Operational Environmental Satellite (GOES) resulted in spatial discontinuities along the boundaries of the GOES product.

1.2. NLDAS-1 Legacy and Impetus for NLDAS-2

[6] Although the validation of NLDAS-1 output products such as skin temperature, soil moisture, soil temperature, snow cover, snow water equivalent, and streamflow found generally good agreement between observations and model

simulations, it also revealed large biases in individual models for particular hydrological processes. As one example, the Noah model greatly underestimated snow water equivalent, and generated overly early snowmelt over mountainous regions. This resulted in a mismatch between daily and monthly observed and simulated streamflow. Such findings encouraged the land modeling community to improve the Noah model. *Slater et al.* [2007] ran a multimodel simulation of pan-arctic hydrology using the ERA-40 reanalysis to force five LSMs, including the Noah and VIC models. They found that SWE output by the Noah model was significantly smaller than that of the other models. This was traced to excessive sublimation in the wintertime driven by an overly large turbulent exchange coefficient (aero-dynamic exchange coefficient). In the real world, a stable boundary layer generally exists in the wintertime in the Arctic and mountainous regions, and therefore the turbulent exchange coefficient is small. *Slater et al.* [2007] imposed a limitation on this coefficient which led to improved Noah LSM output.

[7] Cold-season Noah LSM simulations were also the subject of a study by *Feng et al.* [2008], who found that the Noah model significantly underestimated SWE at three Cold Land Processes Field Experiment (CLPX) sites. This underestimation was mainly due to low maximum snow albedo values that resulted in the input of too much energy into the snowpack, enhanced sublimation, and early snowmelt. The same conclusion was found by *Livneh et al.* [2010]. Updates to the Noah snow sub-model were implemented as a result of these findings and were included in the NLDAS-2 simulations. In preparation for NLDAS-2, the VIC model was also improved through the tuning of model parameters, and SAC-SMA operations in NLDAS were improved through the use of climatologically based (versus Noah-based) potential evaporation forcing data.

[8] NLDAS-1 was limited to three years of retrospective forcing data, and because the first year was used as a model spin-up period, only two years (from October 1, 1997 to September 30, 1999) of model simulations were available for validation and analysis. This length of output is not enough to calculate stable model climatologies and generate hydrological anomalies and percentiles which can be used to form the basis of drought and flood indices. To address this need, NLDAS research partners at NASA extended the forcing data to over 30 years, from January 1, 1979 to the current time (http://www.emc.ncep.noaa.gov/mmb/nldas/LDAS8th/forcing/forcing_narr.shtml). The 30+ years of forcing data were used with the updated configurations of the LSMs described above to produce 30 years of CONUS-wide water and energy cycle products over the same domain, and with the same temporal and spatial resolutions, as those used in NLDAS-1.

[9] Output from these simulations is analyzed below in the first part of a two-part NLDAS-2 paper. Focusing on the intercomparison rather than validation of model output, this first study seeks to highlight the similarities and differences between models, and to assess changes in output from that seen in NLDAS-1. Specifically, we report on the overall model-to-model comparison of results for a 28-year retrospective period (1 October 1979–30 September 2007). Section 2 of the paper describes the general configuration of NLDAS-2, the upgrades made to some of the NLDAS-2

LSMs, and the spin-up of land surface model states. Section 3 provides a model-to-model comparison of water and energy balances, and section 4 analyzes the similarity of the NLDAS-2 model simulations. Section 5 summarizes the improvement from NLDAS-1 to NLDAS-2 by comparing observed and simulated streamflow at 961 small basins across the CONUS. The application of NLDAS-2 model products for drought and large-scale flood monitoring is discussed in section 6. Finally, section 7 provides a summary and conclusions. This text is complemented by the second part of this two-part article [*Xia et al.*, 2012], which focuses, by contrast, on the validation of model-simulated streamflow and evaporation against observations.

2. NLDAS-2 Configuration and Models

2.1. General Configuration

[10] The configuration of NLDAS-2 is very similar to the NLDAS-1 configuration presented in detail by *Mitchell et al.* [2004], and so here only a brief summary is given. NLDAS-2 is an offline data assimilation system featuring uncoupled land surface models which are driven by observation-based atmospheric forcing. With support from the same group of partners as in NLDAS-1, four LSMs (Noah, SAC-SMA, VIC, and Mosaic) are executed over the NLDAS domain with a 1/8° latitude-longitude resolution. Noah and Mosaic operate at a 15 min computational time step, while SAC-SMA and VIC operate at a 60-min time step. The domain covers the conterminous United States (CONUS), the southern part of Canada, and the northern portion of Mexico (125° to 67° W, 25° to 53° N). The majority of NLDAS atmospheric forcing data is derived from the North American Regional Reanalysis (NARR) which features a 32-km spatial resolution and a three-hour temporal resolution. NARR-based variables include 2-m air temperature, 2-m specific humidity, 10-m wind speed, surface pressure, precipitation, incoming solar radiation, and incoming longwave radiation. NLDAS software is used to interpolate the coarse-resolution NARR data to the finer-scale 1/8th degree NLDAS grid and to the one-hour NLDAS temporal resolution.

[11] Several sources of observed data supplement the model-based forcing fields. The CPC unified gauge-based precipitation analysis with monthly Parameter-elevation Regressions on Independent Slopes Model (PRISM) [*Daly et al.*, 1994] adjustments for orographic precipitation impacts is used in place of NARR precipitation over the CONUS whenever possible. The adjusted daily CPC unified gauge-based precipitation data are disaggregated to an hourly time-scale using precipitation estimates from NCEP Stage-II Doppler radar data. Since the NARR assimilates precipitation gauge data, the merged CPC-NARR-based precipitation forcing field is relatively seamless [*Mesinger et al.*, 2006]. Because of the sparseness of gauge-based precipitation data in parts of Canada and Mexico, NARR precipitation is used in these regions. All models use a threshold air temperature of 0°C to partition the precipitation inputs, such that if the air temperature is above this value then the precipitation is considered to be rainfall, and is considered to be snowfall if the air temperature is below this value. NARR downward shortwave radiation features a large positive bias, and so satellite-derived downward shortwave radiation [*Pinker*

Table 1. A Full Description of NLDAS-2 Output Variables for the Four NLDAS-2 LSMs

Variable	Noah	Mosaic	Snow17 ^a	VIC
Net shortwave radiation (W/m ²)	X	X		X
Net longwave radiation (W/m ²)	X	X		X
Latent heat flux (W/m ²)	X	X		X
Sensible heat flux (W/m ²)	X	X		X
Ground heat flux (W/m ²)	X	X		X
Snow phase change heat flux (W/m ²)	X	X		X
Downward shortwave radiation (W/m ²)	X	X		X
Downward longwave radiation (W/m ²)	X	X		X
Snowfall (mm/h)	X	X	X	X
Rainfall (mm/h)	X	X	X	X
Total evapotranspiration (mm/h)	X	X	X	X
Surface runoff (mm/h)	X	X	X	X
Subsurface runoff (mm/h)	X	X	X	X
Snowmelt (mm/h)	X	X	X	X
Surface skin temperature (K)	X	X		X
Surface albedo (%)	X	X		X
Snow water equivalent (mm)	X	X	X	X
Plant canopy surface water (mm)	X	X		X
Layer 1 soil temperature (K)	X			X
Layer 2 soil temperature (K)	X			X
Layer 3 soil temperature (K)	X			X
Layer 4 soil temperature (K)	X			X
Total soil moisture (mm)	X	X	X	X
Top 1 m soil moisture (mm)	X	X	X	X
Root zone soil moisture (mm)	X	X	X	X
Layer 1 soil moisture (mm)	X	X	X ^b	X
Layer 2 soil moisture (mm)	X	X	X ^b	X
Layer 3 soil moisture (mm)	X	X	X ^b	X
Layer 4 soil moisture (mm)	X		X ^b	
Layer liquid soil moisture (mm)	X			
Layer 2 liquid soil moisture (mm)	X			
Layer 3 liquid soil moisture (mm)	X			
Layer 4 liquid soil moisture (mm)	X			
Total column soil wetness (%)	X	X		X
Root zone soil wetness (%)	X	X		X
Canopy water evaporation (W/m ²)	X	X		X
Canopy transpiration (W/m ²)	X	X		X
Bare soil evaporation (W/m ²)				
Sublimation (W/m ²)	X	X		X
Aerodynamic conductance (m/s)	X	X		X
Leaf area index (-)	X	X		X
Snow depth (mm)	X	X	X	X
Snow cover (%)	X	X	X	X
Potential evaporation (mm/h)	X			X
Vegetation cover fraction (%)	X	X		

^aRun with SAC-SMA.^bVariable derived via post-processing.

et al., 2003] is used to adjust the NARR data over each grid cell using the ratio of their monthly average diurnal cycle. Employing enhanced treatment of boundaries, this approach removes both the shortwave spatial discontinuity problem noted in NLDAS-1 and the positive bias which characterizes the NARR data.

[12] The four NLDAS-2 LSMs employ the following common configuration and parameters over the NLDAS domain: a 1/8th degree regular latitude-longitude grid, land mask, terrain elevation, soil texture, vegetation classes and distribution, streamflow network, routing model, and GRIB input and output file format [Mitchell et al., 2004]. They also draw, as appropriate, from the single set of NLDAS-2 forcing variables described above. Depending on each model's capabilities, output fields include such states as soil moisture, soil temperature, and snow water equivalent, and surface fluxes such as latent, sensible, and ground heat flux, and

runoff. A full description of output variables for the four models is given in Table 1. Although common fields of vegetation and soil classes are made available to all four models, they use their own vegetation and soil parameter values such as root depth and density, their own number and thickness of soil layers, and their own seasonal cycle of vegetation characteristics. This is to ensure that the positive legacy of calibration or tuning performed with each model over the past many years is retained.

2.2. Land Surface Models and Routing Scheme

[13] Four LSMs are currently implemented in NLDAS-2. These models represent different approaches to land surface modeling. The Mosaic and Noah models grew from the legacy of surface-vegetation-atmosphere transfer (SVAT) schemes within the coupled climate modeling community. Serving as land components, they were coupled to global climate and regional weather models. Therefore, during their development, an emphasis was placed on the accuracy of the water and energy exchanges between the land surface and the atmospheric boundary through evapotranspiration, latent heat, and sensible heat. By contrast, VIC and SAC-SMA were developed within the hydrological community as uncoupled models. They have been widely used in hydrological simulations at various spatial scales ranging from a single-point to CONUS-wide. Given their heritage, the simulation of streamflow has been the main focus during their development and calibration. Over the past decade, the VIC model has further developed to account for the full energy balance based on the SVAT concept, and SAC-SMA has evolved to include SVAT-type evapotranspiration physics from the Noah LSM [Koren et al., 2007]. In a reciprocal development, research is underway to implement portions of the SAC-SMA and/or VIC hydrological modeling schemes into the Noah LSM. An initial effort to improve streamflow simulations along these lines is underway and detailed by Livneh et al. [2011].

[14] All four models have been implemented in both uncoupled and coupled modes at various spatial scales. They have participated in model intercomparison projects at local to regional scales in the Project for the Intercomparison of Land-Surface Parameterization Schemes (PILPS) [T. H. Chen et al., 1997; Wood et al., 1998; Schlosser et al., 2000; Bowling et al., 2003; Nijssen et al., 2003], and at global scales in the Global Soil Wetness Project (GSWP) [Dirmeyer et al., 1999, 2006]. Formulated as a conceptual hydrological model, SAC-SMA was originally executed in a lumped fashion, but has since been converted into a distributed version which is described in detail below. Noah and SAC-SMA have been tested in phases 1 and 2 of the Distributed Model Intercomparison Project (DMIP) [Smith et al., 2004, 2012].

[15] The Noah model was developed as the land component of the NOAA NCEP mesoscale Eta model [Betts et al., 1997; F. Chen et al., 1997; Ek et al., 2003]. It serves as the land component in the evolving Weather Research and Forecasting (WRF) regional atmospheric model, the NOAA NCEP coupled Climate Forecast System (CFS) and the Global Forecast System. The model simulates the soil freeze-thaw process and its impact on soil heating/cooling and transpiration, following Koren et al. [1999]. The model has four soil layers with spatially invariant thicknesses of 10, 30, 60 and 100 cm. The first three layers form the root zone

in non-forested regions, with the fourth layer included in forested regions. Recent upgrades to the Noah LSM which are utilized in NLDAS-2 are detailed in the next section.

[16] The Mosaic model was developed for use in NASA's global climate model [Koster and Suarez, 1994, 1996; Koster et al., 2000]. It is a SVAT scheme that accounts for the sub-grid heterogeneity of vegetation and soil moisture with a tiling approach. Up to 10 tiles can be used in the current configuration of Mosaic. Each vegetation tile calculates its own energy and water balance, and soil moisture and temperature. Tiles have three soil layers with thicknesses of 10, 30 and 160 cm, the first two of which fall within the root zone. The water storage in each layer of the soil column is calculated as a weighted average of the water storage from all tiles.

[17] The VIC model was developed at the University of Washington and Princeton University as a macroscale, semi-distributed, grid-based, hydrologic model [Liang et al., 1994; Wood et al., 1997]. VIC can be executed in several modes, with the full water and energy balance version chosen for NLDAS-1 and NLDAS-2. VIC features three soil layers, with a 10 cm top layer and spatially varying thicknesses for layers two and three. The root zone can span all three layers and depends on the vegetation type and its associated vertical root distribution. Like the Mosaic model, the VIC model utilizes sub-grid vegetation tiles. The VIC model includes a two-layer energy balance snow model [Cherkauer et al., 2003], which represents snow accumulation and ablation on the ground and in the forest canopy, and which uses sub-grid elevation bands to represent the impact of elevation on temperature, precipitation, and snow. VIC has been widely applied to large river basins in the U.S. [Nijssen et al., 1997; Abdulla et al., 1996; Cherkauer and Lettenmaier, 1999] and elsewhere [Lohmann et al., 1998; Lobmeyr et al., 1999], at national [Maurer et al., 2002] and global scales [Nijssen et al., 2001; Sheffield et al., 2006].

[18] Used operationally at NWS River Forecast Centers, the SAC-SMA model began as a lumped conceptual hydrological model [Burnash et al., 1973; Burnash, 1995]. Seeking to increase the applicability and usability of SAC-SMA in both operational and research environments, OHD developed a distributed version and an accompanying set of nationwide a priori model parameters [Koren et al., 2000, 2004]. The main benefit of these parameters is that they eliminate the necessity of deriving a calibrated set of parameters and facilitate large-scale and geographically diverse implementations of the model. While updates to these parameters have been released by OHD, NLDAS makes use of the first version which is based on relationships between vegetation cover and STATSGO soils data, and has not performed further calibration.

[19] The SAC-SMA model represents water storage using five conceptual water storage components divided into upper and lower zones, which are further separated into tension and free water storage components. Additionally, the model features a sixth variable water storage component that accounts for the effects of varying areas of saturation near streams. Together, these components represent the active part of the total water storage in a grid cell. Differing from the other NLDAS-2 models, the water storage components of the SAC-SMA model are not tied to any soil depth or thickness. This characteristic complicates intercomparison with other

models and validation against soil column observations, and was addressed as part of a recent frozen ground physics upgrade. In this upgrade, SAC-SMA gained the ability to map the conceptual water storages to distinct soil layers (SAC-Heat Transfer (HT)) [Koren et al., 2007]. This upgrade has allowed for the accurate simulation of the vertical profile of soil moisture and soil temperature. SAC-HT has been further upgraded through the inclusion of Noah LSM evapotranspiration physics (SAC-HT-Evapotranspiration (ET)), which allow for the internal computation of potential evapotranspiration [Koren et al., 2010]. While these upgraded versions were not used in the base set of NLDAS-2 simulations, they will be considered for use in future simulations. As an interim solution, soil moisture output was computed for NLDAS-2 at distinct soil layers using post-processing techniques as follows. Model parameter-soil property relationships were used to convert the upper and lower soil moisture capacities into soil moisture contents at a number of soil layers. The physically based heat transfer (HT) component was taken from SAC-HT and used to determine the distribution of liquid/frozen water in the layered soil column. Five layer depths were defined a priori to cover a 2 m soil profile with thinner layers closer to the soil surface. The layered soil moisture contents from the HT component were then interpolated by weighted averaging to the same layers as the Noah model.

[20] To more fully represent the hydrologic cycle in NLDAS-2, SAC-SMA was run together with the distributed version of the Snow-17 model. This model was also used in NLDAS-1, and is a temperature index-based snow model detailed by Anderson [1973]. As is the case with SAC-SMA, a priori model parameters have recently been developed for Snow17. However, as these parameters were not available at the time of the initial NLDAS simulations, constant nationwide values supplied by OHD were used in their place.

[21] The NLDAS-2 forcing variables ingested by SAC-SMA and Snow-17 are smaller in number than those used by the other NLDAS-2 LSMs. In particular, only precipitation and 2-m air temperature are used by Snow-17, and only potential evapotranspiration, precipitation, and Snow17 snowmelt are ingested by SAC-SMA. Furthermore, since SAC-SMA is a conceptual rainfall-runoff water balance model, it does not calculate an energy balance and outputs a list of variables which differs from that of the other LSMs.

[22] Departing from the model configuration utilized in NLDAS-1, climatologically based PE (versus Noah-based PE) was ingested into SAC-SMA for the NLDAS-2 simulations. This change was implemented based on operational experience at OHD, along with research findings at NCEP which showed the approach to be effective at reducing the model's mean annual negative runoff bias found in NLDAS-1.

[23] Output from all four models was run through a common routing scheme. Described in detail in the companion to this paper [Xia et al., 2012], the scheme is based on the linearized Saint-Venant equations, and both calculates the timing of the runoff reaching the outlet of a grid cell and transports water through the river network. Following Lohmann et al. [2004], the routing model was executed in two modes: lumped and distributed, and is the same as that utilized in the NLDAS-1 set of simulations. Water routed from cell-to-cell served as the only form of exchange between cells, since no model featured blowing/drifting

snow or other state transfers between cells. Routed streamflow output is not analyzed in this paper, and readers are referred to the companion paper [Xia *et al.*, 2012] for analysis and validation of this variable.

2.3. Land Surface Model Improvements

[24] As mentioned in preceding sections, the research findings which emerged from NLDAS-1 encouraged the LSM development community to improve output from the NLDAS LSMS through code upgrades, parameter tuning, and alterations in the source of PE forcing data. As NOAA NCEP's main operational LSM, the Noah model has been significantly improved post-NLDAS-1 by a multiinstitutional collaboration between EMC, the University of Washington, Princeton University, the NASA Hydrological Sciences Branch, OHD, and other institutions. The following upgrades have been implemented to reduce water and energy flux simulation biases in the warm season: (1) replacing the constant leaf area index (LAI) with seasonally and spatially varying values, (2) allowing a seasonally varying vertical rooting depth, for which the profile can be fixed (e.g., evergreen trees) or change from month to month (e.g., crops), (3) upgrading the parameterization for the vapor-pressure deficit term in the canopy resistance, (4) changing the minimum stomatal resistance parameter for some vegetation classes, (5) changing the upper threshold of soil moisture at which the vegetation reacts to a soil moisture deficit, (6) upgrading the depiction of the diurnal variation of surface albedo, and (7) changing the parameter formulation of the roughness length for heat to increase the daytime aerodynamic conductance.

[25] The main purpose of these changes was to improve the parameterizations related to the seasonal and diurnal simulations of water fluxes, energy fluxes and state variables. As indicated by Mitchell *et al.* [2004], increasing Noah's aerodynamic conductance value—a product of the surface turbulent exchange coefficient for heat and the wind speed—led to more reasonable land skin temperatures when compared to the GOES values. The aerodynamic conductance was increased by decreasing an adjustable parameter C_z from 0.2 to 0.05. As indicated by Mitchell *et al.* [2004], decreasing C_z increases the roughness length for heat, which increases aerodynamic conductance and the land/atmosphere coupling. Therefore, a value of 0.05 for C_z was selected for NLDAS-2 run. Further expanding upon the changes within Noah, the calculation of seasonal leaf area index (LAI) was altered to depend on the vegetation class and was rewritten as

$$LAI = LAI_{\min} + \alpha(LAI_{\max} - LAI_{\min}) \quad (1)$$

where $\alpha = \frac{f - f_{\min}}{f_{\max} - f_{\min}}$, f is monthly vegetation cover fraction, and f_{\max} and f_{\min} are the maximum and minimum vegetation cover fractions, respectively. Detailed descriptions of these upgrades and sensitivity tests can be found in the work of Wei *et al.* [2011].

[26] Along with the aforementioned upgrades to Noah's representation of warm season processes, several changes were also made to Noah's cold season physics including (1) increasing the maximum snow albedo to reduce the input of solar energy and to extend the snow season, (2) constraining snow sublimation under stable atmospheric conditions, and

(3) taking into consideration the effect of snow age on snowpack properties (e.g., albedo). The main purpose of these changes was to increase the modeled snow water equivalent and snow cover given the underestimation seen in the NLDAS-1 evaluations [Mitchell *et al.*, 2004; Sheffield *et al.*, 2004; Pan *et al.*, 2003], in recent studies from panarctic simulations, and in the CLPX experiment [Slater *et al.*, 2007; Feng *et al.*, 2008; Livneh *et al.*, 2010].

[27] Maximum snow albedo is now calculated as

$$Albedo_{\max}^{new} = Albedo_{\max}^{control} + \beta(0.85 - Albedo_{\max}^{control}) \quad (2)$$

where $Albedo_{\max}^{new}$ is the new maximum snow albedo. The value of 0.85 is the fresh snow albedo that was used in the PILPS project and which is used by VIC model. $Albedo_{\max}^{control}$ is the default maximum snow albedo derived from satellite products [Robinson and Kukla, 1985], and β is a tuning parameter. If $\beta = 1$, $Albedo_{\max}^{new} = 0.85$. If $\beta = 0$, $Albedo_{\max}^{new} = Albedo_{\max}^{control}$. From sensitivity experiments, a value of $\beta = 0.5$ was chosen for NLDAS-2 to represent the mixed maximum snow albedo. To correct the bias in snow sublimation, the value of the surface exchange coefficient, CH, was modified to reflect atmospheric boundary layer stability as quantified by the Richardson number, R_{IB} . In particular, if R_{IB} is larger than 0.0 (stable conditions), but less than or equal to 2.0, $CH = CH \times \max(1.0 - R_{IB}/0.5, 0.05)$.

[28] To better represent the effect of snowpack aging on Noah's maximum snow albedo, the parameterization used in the VIC model [Liang *et al.*, 1994] was introduced into the Noah model. The parameterization assumes that maximum snow albedo decreases with time as the snow darkens. During snow accumulation, the maximum snow albedo decreases slowly with time; during snow ablation, the albedo decreases rapidly. In this study, we use the temperature of the first soil layer (T_1) to distinguish between the accumulation and ablation phases. The albedo is calculated as

$$Albedo_{\max}^{decay} = Albedo_{\max}^{new} A^{B^t} \quad (3)$$

where t is the number of days since the last snowfall, $A = 0.94$ and $B = 0.58$ for the accumulation phase, and $A = 0.82$ and $B = 0.46$ for the ablation phase. More details can be found in the work of Livneh *et al.* [2010].

[29] Upgrades to the Mosaic model consisted only of a decrease in the soil heat capacity from 175,000 to 70,000 J m⁻² K⁻¹. This was done in response to the negative daytime and positive nighttime ground heat flux biases found in NLDAS-1 [Robock *et al.*, 2003], and in response to the cool bias in daytime land surface temperatures during the warm season also found in NLDAS-1 [Mitchell *et al.*, 2004].

[30] An analysis of results from NLDAS-1 determined that the SAC-SMA model underestimated mean annual streamflow for the 1145 small, and nine large, study basins. This underestimation stemmed, in part, from SAC-SMA's use of potential evapotranspiration calculated by the Noah model, which resulted in overly high mean evapotranspiration and correspondingly low mean annual streamflow. This shortcoming was addressed in NLDAS-2, as discussed above, by altering the source of potential evapotranspiration to a climatologically averaged monthly potential evapotranspiration

data set. This leads to improved simulations of mean annual streamflow [Xia *et al.*, 2012].

[31] The VIC model also exhibited streamflow simulation problems in NLDAS-1, overestimating streamflow over most of the small study basins (in particular in the southeastern CONUS) and over eight of the nine large study basins. The NLDAS-1 biases in the VIC model have been attributed to the use of incorrect parameter values. With this in mind, *Troy et al.* [2008] used observed mean annual streamflow from 1997 to 2001 at 1130 small basins over the CONUS to re-calibrate the model. These re-calibrated parameters were used in the NLDAS-2 VIC simulations detailed in this paper.

[32] Many other upgrades and new parameterizations have been added to VIC since NLDAS-1, including new parameterizations of the soil temperature profile, canopy temperature, blowing snow, spatially variable snow cover and frozen ground, dynamic wetland and lake models. However, these were not used in NLDAS-2 because of model structural changes and ongoing testing.

2.4. Model Spin-Up and NLDAS-2 Setup

[33] The spin-up of a land surface model can be described as the movement of its states toward equilibrated values. The equilibrium state of a robust land surface model should be physically realistic, and its behavior in the adjustment period should be physically meaningful. Without appropriate spin-up, and thus appropriate initial forecast conditions, model predictions can be unstable and may exhibit drift as the model states try to move toward equilibrium values for the prevailing climate. Analyses based on these model outputs may give misleading results. *Yang et al.* [1995] discussed the time to equilibrium (spin-up period) of 22 land surface models for grass and forest sites within the PILPS project. The results showed a wide range of spin-up timescales, from one year to over 20 years, depending on the model, state variable and vegetation type. *Cosgrove et al.* [2003b] examined the spin-up behavior of LSMs at continental scales using the four NLDAS-1 land surface models. They selected six sub-regions covering North America, which were chosen to encompass a wide variety of climate, soil and vegetation regimes. The results showed that, in general, all models reached a state of approximate equilibrium within the first one to three years for all six sub-regions. Furthermore, an 11 yearlong recursive experiment with the Mosaic model showed that appreciable drift in large-scale land surface states ceased within approximately one year, and fine-scale (grid-scale) equilibrium was reached within 5.5 years for total soil moisture and root zone soil moisture. VIC and Noah displayed longer spin-up times than Mosaic and in some cases took longer than 10 years to reach fine scale equilibrium [*Cosgrove et al.*, 2003b].

[34] To minimize the impact of model spin-up on the NLDAS-2 analyses, we used a two-step method to generate the initial states for each of the four land surface models. First, the last eight years of output from the 11-year NLDAS-1 real-time simulation from 1 October 1996 to 31 December 2007 were averaged for each model to produce climatological initial land states for 00Z October 1. This first step disregards the first three years of the real-time simulation that are subject to spin up effects in the NLDAS-1 simulation. The NLDAS-2 forcing differs from the NLDAS-1 forcing in

terms of derivation methods and underlying data sets and the NLDAS-1 real-time models did not include the upgrades in section 2.3, and so in the second step a spin-up simulation was carried out using the NLDAS-2 forcing as follows: The climatological 00Z October 1 states from NLDAS-1 were used to initialize a 15.25 year spin-up simulation for each model from 1 October 1979 to 1 January 1995 using the NLDAS-2 forcing data. The January 1st states from the last 10 years of this spin-up simulation were then averaged together to provide initial states for the final NLDAS-2 simulation. In this final analysis run, the four models were then executed from 1 January 1979 to the present using the NARR- and observation-based NLDAS-2 forcing. Because the routing code that translates gridded model runoff to streamflow needs 6 months of spin-up, only 28 years of model output from 1 October 1979 to 1 September 2007 is analyzed in this study.

3. Model-to-Model Comparison of Water and Energy Budgets

[35] Results from the NLDAS-2 series of simulations are analyzed in the context of the surface water balances of SAC-SMA, VIC, Noah, and Mosaic, and through the energy balances of the three energy balance models: VIC, Noah, and Mosaic. For a given time period, the surface water balance equation can be written as

$$\frac{dS_1}{dt} + \frac{dS_c}{dt} + \frac{dS_n}{dt} = P - E - R_1 - R_2 \quad (4)$$

where P , E , R_1 , and R_2 are mean precipitation, evapotranspiration, surface runoff and subsurface runoff, respectively. $\frac{dS_1}{dt}$, $\frac{dS_c}{dt}$, and $\frac{dS_n}{dt}$ are the change in storage of total column soil moisture, canopy interception, and snowpack.

[36] Similarly, the surface energy balance equation for the three NLDAS-2 models (Noah, VIC, and Mosaic) that simulate the energy balance is

$$R_n - LE - SH - G - SF = 0 \quad (5)$$

where $R_n = (S^\downarrow - S^\uparrow + IR^\downarrow - IR^\uparrow)$ is the surface net radiation, and S^\downarrow , S^\uparrow , IR^\downarrow , and IR^\uparrow are the surface downward shortwave radiation, surface upward shortwave radiation, surface downward longwave radiation, and surface upward longwave radiation fields, respectively. Additionally, LH is the latent heat flux, SH is the sensible heat flux, G is the ground heat flux, and SF is the snow phase change heat flux.

[37] Upward shortwave (S^\uparrow) and longwave (IR^\uparrow) radiation are equivalent to

$$S^\uparrow = \alpha_{ave} S^\downarrow \quad (6a)$$

$$IR^\uparrow = \varepsilon \sigma T_{skin}^4 + (1 - \varepsilon) IR^\downarrow \quad (6b)$$

where IR^\downarrow is downward longwave radiation, α_{ave} is the area-averaged albedo for a given grid box, ε is land surface emissivity, and σ is the Stefan–Boltzmann constant. It should be noted that the representation of upward shortwave radiation given by equation (6a) is appropriate for the Noah and Mosaic models. The VIC model includes absorption of the shortwave radiation by the vegetation canopy, and so

Table 2. A Summary of the 28-Year Average Annual Energy Budget Analysis

Model	NETLONG (W/m ²)	NETSHORT (W/m ²)	RNET (W/m ²)	LH (W/m ²)	SH (W/m ²)	G (W/m ²)	SF (W/m ²)	Res (W/m ²)
<i>Continental United States (CONUS)</i>								
Noah	-68.1	152.2	84.1	37.4	46.5	0.4	-0.6	0.0
Mosaic	-66.0	154.5	88.5	50.3	37.6	0.0	-0.9	0.0
VIC	-74.5	150.9	76.4	42.1	34.9	1.2	-0.6	0.0
<i>NW CONUS</i>								
Noah	-72.1	140.5	68.4	25.3	42.2	1.0	-1.9	0.0
Mosaic	-69.7	141.3	71.5	35.8	34.8	0.0	-0.9	0.0
VIC	-78.3	137.6	59.3	29.1	30.2	1.12	-1.12	0.0
<i>NE CONUS</i>								
Noah	-51.8	127.8	76.0	36.6	39.2	0.1	-0.3	0.0
Mosaic	-51.3	136.8	85.5	58.7	25.8	0.0	-1.0	0.0
VIC	-53.7	125.1	71.3	42.4	30.0	2.1	-1.0	0.0
<i>SW CONUS</i>								
Noah	-89.0	172.5	83.6	26.5	57.0	0.9	-1.0	0.0
Mosaic	-85.8	168.4	82.6	31.6	50.7	0.0	-0.3	0.0
VIC	-98.7	172.8	74.1	30.8	43.6	0.9	-0.6	0.0
<i>SE CONUS</i>								
Noah	-54.8	159.1	104.4	59.5	45.2	0.3	0.0	0.0
Mosaic	-52.9	164.8	111.9	76.4	35.5	0.0	0.0	0.0
VIC	-61.0	158.7	97.8	64.7	33.9	0.8	0.0	0.0

upward shortwave radiation in this model depends also on multiple reflection and absorption processes between the canopy and the ground. For the three models that simulate the surface energy balance (Noah, Mosaic and VIC), the infrared reflectance term (second term in equation (6b)) is set to zero. In the absence of snow, all models use an emissivity of 1 and the infrared reflectance term is zero. In the presence of snow, the emissivity is weighted by a fixed snow emissivity multiplied by the snow covered area fraction, but the impact of this on reflectance is ignored.

3.1. Mean Annual Comparison

[38] We begin the model-to-model comparison by presenting the water and energy budgets of the models at an annual time scale. Tables 2 and 3 give the mean annual energy budgets for three models (Noah, Mosaic, and VIC) and mean annual water budgets for four models (Noah, Mosaic, SAC-SMA, and VIC) over four quadrants of the CONUS from 1 October 1979 to 30 September 2007. The four analysis quadrants are the Northwestern CONUS (NW, 98–125°W, 40–50°N), Northeastern CONUS (NE, 67–98°W, 40–50°N), Southwestern CONUS (SW, 98–125°W, 25–40°N), and Southeastern CONUS (SE, 67–98°W, 25–40°N), as defined by Lohmann *et al.* [2004]. The area-averaged mean annual energy budget analysis shows that the Noah, Mosaic and VIC models conserve energy well, with each featuring zero energy residual. Mosaic produces the most net radiation over almost all regions because it generally has the largest net shortwave radiation and smallest net longwave radiation. VIC has the smallest net radiation over all of the analysis areas as it features the smallest net longwave radiation. The Noah LSM's net radiation falls in-between that from VIC and Mosaic for the CONUS and over three of the four quadrants. The exception is over the SW where it produces the largest net radiation because of its large net shortwave radiation and small net longwave

radiation value. Downward shortwave and longwave radiation are provided to each model in the NLDAS-2 forcing data set and are the same for all three models. Therefore the differences in net radiation are due to differences in shortwave radiation surface albedo (plus canopy effects in VIC),

Table 3. A Summary of the 28-Year Average Water Budget Analysis

Model	P (mm/yr)	ET (mm/yr)	R (mm/yr)	P – ET – R (mm/yr)	R/P
<i>Conterminous United States (CONUS)</i>					
Noah	773.6	463.0	311.0	-0.4	0.40
Mosaic	773.6	635.1	139.1	-0.6	0.18
SAC-SMA	773.6	534.4	239.1	0.1	0.31
VIC	773.6	537.3	267.0	-30.7	0.35
<i>NW CONUS</i>					
Noah	587.0	312.9	273.9	0.2	0.46
Mosaic	587.0	451.7	135.2	0.1	0.23
SAC-SMA	587.0	384.9	201.2	0.9	0.34
VIC	587.0	373.6	238.5	-25.1	0.41
<i>NE CONUS</i>					
Noah	896.8	451.9	445.3	-0.4	0.49
Mosaic	896.8	740.1	157.0	-0.3	0.18
SAC-SMA	896.8	637.3	258.8	0.7	0.29
VIC	896.8	541.0	380.0	-24.2	0.42
<i>SW CONUS</i>					
Noah	436.2	328.7	107.4	0.1	0.25
Mosaic	436.2	400.0	36.0	0.2	0.08
SAC-SMA	436.2	362.5	72.2	-2.5	0.17
VIC	436.2	393.2	78.1	-35.1	0.18
<i>SE CONUS</i>					
Noah	1192.7	742.1	451.2	-0.6	0.38
Mosaic	1192.7	966.0	230.7	-4.0	0.19
SAC-SMA	1192.7	772.8	421.3	-1.4	0.35
VIC	1192.7	825.2	402.6	-35.1	0.34

and to differences in the surface temperature and emissivity for longwave radiation, as indicated by equations (6a) and (6b), respectively.

[39] Figure 1 shows the mean annual surface albedo (Figures 1a–1c) and skin temperature (Figures 1d–1f) from Noah, Mosaic and VIC calculated over the 28-year NLDAS-2 study period. Mosaic's albedo is smaller than Noah's for all regions except for the SW, in particular over the SE and the west coast. Low albedo values result in lower upward shortwave radiation, which is consistent with the higher net shortwave radiation in Mosaic. VIC has a larger surface albedo and smaller net short shortwave radiation than Noah for western mountainous regions. In most regions of the southeastern and southwestern United States, VIC has a smaller surface albedo than Noah. However, VIC features higher skin temperatures over the CONUS than do Noah and Mosaic (Figures 1d–1f), and this is reflected in the larger upward longwave radiation and smaller net longwave radiation values listed in Table 2. The values of surface albedo cannot explain the CONUS warm bias in the VIC land surface skin temperature in areas outside of the western mountainous regions. Further analysis showed that the bias is mainly due to the large ground heat flux during the nighttime, when the soil releases too much heat to the land surface. The reason for this is unclear at present and will be addressed in future studies.

[40] Complementing the preceding energy budget analysis, the area-averaged mean annual water budget analysis (Table 3) shows that Noah, Mosaic, and SAC-SMA generally conserve water over all regions. However, VIC displays a small water imbalance (<4%) over the CONUS and all quadrants except for the SW where it is about 8%. For a 28-year average, stored water should equal lost water so that the model water fluxes are balanced. However, for the VIC model, lost water is much larger than stored water. The VIC model water budget imbalance was traced back to the model parameterization for sub-grid precipitation, which is used to simulate the fractional grid coverage of a storm as a function of storm total [Gong *et al.*, 1994]. This parameterization is an optional process in the model and was run in NLDAS-1 and remained switched on in NLDAS-2. We re-ran three selected grid cells which had large water balance errors in the original simulation, with this parameterization switched on (the same as the original simulation) and off. The water balance errors with the sub-grid precipitation parameterization switched on produced unacceptably large errors, but were zero or close to zero when the parameterization was switched off. Examination of time series of the water budget variables showed spurious jumps in soil moisture for the sub-grid precipitation simulation which is likely due to model code errors when the water storage terms in the wet and dry sub-grid fractions are merged after a storm has passed. The full simulation will be re-run without this parameterization. Overall, Noah (Mosaic) features the smallest (largest) evaporation and largest (smallest) total runoff for all regions. The evaporation and total runoff generated by SAC-SMA and VIC lie between the other two models, with VIC producing more total runoff than SAC-SMA except over the SE. Noah and VIC have larger mean annual runoff ratios than do Mosaic and SAC-SMA over the CONUS and NW, NE and SW regions. Mosaic is characterized by the lowest mean annual runoff ratio of the four models.

[41] Figure 2 shows the spatial distribution of the 28-year mean annual runoff ratio along with the model spread calculated from the four-model ensemble mean (EM). Model spread is defined as the magnitude of the standard deviation between the individual models in the ensemble mean. Large runoff ratios over the east and the west coasts, and smaller values over the interior states, are driven by gradients of mean annual precipitation. Echoing the results depicted in Table 3, Noah produces the largest runoff ratio and Mosaic the smallest. The values from SAC-SMA and VIC are close to those of the ensemble mean, although VIC features larger runoff ratios than does SAC-SMA. The largest differences among the four models are located in the NE and in the western mountains, where Noah and VIC produce large runoff ratios, and Mosaic and SAC-SMA produce small runoff ratios.

[42] Mean annual evaporation from the four models, the ensemble mean, and the model spread are shown in Figure 3. Mosaic generates the largest evaporation over the CONUS (Figure 3c) and Noah generates the smallest evaporation (Figure 3b). As was the case with the runoff ratio, SAC-SMA and VIC lie in-between (Figures 3d and 3e) and are close to the ensemble mean (Figure 3a). The disparity in mean annual evaporation among the four LSMs is most obvious in the NE and over the western mountainous regions, a finding which is in line with the runoff ratio results. Overall, the disparity in evaporation among the four models is much less than that found in the NLDAS-1 simulations shown by Mitchell *et al.* [2004, Figure 2]. This tightening of results is due to the model upgrades, parameter calibrations, and forcing alterations outlined in section 2.3. In particular, the re-calibration of VIC has removed the over-estimation of ET in densely vegetated regions such as the SE, and the use of climatologically averaged versus Noah-supplied potential evaporation (PE) as a forcing for the SAC-SMA model has removed the underestimation of ET over the same region.

3.2. Annual Cycle Comparison

[43] In general, all of the components of the energy balance show a clear seasonal variation over the CONUS (for the three models that simulate the energy cycle and their ensemble mean, Figure 4). The seasonal cycle of surface net radiation (solid line) varies little between models and is mainly driven by sun-earth geometry. The seasonal cycles of latent heat flux (dotted line) and sensible heat flux (dashed line) loosely follow the seasonal course of surface net radiation, but do vary between models. Noah exhibits a larger sensible heat flux than latent heat flux, particularly in the spring and early summer (Figure 4a), while Mosaic displays the opposite (Figure 4b). The VIC model produces a larger latent heat flux than sensible heat flux in the winter and summer, a smaller latent heat flux in the spring (Figure 4c), and is generally close to the ensemble mean (Figure 4d), except for small differences in December and January and an overestimation of latent heat flux in the summer. Noah, Mosaic, and VIC each display a negative ground heat flux (dashed-dotted line), representing a release of energy from the ground to the atmosphere, during the winter, and a positive ground heat flux, representing a transfer of heat energy from the atmosphere to the ground, in the summer. The snow phase change heat flux is quite small over the CONUS (not shown). The energy residual averaged over the CONUS

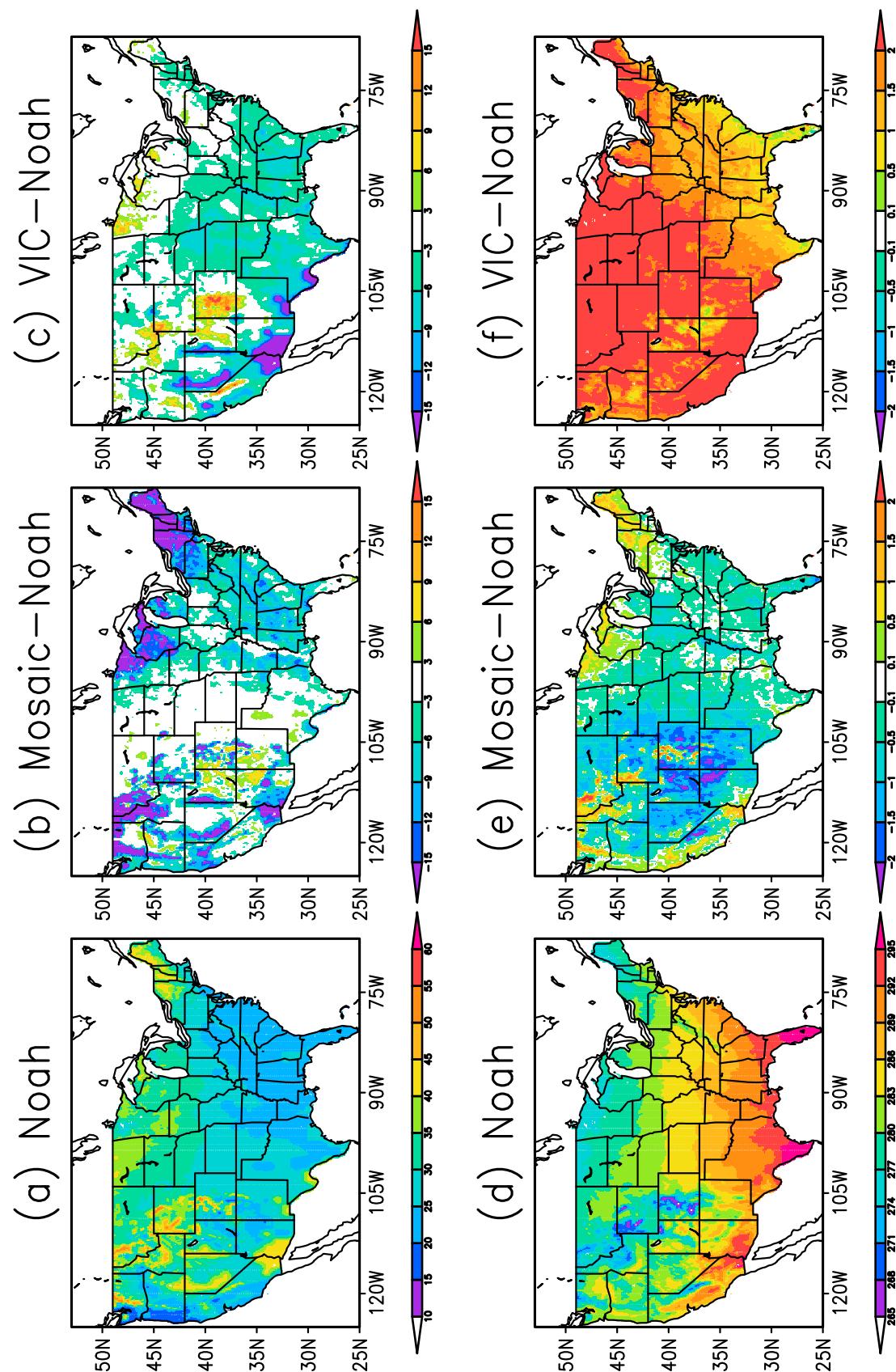


Figure 1. The 28-year mean annual surface albedo (%) for (a) Noah model, (b) difference between Mosaic and Noah models, and (c) difference between VIC and Noah models; and 28-year mean annual surface skin temperature (K) for (d) Noah model, (e) difference between Mosaic and Noah models, and (f) difference between VIC and Noah models.

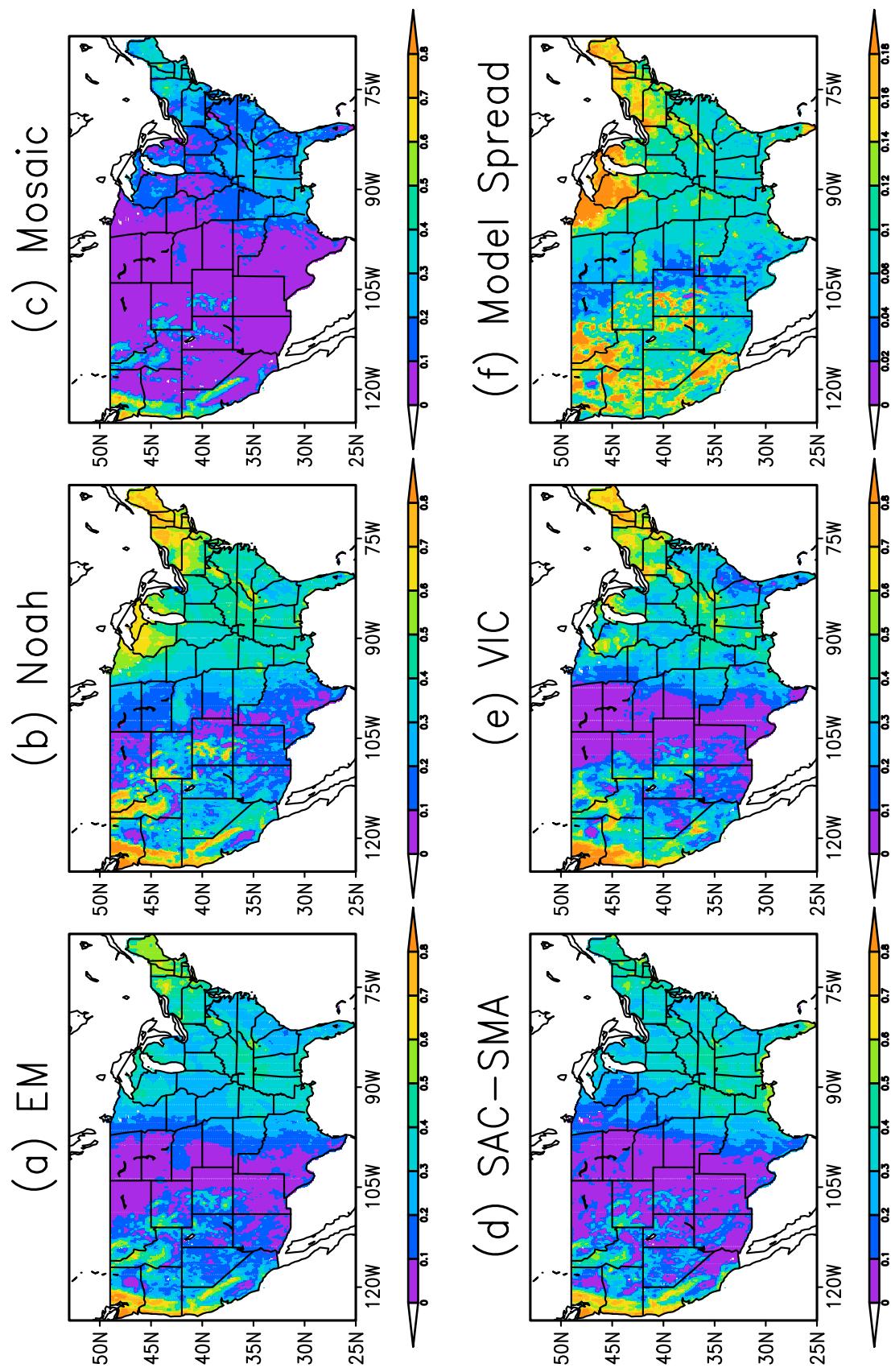


Figure 2. The 28-year mean annual runoff ratio for (a) ensemble mean EM, (b) Noah model, (c) Mosaic model, and (d) SAC-SMA model, (e) VIC model and (f) their spread.

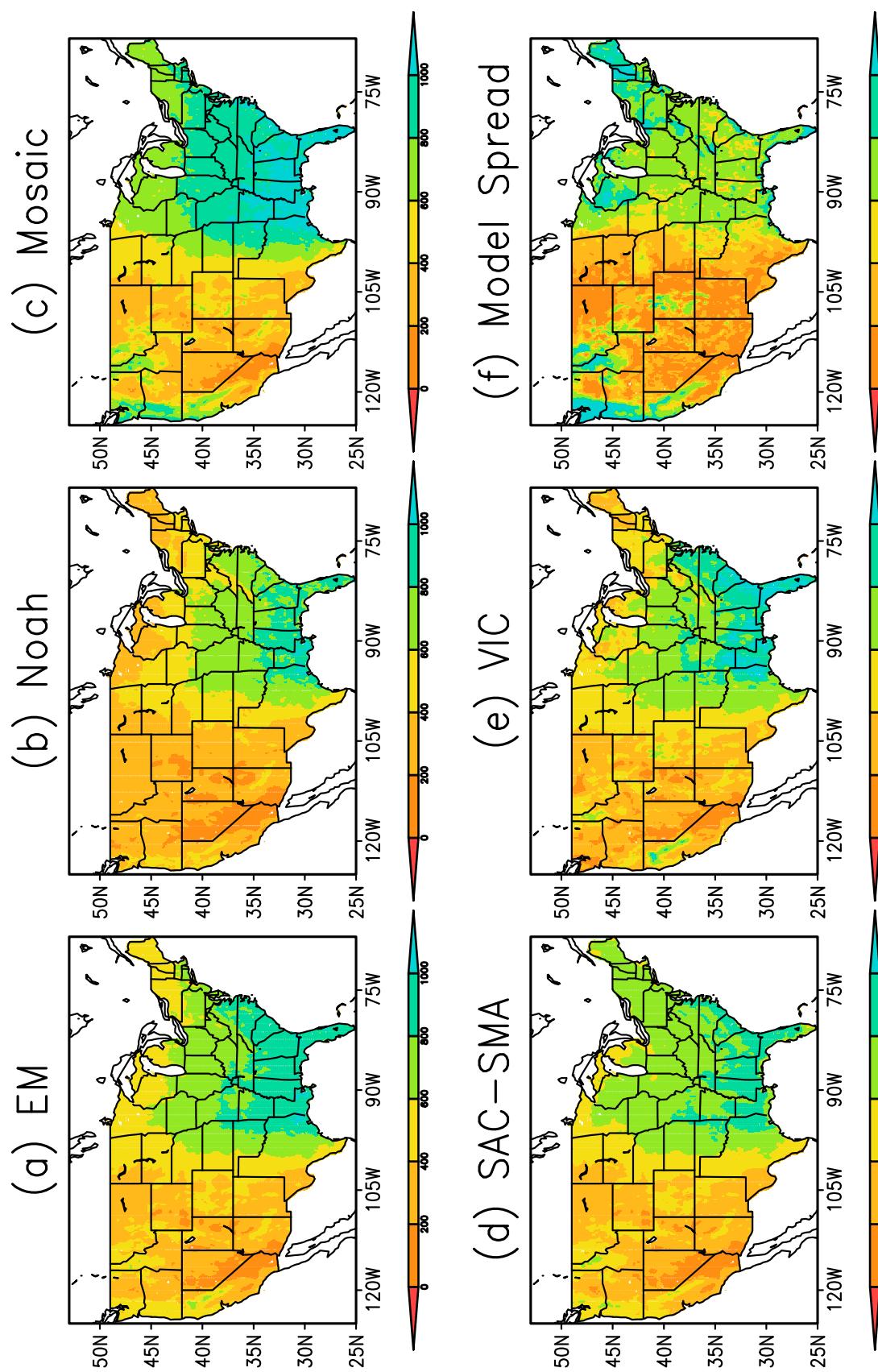


Figure 3. The 28-year mean annual evaporation (mm/year) for (a) ensemble mean, (b) Noah model, (c) Mosaic model, and (d) SAC-SMA model, (e) VIC model and (f) their spread.

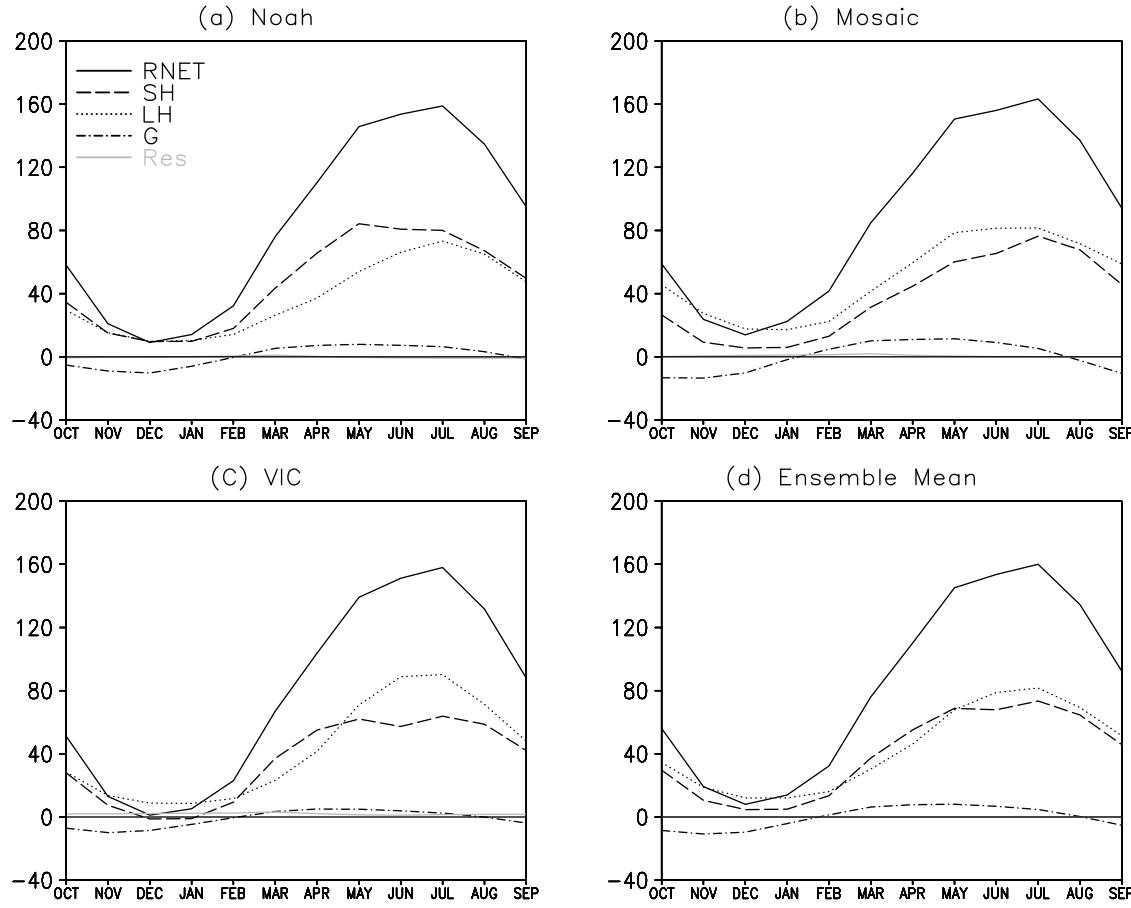


Figure 4. Seasonal cycle of 28-year area-averaged mean annual energy balance components (W/m^2) over the CONUS for (a) Noah model, (b) Mosaic model, (c) VIC model, and (d) three-model ensemble mean (thick solid line: net radiation RNET, dotted line: latent heat flux LH, dashed line: sensible heat flux SH, dashed-dotted line: ground heat flux G, gray solid line: energy balance residual Res).

(dashed-dotted-dotted line) is close to zero, implying that all three models close the energy balance very well. The differences in the seasonal cycle of the ratio of sensible heat / latent heat fluxes are fairly large among models. This was also shown in the NLDAS-1 model evaluations against measurements in the southern Great Plains [Robock *et al.*, 2003]. They found the seasonal cycles of latent heat to be slightly low for VIC with a commensurate slight high bias in sensible heat, but a very high bias in latent heat for Mosaic that was mainly due to an exaggerated ground heat flux. The latter was due to a high soil heat capacity and this was corrected in NLDAS-2 (section 2.3). The Noah model gave reasonable results compared to the observations. Further investigation into the differences in the NLDAS-2 models shown in Figure 4 using observed sensible and latent heat fluxes from all available flux towers in the United States is underway and preliminary results suggest that no single model does better than the others, and that the three-model ensemble mean is more realistic than any individual model compared with observations.

[44] The annual cycles of NLDAS-2 LSM land surface water fluxes and storages, averaged over the CONUS for 1979–2007, are shown in Figure 5. Total precipitation, which is identical for all models, is largest in the summer, reflecting the dominance of the North American monsoon

(solid line). All models exhibit a summer maximum of evapotranspiration that exceeds precipitation, although the month varies among the models between June and July (dotted line). The Noah model has the smallest evapotranspiration and the largest total runoff of all models. By contrast, the Mosaic model has the largest evapotranspiration and smallest runoff. The runoff peaks in March for all models (dashed line) in response to spring snowmelt and the interplay of precipitation minus evapotranspiration. The accumulation of snowpack creates a residual in the water balance in the winter, which is depleted in the summer mainly by evapotranspiration, as reflected by changes in the soil moisture (dashed-dotted). The model-to-model differences in water storage change are most obvious in the summertime, with the VIC model featuring the largest changes (Figure 5d).

[45] Focusing on soil moisture, Figure 6 displays the annual cycle of area-averaged monthly mean top-2 m soil moisture over the four quadrants of the U.S. These annual cycles are characterized by peaks in the winter and spring, and lower values in the summer. In the NW, the Noah (solid line) and SAC-SMA models (dotted line) have larger top 2 m soil moisture values than do Mosaic (dashed) and VIC (dashed-dotted line). In the NE quadrant, Noah, Mosaic and SAC-SMA have comparable levels of soil moisture, all of

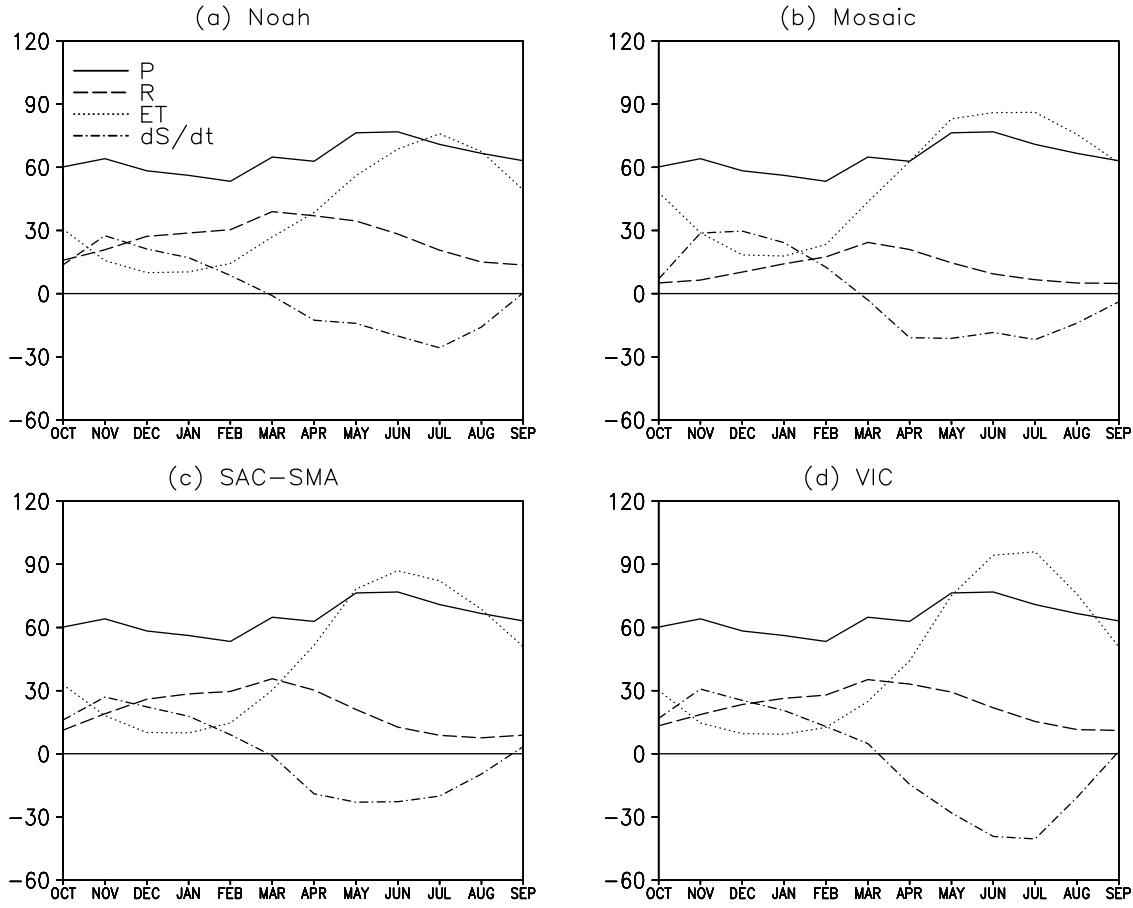


Figure 5. Seasonal cycle of 28-year area-averaged mean annual water balance components (mm/month) over the CONUS for (a) Noah model, (b) Mosaic model, (c) SAC-SMA model, and (d) VIC model (Thick solid line: precipitation P, dotted line: evaporation ET, dashed line: total runoff R, dashed-dotted line: water storage change dS/dt).

which are larger than what VIC features. Mosaic and VIC soil moisture values are comparable, and are smaller than those produced by Noah and SAC-SMA. And finally, in the SE, Mosaic and SAC-SMA soil moisture levels are comparable, and are smaller than those simulated by Noah and VIC. Overall, larger differences among the models appear over the wetter northeastern and northwestern quadrants, and smaller differences among the models appear in the southwestern quadrants. Generally speaking, there are larger differences among the models in the North than in the South.

[46] Figure 7 shows the 28-year mean May through September change in top 2 m soil moisture. This period was chosen to represent the approximate soil moisture dry-down season. Positive values denote soil drying and negative values denote soil moistening. Soil moistening in Florida, western Texas and eastern New Mexico is out of phase with the dry-down seen over most of the CONUS, since the wet season in these regions occurs during the summertime [Mitchell *et al.*, 2004]. While this is generally consistent with the results from NLDAS-1, the spatial extent of the areas of moistening is much smaller than seen in NLDAS-1 [see Mitchell *et al.*, 2004, Figure 8]. This change is due partly to the use of the longer 28-year mean in this study (Mitchell *et al.* [2004] looked only at 1999), and partly to the lingering effects of model spin-up in NLDAS-1. There is

also less disparity between the NLDAS-2 models than that seen in NLDAS-1, owing to the improvements in model code, parameter tuning, and changes to forcing data outlined above.

[47] Focusing on a cold-season variable, Figure 8 shows the annual cycle of area-averaged snow water equivalent (SWE) for all four quadrants and models. The area-average includes all grid cells with both snow and snow-free cells. As noted in section 2.2, while VIC, Mosaic, and Noah include snow sub-models within their code, snowpack is not internally simulated by SAC-SMA, but rather is simulated in parallel by the Snow17 model. The SWE values produced by each model show a similar seasonal cycle over all four quadrants, although the peak time and magnitude of SWE differ substantially between models. The SWE generated by the Snow17 model (dotted line) is largest for all four quadrants partly because Snow17 assumes that snow sublimation is zero. SWE in Noah (solid line) is comparable to that in VIC (dashed dotted) and Mosaic (dashed) for all four quadrants except for the SW, where Noah features lower SWE values and produces earlier snowmelt. This behavior is a departure from NLDAS-1 in which Noah had the smallest SWE and displayed early snowmelt in all four quadrants [Mitchell *et al.*, 2004], and stems from improvements in Noah's cold-season parameterizations [Livneh *et al.*, 2010].

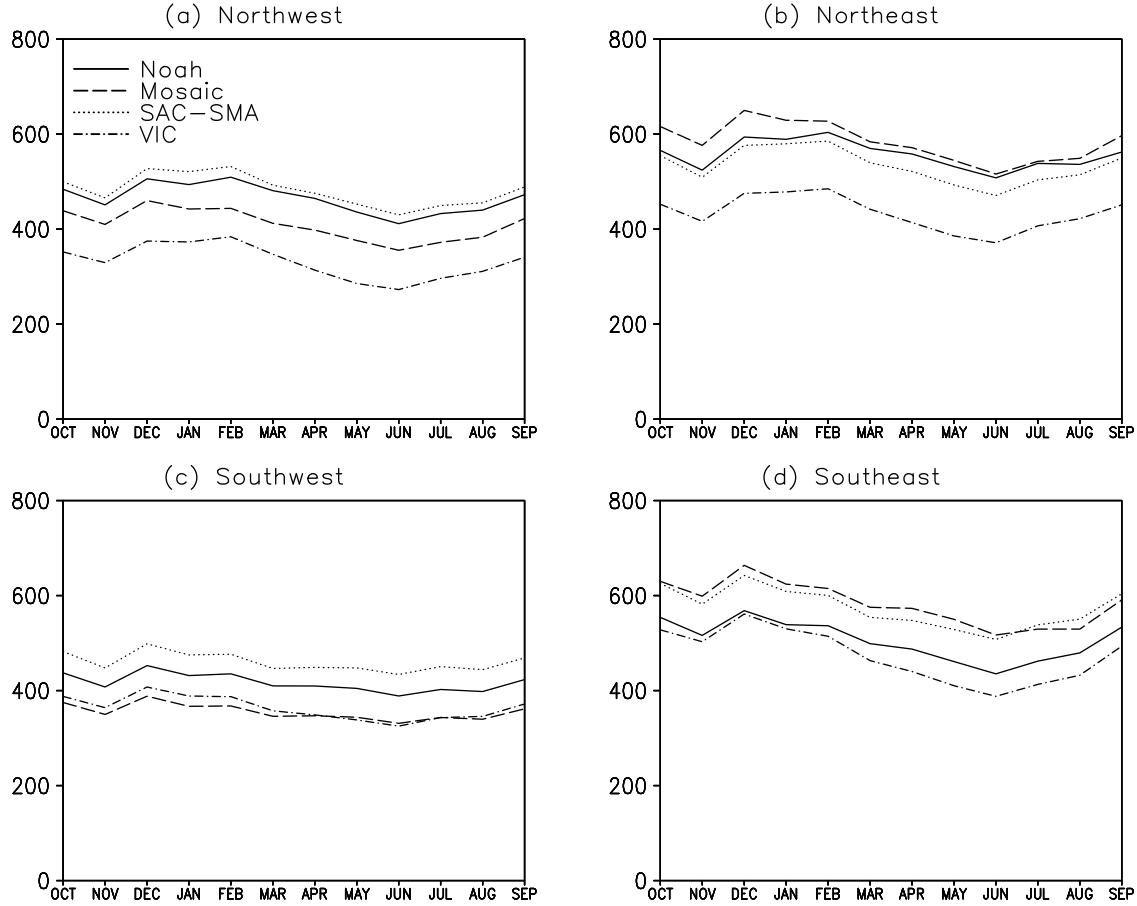


Figure 6. Seasonal cycle of 28-year mean annual top 2 m soil moisture (mm) area-averaged over four CONUS quadrants: (a) NW, (b) NE, (c) SW, and (d) SE (solid line: Noah model, dashed line: Mosaic model, dotted line: SAC-SMA model, dashed-dotted line: VIC model).

The remaining low bias in the SW region is relatively small, but requires further investigation.

4. Analysis of Four-Model Simulation Similarity

[48] Each of the four participating land surface models drew their own required set of forcing variables from a single source of meteorological forcing data produced as part of NLDAS-2. Even with this source of output differences eliminated, differences in the parameterizations of each model can give rise to a very large degree of variation in output. These differences depend not only on the models themselves, but also on the particular variables studied. As illustrated by the multimodel analysis carried out under the Global Soil Wetness Project-2 (GSWP-2) [Dirmeyer *et al.*, 2006], the LSM variables that are highly constrained by forcing data, such as net shortwave radiation and skin temperature, feature large similarity and small spread across models. By contrast the variables associated with snow processes (i.e., snow water equivalent) and soil water (i.e., soil moisture in the lower layers) display small similarity and large spread. To quantify the similarity of the four models in this study, we used two criteria based on monthly anomalies. The monthly timescale was chosen over other durations due to the success of this approach in GSWP-2 [Dirmeyer *et al.*, 2006] and the multimodel study of Wang *et al.* [2009]. The

use of monthly data also proved effective in NLDAS-1, where it was used to quantify and explain the main differences in model output while filtering out the noise contribution of day-to-day weather patterns.

[49] The first monthly anomaly based criterion is similar to that defined in GSWP-2 [Dirmeyer *et al.*, 2006]: For a given model and variable, X , the monthly anomaly is defined as

$$X_{n,m,y}^{Anom} = X_{n,m,y} - \bar{X}_{n,m}, \quad \bar{X}_{n,m} = \frac{1}{Y} \sum_{y=1}^Y X_{n,m,y} \quad (7)$$

where $\bar{X}_{n,m}$ is the mean monthly value for model n and month m , and $X_{n,m,y}^{Anom}$ is the anomaly for year y , and $Y = 28$ is the total number of years. We can define the multimodel mean as

$$\bar{X}_{m,y}^{Anom} = \frac{1}{N} \sum_{n=1}^N X_{n,m,y}^{Anom} \quad (8)$$

where $N = 4$ is the total number of models and the intermodel standard deviation is simply

$$\sigma_{m,y} = \sqrt{\frac{1}{N} \sum_{n=1}^N (X_{n,m,y}^{Anom} - \bar{X}_{m,y}^{Anom})^2} \quad (9)$$

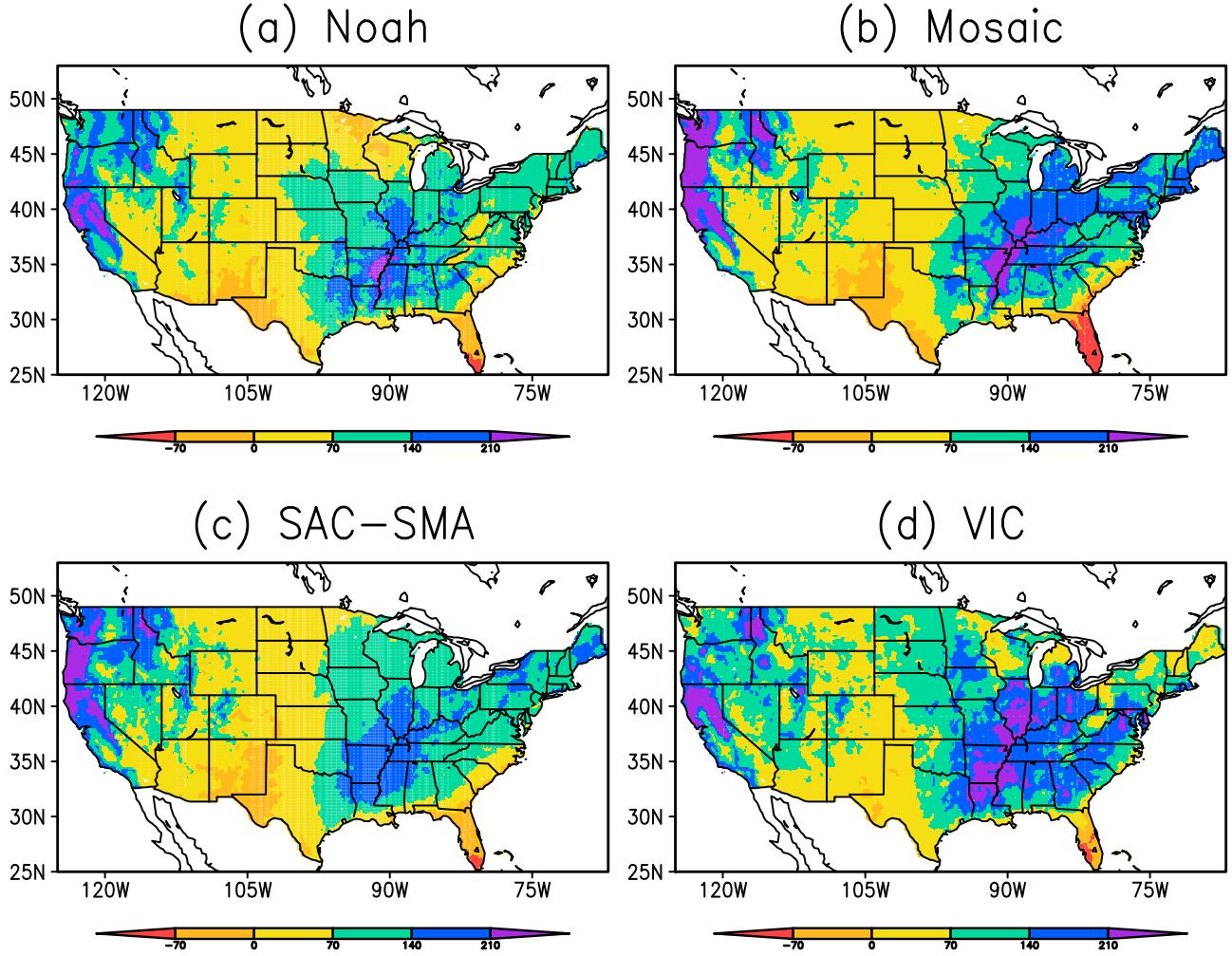


Figure 7. Warm season storage change (mm) of top 2 m soil moisture, calculated as 28 year mean of top 2 m soil moisture in April minus that in September for (a) Noah model, (b) Mosaic model, (c) SAC-SMA model, and (d) VIC model (positive denotes net drying and negative denotes net moistening).

The time average of $\sigma_{m,y}$ is

$$\bar{\sigma} = \frac{1}{MY} \sum_{m=1,y=1}^{m=M,y=Y} \sigma_{m,y} \quad (10)$$

where $M=12$ is the number of months in a year, and the total temporal standard deviation of the multimodel analysis is defined as

$$\sigma_{Total} = \sqrt{\frac{1}{MY} \sum_{m=1,y=1}^{m=M,y=Y} (\bar{X}_{m,y}^{Anom} - \bar{X})^2} \quad (11)$$

where $\bar{X} = \frac{1}{MY} \sum_{m=1,y=1}^{m=M,y=Y} \bar{X}_{m,y}^{Anom}$, which is the mean value of

the monthly anomaly time series for a given variable. To avoid a wide range of values for inter-model variance compared to the other variance metrics, we define an anomaly ratio as the ratio of inter-model standard deviation to total standard deviation as used in GSWP-2:

$$R_{ratio}^{anomaly} = \frac{\bar{\sigma}}{\sigma_{Total}} \quad (12)$$

This quantity measures the inter-model simulation similarity. Besides using GSWP-2's anomaly ratio approach, we also use their definition for the original monthly mean time series, $R_{ratio}^{original}$ (it is represented as R by Dirmeyer *et al.* [2006]). The two ratios quantify the magnitude of the difference between the models, and are calculated for each individual grid cell within the NLDAS-2 domain.

[50] The second metric we use to quantify model simulation similarity is the mean monthly anomaly correlation coefficient between all six combinations of the four LSMs, as used by Wang *et al.* [2009]. This value quantifies how similarly the four models vary in time for a given variable.

[51] The ratios of the inter-model standard deviation to the total standard deviation for monthly anomaly ($R_{ratio}^{anomaly}$) and monthly mean ($R_{ratio}^{original}$) time series are listed in Table 4 for six variables. Following Dirmeyer *et al.* [2006], these variables are rated from “A” to “D” based on the level of model agreement, from high to low, respectively. Examining the monthly anomaly based results, it can be seen that snowmelt is rated an “A,” and that all other variables are rated a “B,” indicating a high level of agreement among the models. Focusing on the results of analyzing the monthly mean data,

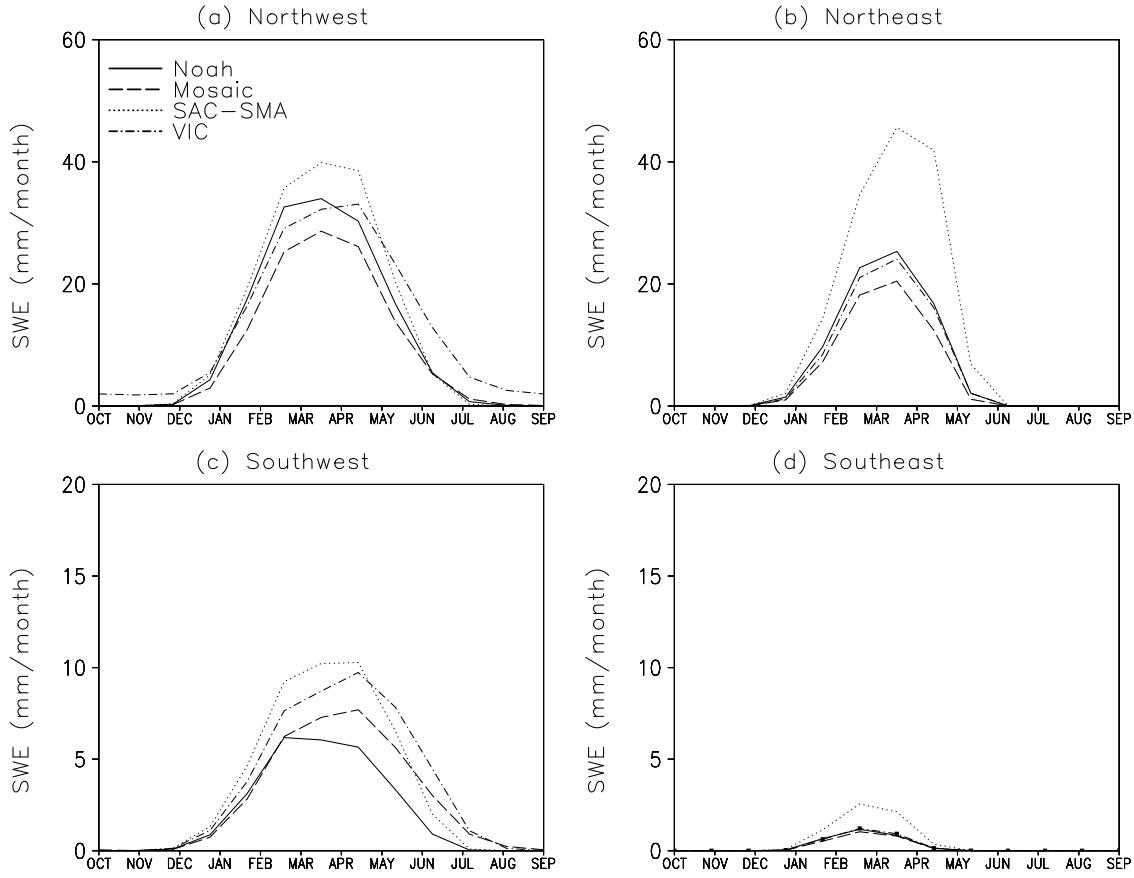


Figure 8. Seasonal cycle of 28-year mean annual snow water equivalent (mm) for four CONUS quadrants: (a) NW, (b) NE, (c) SW, and (d) SE (solid line: Noah model, dashed line: Mosaic model, dotted line: SAC-SMA model, dashed-dotted line: VIC model).

the top 2 m soil moisture and surface runoff variables drop to “C” ratings, and subsurface runoff drops to a “D” rating, indicating a large level of disagreement among the models. These results differ from Dirmeyer *et al.* [2006], who did not find as high a level of agreement between the output variables associated with soil physics and snow processes. The reasons behind this increased level of agreement include the different selection of LSMs used in NLDAS-2, as well as the use of a CONUS versus global study domain. Overall, the results indicate good general agreement among the models.

[52] Figures 9 and 10 show the spatial distributions of $R_{ratio}^{Anomaly}$ and the anomaly correlation, respectively, for the same six variables as in Table 4. The ratio represents the inter-model variability/disparity, and the anomaly correlation represents the inter-model similarity for a given variable. Large values of the ratio and small values of the anomaly correlation indicate large inter-model disparity or vice versa. The criteria for different classifications of disparity are defined following Dirmeyer *et al.* [2006] for the ratio and on the significance level for the anomaly correlations. For evaporation (Figures 9a and 10a), the largest spatial disparities appear over coastal and mountainous areas, in particular in the NE and Lake Superior regions, while the smallest disparities appear over the interior states and the desert SW. For surface runoff (Figures 9b and 10b), large disparities are observed in several states scattered

across the interior region and in the East. Correlations are higher in the SE and along the western coast, and are lower in the NE and western mountains (Figure 10b). The subsurface runoff anomaly (Figure 9c) shows the largest disparity among the six variables, and except for the SE and western coast, most regions exhibit large disparities and small anomaly correlations (Figure 10c). This is particularly true over very dry interior regions such as northern Texas, northeastern Montana, and northwestern North Dakota (Figure 9c). The top 2 m soil moisture anomaly displays the largest disparity in the NE and Lake Superior region, with some large disparities also present across the western mountains (Figure 9d). These findings are echoed by the anomaly correlations in Figure 10d. Differences in SWE and

Table 4. Summary of $R_{ratio}^{Anomaly}$ and $R_{ratio}^{original}$ Averaged Over the Conterminous United States for Six Variables^a

Variable	$R_{ratio}^{Anomaly}$	$R_{ratio}^{original}$
Evaporation	0.58 (B)	0.48 (B)
Surface Runoff	0.50 (B)	1.18 (C)
Subsurface Runoff	0.62 (B)	4.33 (D)
Top 2 m Soil Moisture	0.54 (B)	1.83 (C)
Snow Water Equivalent	0.44 (B)	0.70 (B)
Snowmelt	0.24 (A)	0.10 (A)

^aA, B, C, and D ranks are defined following Dirmeyer *et al.* [2006].

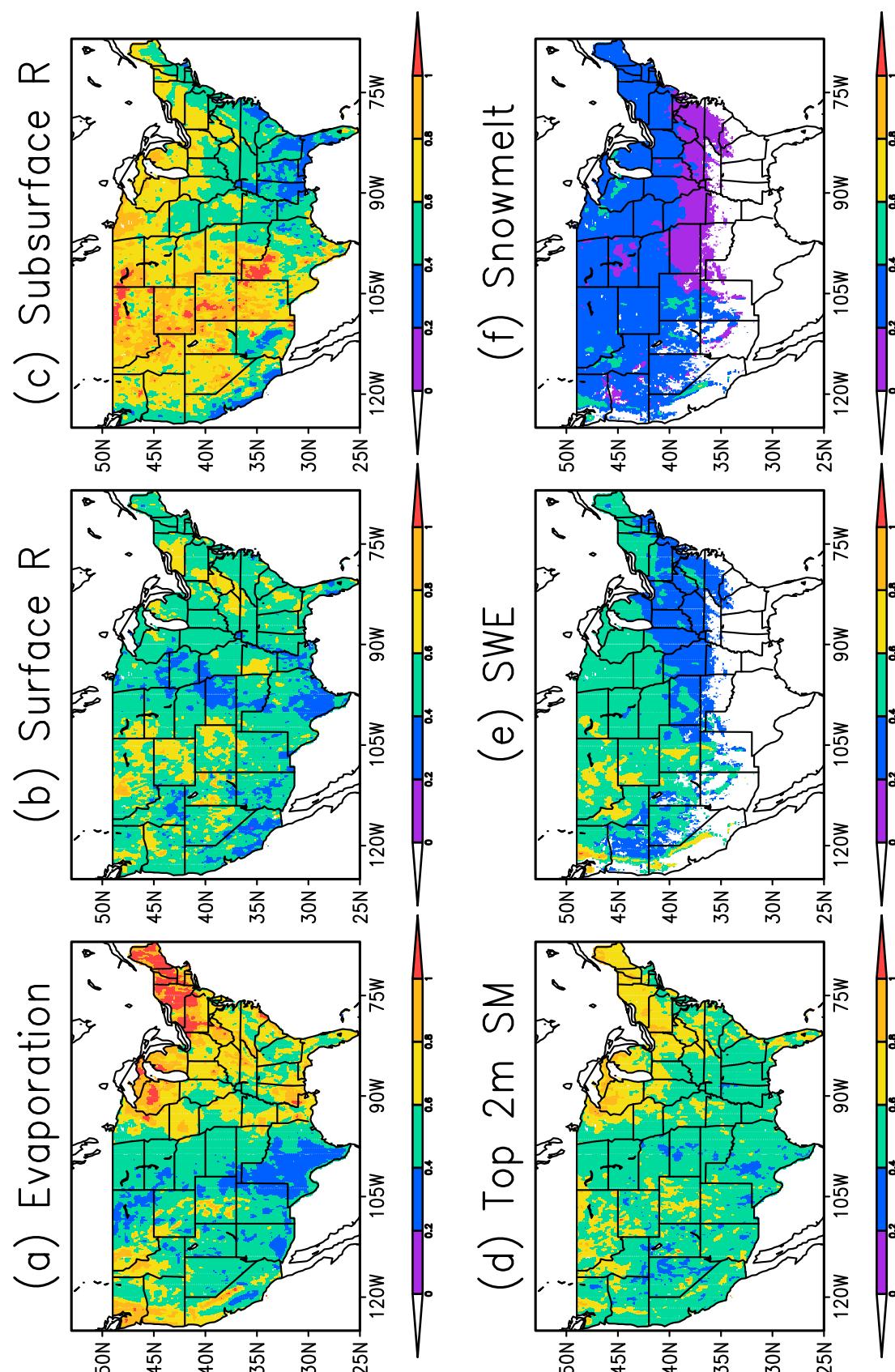


Figure 9. The spatial distribution of the ratio ($R_{ratio}^{anomaly}$) of the inter-model standard deviation to the total standard deviation for monthly mean anomalies of (a) evaporation, (b) runoff, (c) subsurface R, (d) top 2 m SM, (e) SWE, and (f) snowmelt (R denotes runoff, SM denotes soil moisture, SWE denotes snow water equivalent, small values represent high level of agreement, large values represent significant disparity).

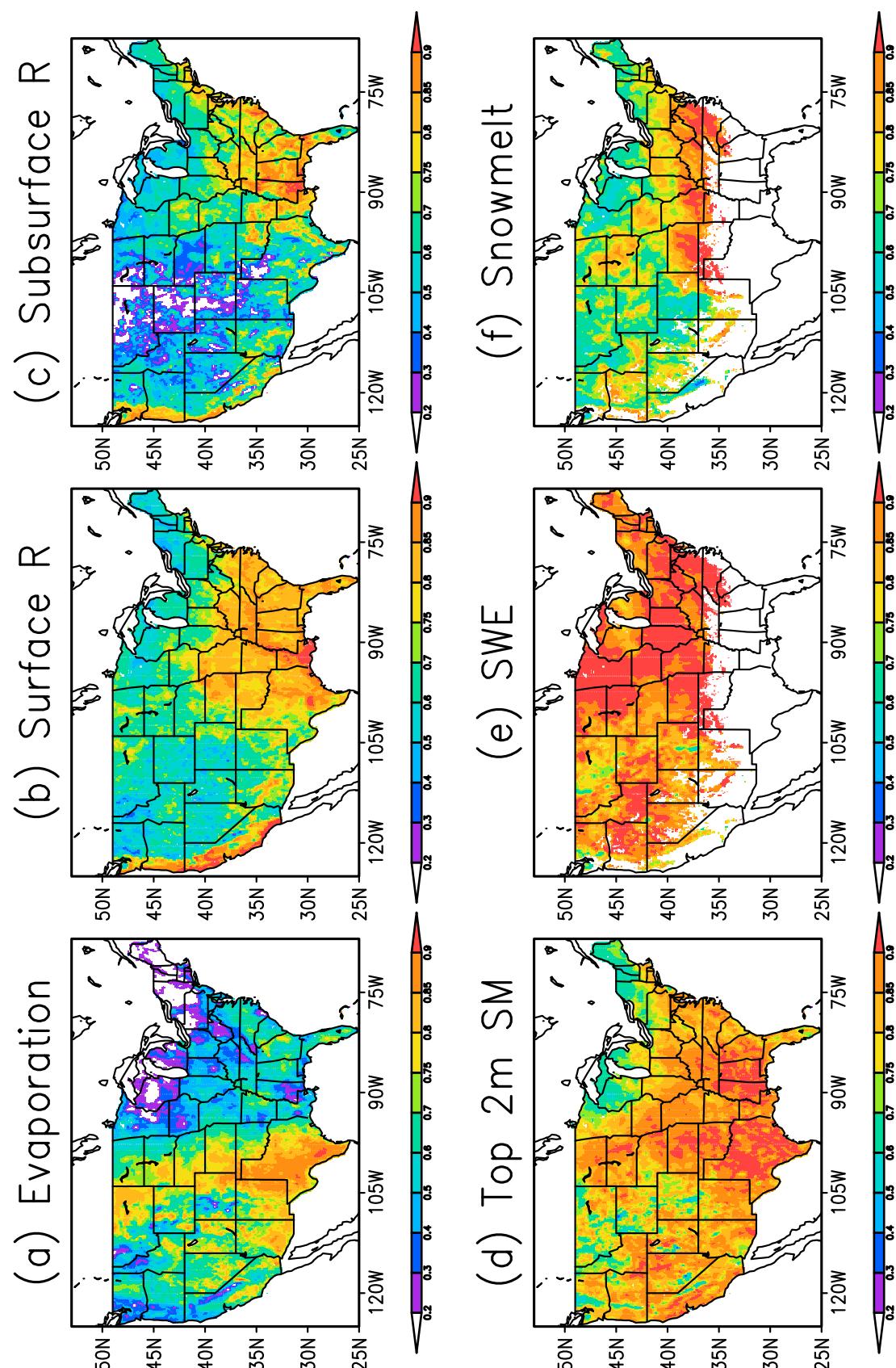


Figure 10. Averaged from the six possible combinations of NLDAS-2 models, the spatial distribution of the anomaly correlation of (a) evaporation, (b) surface R, (c) subsurface R, (d) top 2 m SM, (e) SWE, and (f) snowmelt (R denotes runoff, SM denotes soil moisture, SWE denotes snow water equivalent).

snowmelt anomalies exhibit a north-south gradient except for scattered portions of the interior states in which there is large disparity between models (Figures 9e and 9f). The corresponding anomaly correlations display a general east-west gradient for SWE and a north-south gradient for snowmelt (Figures 10e and 10f).

[53] Overall, the results of the six-variable hydrological analysis show that the model-to-model variation of the monthly anomaly is less, in most cases, than the model-to-model variation of the monthly mean. The disparity between models exhibits a clear spatial distribution which is unique to each variable. Subsurface runoff, in particular, features large model-to-model differences which need to be further investigated.

5. Improvement From NLDAS-1 to NLDAS-2

[54] Figure 11 compares the relative bias of the mean annual runoff for the three models (Noah, SAC-SMA, VIC) that were upgraded between NLDAS-1 and NLDAS-2. Results for Mosaic are not shown because its upgrades were minor. We retrospectively ran both NLDAS-1 and NLDAS-2 model versions from 1 January 1979 to 31 December 2007 using the NLDAS-2 forcing. The mean annual runoff relative bias [(model-observed)/observed] is calculated for 961 small basins (23 km^2 to $10,000 \text{ km}^2$) for 1 October 1979 to 30 September 2007. The observed runoff for a given basin is calculated from the basin area and observed basin-outlet stream discharge information provided by the U.S. Geological Survey (USGS). Generally, the mean model spread or uncertainty for the three models can be represented by the standard deviation among models of each quantity (e.g., runoff). We calculated the spread ratio between NLDAS-1 and NLDAS-2 to identify the change in model spread. If the spread ratio is smaller than 1, this indicates that model uncertainty is reduced, and otherwise this indicates that model uncertainty is increased. For the NLDAS-1 run, Noah overestimates mean annual runoff in the western CONUS and NE (Figure 11a), SAC overestimates mean annual runoff over most of CONUS except for the Great Plains (Figure 11c), and VIC underestimates mean annual runoff over most of CONUS, in particular for the SE (Figure 11c). For the NLDAS-2 run, all three models show an improvement from NLDAS-1 to NLDAS-2, which includes the reduction of negative bias in Noah (Figure 11b) and SAC (Figure 11d) and the positive bias in VIC (Figure 11f) for most regions of CONUS. The spread ratio is smaller than 1 for most regions of CONUS, except for the Great Plains and locations near Lake Superior (Figure 11g), suggesting that model uncertainty is reduced from NLDAS-1 to NLDAS-2 overall. The regions failing to reduce model uncertainty are mainly located in arid and semi-arid climate zones, which have small runoff values. For these regions, all models show an inability to simulate mean annual runoff (Figures 11a–11f) for both NLDAS-1 and NLDAS-2. The possible reasons for this are (1) lack of representation of irrigation in all models, (2) lack of groundwater process in all models, (3) errors in precipitation, and/or inappropriate model parameters. Into these models, calibration may help to improve their simulations in arid and semi-arid regions. Detailed evaluations of the NLDAS-2 simulations against

observations are given in our companion paper (i.e., streamflow) [Xia et al., 2012] and in a series of papers in preparation (e.g., soil moisture, soil temperature, land surface temperature, energy fluxes, and snow).

6. Application of NLDAS-2 Products to Hydrological Monitoring

[55] The preceding analyses showed that the monthly anomalies of the NLDAS-2 hydrological products from the four models have a high level of agreement. This is an important factor in their usefulness for the identification of drought conditions. In general, the top 2 m soil moisture anomaly is indicative of agricultural drought, and the total runoff anomaly is a useful metric of hydrological drought [Sheffield et al., 2004; Andreadis et al., 2005]. Both soil moisture and total runoff are indicators of wet conditions and flood potential.

[56] Figures 12 and 13 show the area-averaged top 2 m soil moisture and total runoff anomalies from the four NLDAS-2 LSMs, for all four quadrants of the CONUS. The results depicted in Figure 12 show that agricultural drought occurred in the northeast quadrant in 1988, the SE quadrant in 1981, 1988, 2000, and 2007, and over the SW quadrant in 1990 and 2002, which is consistent with documented drought events. Abnormally wet years are also identified as 1993 in the northeast and SW quadrants, and 2005 in the SW quadrant. The models exhibit remarkable similarity, although the Mosaic model produces larger anomalies in extreme years (e.g., 1988 in the northeast quadrant) than do the other three models.

[57] Figure 13 depicts such conditions in 1981, 1988, 2000 and 2007 in the SE and in 1988 over the NE. This is consistent with the timing of the agricultural droughts shown in Figure 12. A similar consistency between soil moisture- and runoff-based analyses was found by Andreadis et al. [2005], who used the VIC model in an analysis of long-term U.S. drought. Figure 13 also depicts wet conditions in the NW in 1984, the mid to late 1990s, and in 2006 (Figure 13a), in the NE in 1984 and 1993 (Figure 13b), in the SW in 1983, 1993 and 1995 (Figure 13c), and in the SE in 1983 and 1998 (Figure 13d).

[58] Maps of the 1998 drought and 1993 flood events for the four models are shown in Figures 14 and 15 for soil moisture and runoff, respectively. The maps show large negative anomalies in 1988 in the central U.S. for all four models (Figure 14, left), implying serious drought conditions, although the anomaly magnitudes differ between the models. This is accompanied by lower total runoff than the 28-year average July value (Figure 15, left). The large positive soil moisture anomaly in 1993 (Figure 14, right) appears in the northern states, and over the upper Mississippi and Missouri River basins in all four models, and is accompanied by a large positive total runoff anomaly (Figure 15, right), consistent with observed flood conditions in those regions. The anomaly magnitude and size of the abnormally wet region are different for the four models, with the Noah model tending to have more extreme runoff anomalies. In general, however, all models capture the record 1988 drought over the central U.S., and flooding in 1993 over the upper Mississippi and Missouri River basins.

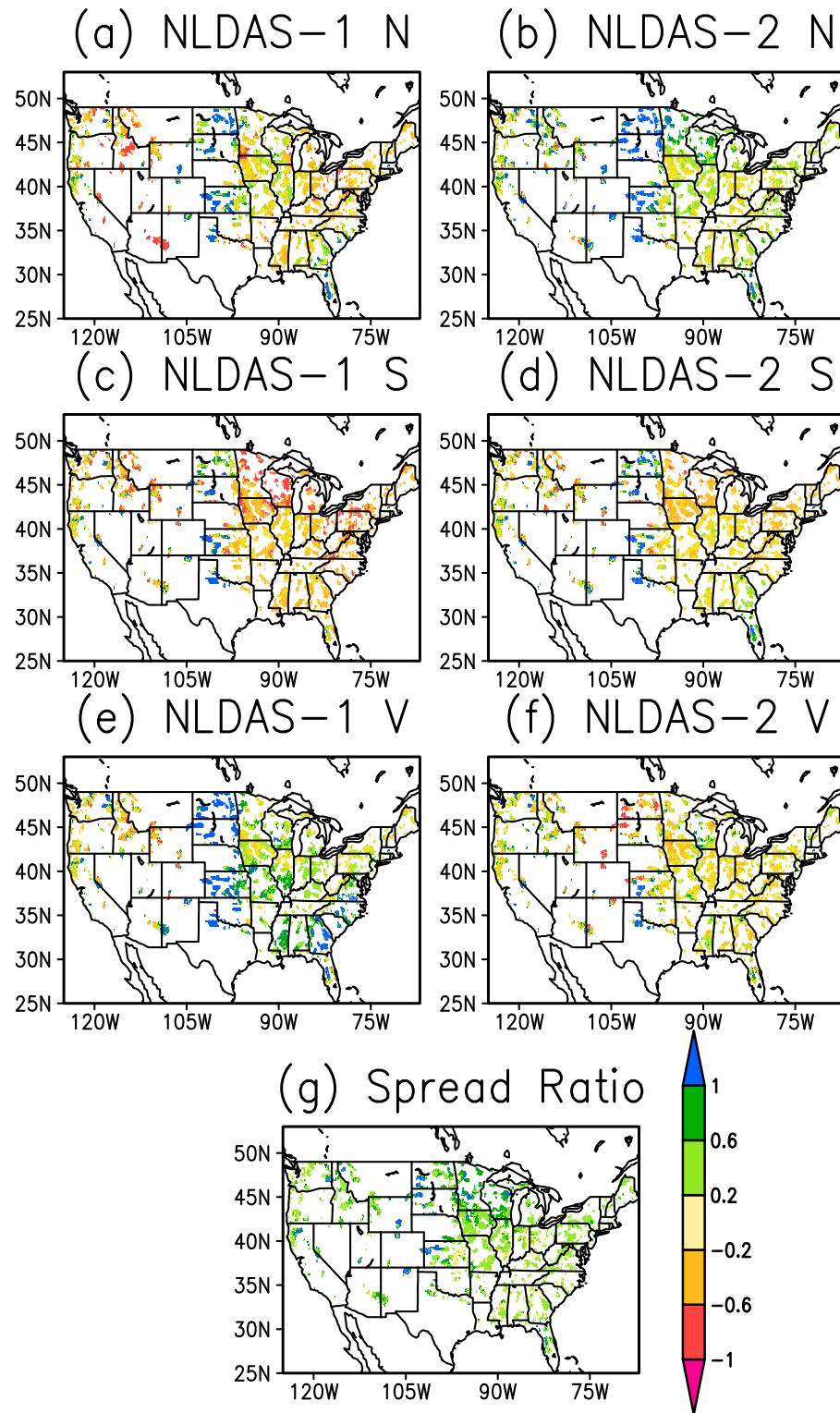


Figure 11. Mean annual runoff relative bias [(model-observed)/observed] for the 961 small basins for 1 October 1979 to 30 September 2007 from (a, b) Noah, (c, d) SAC-SMA, (e, f) VIC and (g) spread ratio (definition in section 3.3). Figures 11a, 11c, and 11e show NLDAS-1 model version simulation, and Figures 11b, 11d, and 11f show NLDAS-2 model version simulation using NLDAS-2 forcing. The observed runoff for a given basin is calculated from basin area and observed basin-outlet stream discharge information provided by U.S. Geological Survey.

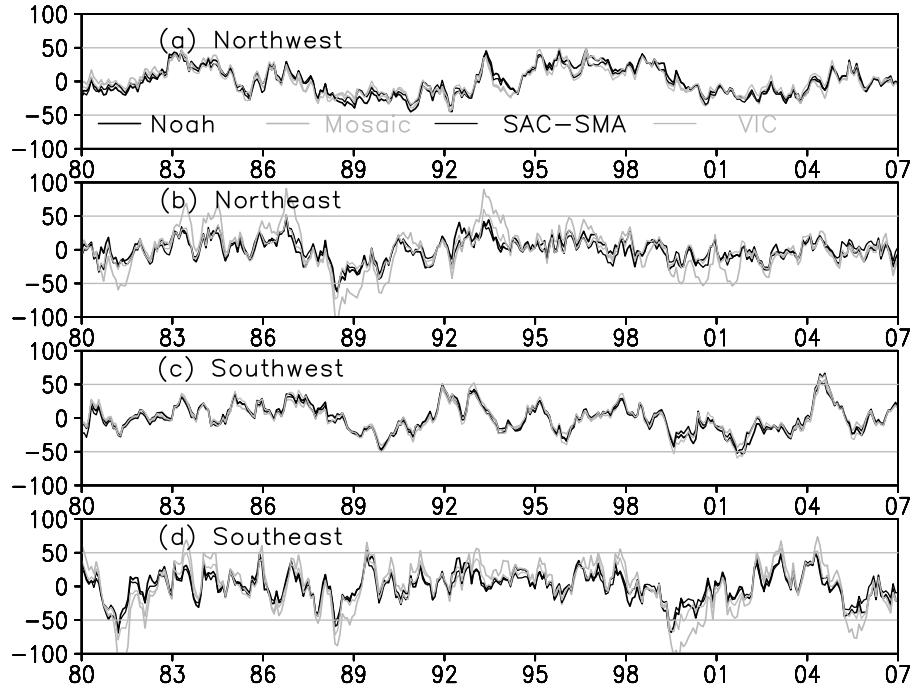


Figure 12. Time series of 28-year area-averaged monthly top 2 m soil moisture anomaly (mm) in NLDAS-2 for four CONUS quadrants: (a) NW, (b) NE, (c) SW, (d) SE (thick black line: Noah model, thick gray line: Mosaic model, thin black line: SAC-SMA model, thin gray line: VIC model).

[59] Monthly anomalies have been the focus of the inter-comparison studies in this paper and are useful for monitoring extended droughts and very large, long-duration floods. However, it should be noted that the NLDAS-2

website also features anomaly and percentile products on shorter timescales (i.e., daily to weekly). These products are more appropriate than the monthly statistics for monitoring

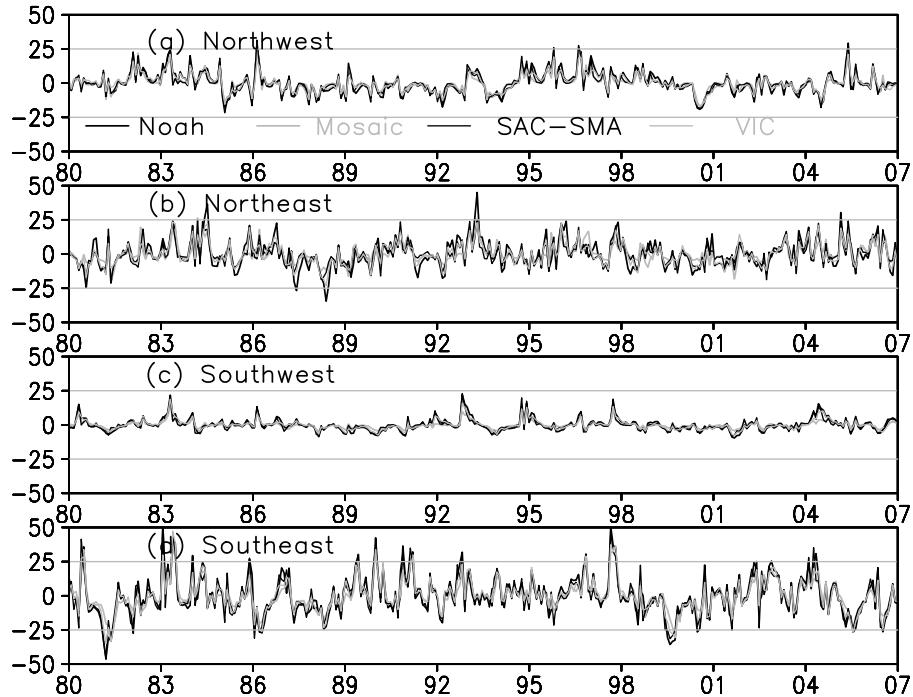


Figure 13. Time series of 28-year area-averaged monthly total runoff anomaly (mm/month) in NLDAS-2 for four CONUS quadrants: (a) NW, (b) NE, (c) SW, (d) SE (thick black line: Noah model, thick gray line: Mosaic model, thin black line: SAC-SMA model, thin gray line: VIC model).

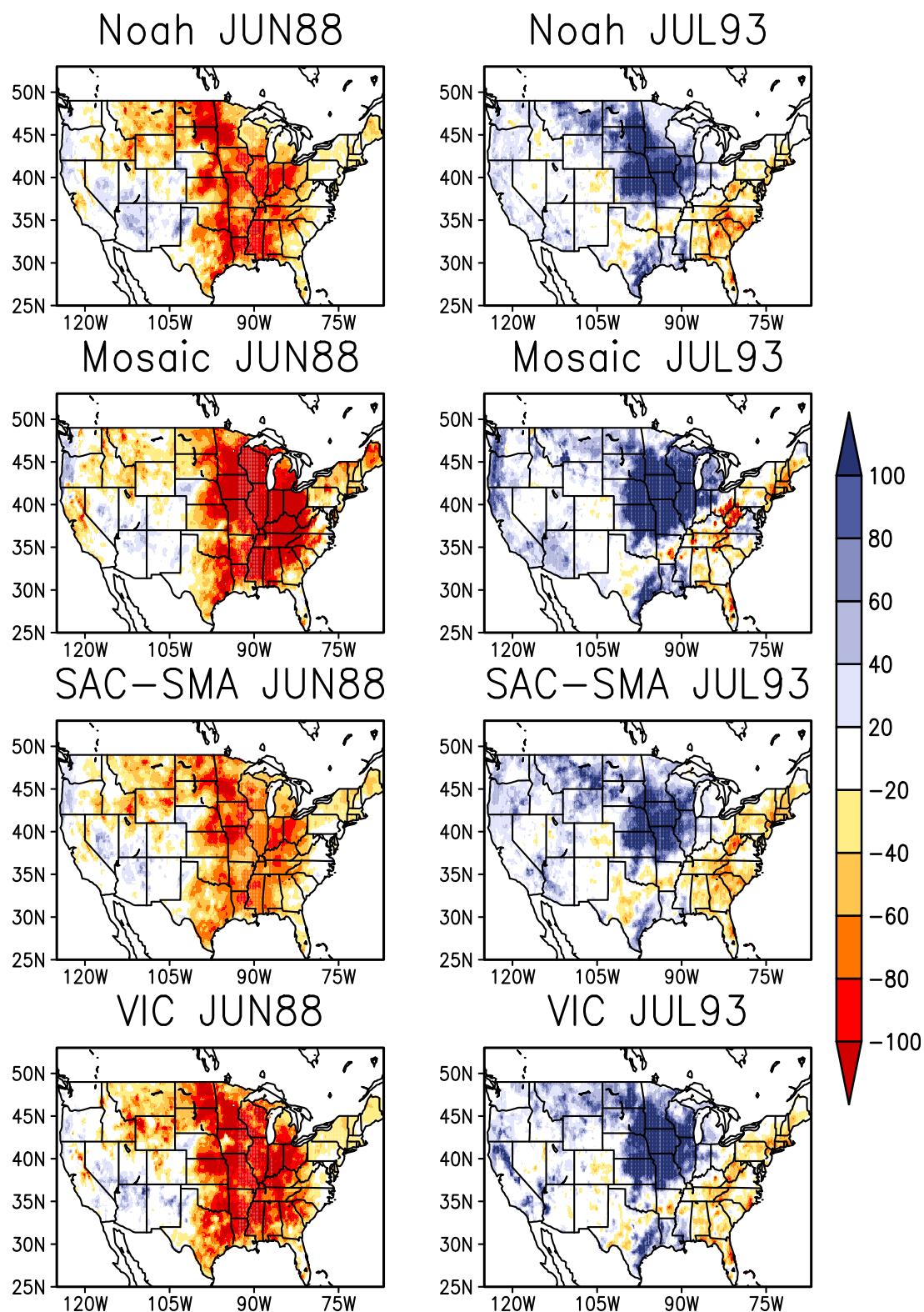


Figure 14. Top 2 m soil moisture anomaly (mm) for (left) June 1988 (drought year) and (right) July 1993 (wet year) for (top to bottom) Noah, Mosaic, SAC-SMA and VIC model.

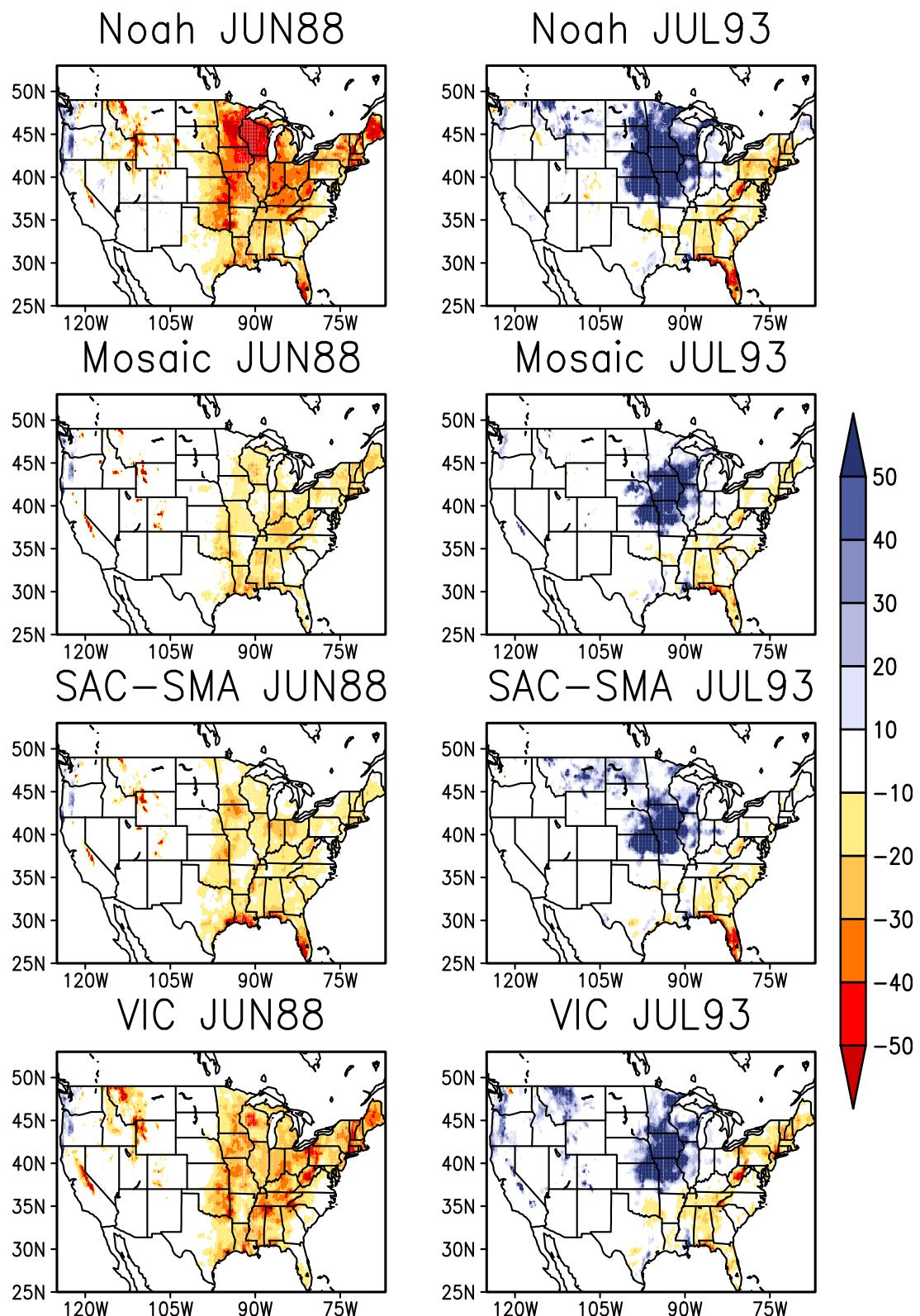


Figure 15. Same as Figure 14 except for monthly mean total runoff anomaly.

short-term yet severe droughts, as well as the majority of flood events.

7. Summary and Conclusions

[60] This paper is part one of a two-part article detailing the analysis and validation of output from the four LSMs executed as part of the NLDAS-2 project. Part one focused on a model-to-model comparison and on the application of multimodel products, while the second part [Xia et al., 2012] focuses on the validation of multimodel simulated streamflow and evaporation against observations. In this study we have intercompared NLDAS-2 multimodel mean annual and seasonal values of water fluxes, energy fluxes, and state variables, and have investigated the level of multimodel similarity in terms of magnitude and temporal variability.

[61] NLDAS-2 output offers many improvements over the original NLDAS-1 data, and is a testament to the progress made in NLDAS-1 in identifying errors in the forcing data and models, and in instigating related improvements. This improvement from NLDAS-1 to NLDAS-2 has been briefly discussed and summarized in section 5. The main distinctions between the two NLDAS phases are as follows:

[62] 1. The source of forcing data and the biases inherent in the forcing: NLDAS-1 utilized EDAS-based data as the underlying backbone for its surface forcing data set, while NLDAS-2 fills this role with NARR data. Both phases of NLDAS made use of gauge-based daily precipitation analyses. The daily precipitation data are disaggregated to an hourly timescale using precipitation estimates from NCEP Stage-II Doppler radar data. NLDAS-1, however, did not make any adjustments for the effects of topography on the gauge-based daily precipitation, which resulted in large biases in streamflow and SWE [Lohmann et al., 2004; Pan et al., 2003]. Based on these findings, the gauge-based daily precipitation was adjusted for such effects in NLDAS-2 using the PRISM data set [Daly et al., 1994]. The second problem which characterized the NLDAS-1 forcing data was the spatial discontinuity in downward shortwave radiation caused by the merging of GOES satellite data and EDAS model data. Addressing this issue in NLDAS-2, a ratio-based correction algorithm with special edge treatment was used to bias-correct the NARR downward shortwave radiation using five years (1996–2000) of GOES data.

[63] 2. Model upgrades, parameter re-calibration, and changes to PE forcing data: The results of NLDAS-1 led to improvements in model code, parameters, and PE forcing data within NLDAS-2. These alterations in turn led to a higher level of agreement among the four models for all variables when compared against results from NLDAS-1.

[64] 3. The time period of the forcing data and simulations: The short three-year simulation period (1996–1999) of NLDAS-1 prevented the calculation of anomalies and percentiles which can be used to support drought and flood analysis and monitoring. By contrast, the 30 year period of NLDAS-2 has allowed for an extensive analysis of NLDAS land surface products, especially in the context of longer term variability and drought monitoring.

[65] 4. The derivation and application of multimodel ensemble products: The longer time period of NLDAS-2 has supported a more robust assessment of multimodel products, including the evaluation of model disparity for different

variables. The model products have been demonstrated to show consistency which is important for effective drought monitoring, and the ensemble mean has been shown to offer increased simulation skill over individual models [Xia et al., 2012].

[66] As noted above, there is a higher level of agreement among the four models in NLDAS-2 than in NLDAS-1. However, substantial differences remain in certain flux and state variables. The most noticeable differences are located in the NE, and the Lake Superior and western mountainous regions, where large disparities exist in cold-season variables such as snowpack and frozen soil. The treatment of the processes governing these variables differs among the four models. For example, only VIC includes sub-grid elevation banding which tends to make snow persist for longer periods in topographically complex regions. Additionally, sublimation is included in the Noah, Mosaic and VIC models, but not in Snow17, the snow model companion of SAC-SMA.

[67] Other differences are apparent in the evaporation and subsurface runoff fields and are related to how soil hydrologic processes are simulated in each model. While the Noah, Mosaic and VIC models all have surface infiltration schemes which account for sub-grid variability in soil moisture and precipitation, the methods differ, as do the approaches for modeling drainage. Furthermore, although all three of these models also include direct evaporation from bare soil, transpiration from vegetation, evaporation of interception, and snow sublimation, the formulation of these processes differs, along with their canopy resistance parameters, their vegetation phenology, and their root profiles. The SAC-SMA model differs even more, using a “two-reservoir” soil water storage structure (a shallow upper reservoir and a deeper lower reservoir) and calculating evaporation as a fraction of input potential evaporation. These different treatments lead to disparities in evaporation and subsurface runoff among the four models, and further work is needed to isolate the contribution of each process and method to the differences seen.

[68] A major goal of NLDAS-2 is to support the activities of the National Integrated Drought Information System (NIDIS) through the provision of land surface anomalies and percentiles. With this in mind, examples of multimodel output were shown in this study for the record drought event over the central U.S. in June of 1988, and the record large-scale flood event over the upper Mississippi and Missouri basins in 1993. The positive results from the monthly timescale analyses detailed in this paper provide confidence in the suitability of the multimodel products for drought monitoring activities. Additionally, the NLDAS-2 website provides anomaly and percentile fields at the shorter daily to-weekly timescales that are more suited for monitoring floods and short-term droughts.

[69] CPC scientists have used NLDAS-2 soil moisture and total runoff to analyze and monitor droughts for their monthly drought briefings, and have used NLDAS-2 soil moisture percentiles to assist in the construction of their seasonal drought outlooks. Furthermore, NLDAS-2 soil moisture and total runoff percentiles have been used by authors of the U.S. Drought Monitor (USDM, <http://www.drought.gov>) as one of their reference data sources to construct the USDM analyses. With the growing need to support such operational activities, NLDAS-2 will soon be switched from a quasi-operational run mode at NCEP/EMC, to a fully

operational, more robust implementation. At the same time, we will also initiate an intensive evaluation strategy to assess NLDAS-2 fluxes and state variables using in situ measurements and satellite-based products to identify model shortcomings and thus guide the direction of future model improvements.

[70] All NLDAS-2 forcing and model output data are freely available from the NOAA/NCEP (<http://www.emc.ncep.noaa.gov/mmb/nldas/>) and NASA/GSFC (<http://disc.sci.gsfc.nasa.gov/hydrology/data-holdings>) public data servers.

[71] **Acknowledgments.** This work was supported by the NOAA/CPO/CPA core project.

References

- Abdula, F. A., D. P. Lettenmaier, E. F. Wood, and J. A. Smith (1996), Application of a macroscale hydrologic model to estimate the water balance of the Arkansas-Red River basin, *J. Geophys. Res.*, **101**, 7449–7459, doi:10.1029/95JD02416.
- Anderson, E. (1973), National Weather Service river forecast system—Snow accumulation and ablation model, *NOAA Tech. Memo. NWS HYDRO-17*, 217 pp., Natl. Oceanic and Atmos. Admin., Silver Spring, Md.
- Andreidis, K. M., E. A. Clark, A. W. Wood, A. F. Hamlet, and D. P. Lettenmaier (2005), 20th century drought in the conterminous United States, *J. Hydrometeorol.*, **6**, 985–1001, doi:10.1175/JHM450.1.
- Betts, A., F. Chen, K. Mitchell, and Z. Janic (1997), 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, doi:10.1175/1520-0493(1997)125<2896:AOTLSA>2.0.CO;2.
- Bowling, L. C., et al. (2003), Simulation of high latitude processes in the Torne-Kalix basin: PILPS Phase 2(e): 1: Experiment description and summary comparisons, *Global Planet. Change*, **38**, 1–30, doi:10.1016/S0921-8181(03)00003-1.
- Brock, F. V., K. C. Crawford, R. L. Elliott, G. W. Cuperus, S. J. Stadler, H. Johnson, and M. D. Ellits (1995), The Oklahoma Mesonet: A technical overview, *J. Atmos. Oceanic Technol.*, **12**, 5–19, doi:10.1175/1520-0426(1995)012<0005:TOMATO>2.0.CO;2.
- Burnash, R. J. C. (1995), The NWS river forecast system—catchment modeling, in *Computer Models of Watershed Hydrology*, edited by V. P. Singh, pp. 311–366, Water Resour. Publ., Littleton, Colo.
- Burnash, R. J. C., R. L. Ferral, and R. A. McGuire (1973), A generalized streamflow simulation system: Conceptual models for digital computer, technical report, Joint Fed. State River Forecast Cent., Sacramento, Calif.
- Chen, F., Z. Janic, and K. Mitchell (1997), Impact of atmospheric surface-layer parameterizations in the new land-surface scheme of the NCEP mesoscale Eta model, *Boundary Layer Meteorol.*, **85**, 391–421, doi:10.1023/A:1000531001463.
- Chen, T. H., et al. (1997), Cabauw experimental results from the Project for Intercomparison of Landsurface Parameterization Scheme (PILPS), *J. Clim.*, **10**, 1194–1215, doi:10.1175/1520-0442(1997)010<1194:CERFTP>2.0.CO;2.
- Cherkauer, K. A., and D. P. Lettenmaier (1999), Hydrological effects of frozen soils in the upper Mississippi River basin, *J. Geophys. Res.*, **104**, 19,599–19,610, doi:10.1029/1999JD900337.
- Cherkauer, K. A., L. C. Bowling, and D. P. Lettenmaier (2003), Variable Infiltration Capacity (VIC) cold land process model updates, *Global Planet. Change*, **38**, 151–159, doi:10.1016/S0921-8181(03)00025-0.
- Cosgrove, B. A., et al. (2003a), 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.
- Cosgrove, B. A., et al. (2003b), Land surface model spin-up behavior in the North American Land Data Assimilation System (NLDAS), *J. Geophys. Res.*, **108**(D22), 8845, doi:10.1029/2002JD003316.
- Daly, C., R. P. Neilson, and D. L. Phillips (1994), A statistical-topographic model for mapping climatological precipitation over mountainous terrain, *J. Appl. Meteorol.*, **33**, 140–158, doi:10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2.
- Dirmeyer, P. A., A. J. Dolman, and N. Sato (1999), The pilot phase of the global soil wetness project, *Bull. Am. Meteorol. Soc.*, **80**, 851–878, doi:10.1175/1520-0477(1999)080<0851:TPPOTG>2.0.CO;2.
- Dirmeyer, P. A., X. Gao, M. Zhao, Z. Guo, T. Oki, and N. Hanasaki (2006), Multimodel analysis implications for our perception of the land surface, *Bull. Am. Meteorol. Soc.*, **87**, 1381–1397, doi:10.1175/BAMS-87-10-1381.
- Ek, M. B., K. E. Mitchell, Y. Lin, E. Rodgers, P. Grunman, V. Koren, G. Gayno, and J. D. Tarpley (2003), Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model, *J. Geophys. Res.*, **108**(D22), 8851, doi:10.1029/2002JD003296.
- Feng, X., A. Sahoo, K. Arsenault, P. Houser, Y. Luo, and T. J. Troy (2008), The impact of snow model complexity at three CLPX sites, *J. Hydrometeorol.*, **9**, 1464–1481, doi:10.1175/2008JHM860.1.
- Gong, G., D. Entekhabi, and G. D. Salvucci (1994), Regional and seasonal estimates of fractional storm coverage based on station precipitation observations, *J. Clim.*, **7**, 1495–1505, doi:10.1175/1520-0442(1994)007<1495:RASEOF>2.0.CO;2.
- Koren, V., J. Schaake, K. E. Mitchell, Q. Duan, F. Chen, and J. Baker (1999), A parameterization of snowpack and frozen ground intended for NCEP weather and climate models, *J. Geophys. Res.*, **104**, 19,569–19,585, doi:10.1029/1999JD900232.
- Koren, V., M. Smith, D. Wang, and Z. Zhang (2000), Use of soil property data in the derivation of conceptual rainfall-runoff model parameters, paper presented at 15th Conference on Hydrology, Am. Meteorol. Soc., Long Beach, Calif., 10–14 Jan.
- Koren, V., S. Reed, M. Smith, Z. Zhang, and D.-J. Seo (2004), Hydrology laboratory research modeling system (HL-RMS) of the US National Weather Service, *J. Hydrol.*, **291**, 297–318, doi:10.1016/j.jhydrol.2003.12.039.
- Koren, V., M. Smith, Z. Cui, and B. Cosgrove (2007), Physically based modifications to the Sacramento Soil Moisture Accounting model: Modeling the effects of frozen ground on the runoff generation process, *NOAA Tech. Rep.*, **52**, 43 pp., Natl. Oceanic and Atmos. Admin., Silver Spring, Md.
- Koren, V., M. Smith, Z. Cui, B. Cosgrove, K. Warner, and R. Zamora (2010), Modification of Sacramento Soil Moisture Accounting Heat Transfer component (SAC-HT) for enhanced evapotranspiration, *NOAA Tech. Rep.*, **53**, 66 pp., Natl. Oceanic and Atmos. Admin., Silver Spring, Md.
- Koster, R., and M. Suarez (1994), The components of the SVAT scheme and their effects on a GCM's hydrological cycle, *Adv. Water Resour.*, **17**, 61–78, doi:10.1016/0309-1708(94)90024-8.
- Koster, R., and M. Suarez (1996), Energy and water balance calculations in the Mosaic LSM, *NASA Tech. Memo.*, NASA TM-104606, vol. 9, 60 pp.
- Koster, R., M. Suarez, and M. Heiser (2000), Variance and predictability of precipitation at seasonal-to-interannual timescales, *J. Hydrometeorol.*, **1**, 26–46, doi:10.1175/1525-7541(2000)001<0026:VAPOPA>2.0.CO;2.
- Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges (1994), A simple hydrologically based model of land surface water and energy fluxes for GCMs, *J. Geophys. Res.*, **99**, 14,415–14,428, doi:10.1029/94JD00483.
- Livneh, B., Y. Xia, M. B. Ek, K. E. Mitchell, and D. Lettenmaier (2010), Noah LSM snow model diagnostics and enhancements, *J. Hydrometeorol.*, **11**, 721–738, doi:10.1175/2009JHM1174.1.
- Livneh, B., P. J. Restrepo, and D. P. Lettenmaier (2011), Development of a unified land model for prediction of surface hydrology and land-atmosphere interactions, *J. Hydrometeorol.*, **12**, 1299–1320.
- Lobmeyr, M., D. Lohmann, and C. Ruhe (1999), An application of a large scale conceptual hydrological model over the Elbe region, *Hydrol. Earth Syst. Sci.*, **3**, 363–374, doi:10.5194/hess-3-363-1999.
- Lohmann, D., E. Raschke, B. Nijssen, and D. P. Lettenmaier (1998), Regional scale hydrology, part II: Application of the VIC-2 L model to the Weser River, Germany, *Hydrol. Sci. J.*, **43**, 143–158, doi:10.1080/026266980942108.
- Lohmann, D., et al. (2004), Streamflow and water balance intercomparisons of four land surface models in the North American Land Data Assimilation System project, *J. Geophys. Res.*, **109**, D07S91, doi:10.1029/2003JD003517.
- Luo, L., et al. (2003), 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.
- Maurer, E. P., A. W. Wood, J. C. Adam, D. P. Lettenmaier, and B. Nijssen (2002), A long-term hydrologically based data set of land surface fluxes and states for the conterminous United States, *J. Clim.*, **15**, 3237–3251, doi:10.1175/1520-0442(2002)015<3237:ALTHBD>2.0.CO;2.
- Mesinger, F., et al. (2006), North American regional reanalysis, *Bull. Am. Meteorol. Soc.*, **87**(3), 343–360, doi:10.1175/BAMS-87-3-343.
- Mitchell, K. E., et al. (2004), The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system, *J. Geophys. Res.*, **109**, D07S90, doi:10.1029/2003JD003823.
- Nijssen, B., D. P. Lettenmaier, X. Liang, S. W. Wetzel, and E. F. Wood (1997), Streamflow simulation for continental-scale river basins, *Water Resour. Res.*, **33**, 711–724, doi:10.1029/96WR03517.

- Nijssen, B., G. M. O'Donnell, D. P. Lettenmaier, D. Lohmann, and E. F. Wood (2001), Predicting the discharge of global rivers, *J. Clim.*, 14, 3307–3323, doi:10.1175/1520-0442(2001)014<3307:PTDOGR>2.0.CO;2.
- Nijssen, B., et al. (2003), Simulation of high latitude hydrological processes in the Torne-Kalix basin: PILPS Phase 2e. 2: Comparison of model results with observations, *Global Planet. Change*, 38, 31–53, doi:10.1016/S0921-8181(03)00004-3.
- Pan, M., et al. (2003), 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.
- Pinker, R. T., et al. (2003), 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.
- Robinson, D. A., and G. Kukla (1985), Maximum surface albedo of seasonally snow covered lands in the Northern Hemisphere, *J. Clim. Appl. Meteorol.*, 24, 402–411.
- Robock, A., et al. (2003), 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.
- Schlosser, C. A., et al. (2000), Simulations of a boreal grassland hydrology at Valdai, Russia: PILPS 2(d), *Mon. Weather Rev.*, 128, 301–321, doi:10.1175/1520-0493(2000)128<0301:SOABGH>2.0.CO;2.
- Sheffield, J., et al. (2003), Snow process modeling in the North American Land Data Assimilation System (NLDAS): 1. Evaluation of model-simulated snow cover extent, *J. Geophys. Res.*, 108(D22), 8849, doi:10.1029/2002JD003274.
- Sheffield, J., G. Goteti, F. Wen, and E. F. Wood (2004), A simulated soil moisture based drought analysis for the United States, *J. Geophys. Res.*, 109, D24108, doi:10.1029/2004JD005182.
- Sheffield, J., G. Goteti, and E. F. Wood (2006), Development of a 50-yr high-resolution global dataset of meteorological forcings for land surface modeling, *J. Clim.*, 19, 3088–3111, doi:10.1175/JCLI3790.1.
- Slater, A. G., T. J. Bohn, J. L. McCreight, M. C. Serreze, and D. P. Lettenmaier (2007), A multimodel simulation of Pan-Arctic hydrology, *J. Geophys. Res.*, 112, G04S45, doi:10.1029/2006JG000303.
- Smith, M. B., D.-J. Seo, V. I. Koren, S. M. Reed, Z. Zhang, Q. Duan, F. Moreda, and S. Cong (2004), The distributed model intercomparison project (DMIP): Motivation and experiment design, *J. Hydrol.*, 298, 4–26.
- Smith, M., et al. (2012), Results of the DMIP 2 Oklahoma Experiments, *J. Hydrol.*, doi:10.1016/j.jhydrol.2011.08.056, in press.
- Stokes, G. M., and S. E. Schwartz (1994), The Atmospheric Radiation Measurement (ARM) program: Programmatic background and design of the cloud and radiation testbed, *Bull. Am. Meteorol. Soc.*, 75, 1201–1221, doi:10.1175/1520-0477(1994)075<1201:TARMPP>2.0.CO;2.
- Troy, T. J., E. F. Wood, and J. Sheffield (2008), An efficient calibration method for continental-scale land surface modeling, *Water Resour. Res.*, 44, W09411, doi:10.1029/2007WR006513.
- Wang, A., T. J. Bohn, S. P. Mahanama, R. D. Koster, and D. P. Lettenmaier (2009), Multimodel ensemble reconstruction of drought over the continental United States, *J. Clim.*, 22, 2694–2712, doi:10.1175/2008JCLI2586.1.
- Wei, H., Y. Xia, K. E. Mitchell, and M. B. Ek (2011), Improvement of Noah land surface model for the warm season processes: Assessment of water and energy flux simulation, *Hydrolog. Processes*, in press.
- Wood, E. F., D. P. Lettenmaier, X. Liang, B. Nijssen, and S. W. Wetzel (1997), Hydrological modeling of continental-scale basins, *Annu. Rev. Earth Planet. Sci.*, 25, 279–300, doi:10.1146/annurev.earth.25.1.279.
- Wood, E., et al. (1998), The Project for Intercomparison of Land-Surface Parameterization Scheme (PILPS) Phase 2(c) Red-Arkansas River Basin Experiment: 1. Experiment description and summary intercomparisons, *Global Planet. Change*, 19, 115–135, doi:10.1016/S0921-8181(98)00044-7.
- Xia, Y., et al. (2012), Continental-scale water and energy flux analysis and validation for North-American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of simulated streamflow, *J. Geophys. Res.*, 117, D03110, doi:10.1029/2011JD016051.
- Yang, Z.-L., R. E. Dickinson, A. Henderson-Sellers, and A. J. Pitman (1995), Preliminary study of spin-up processes in land surface models with the first stage data of Project for Intercomparison of Land-Surface Parameterization Scheme Phase 1(a), *J. Geophys. Res.*, 100, 16,553–16,578.
- C. Alonge, AWS Truewind, LLC, 463 New Karner Rd., Albany, NY 12205, USA.
- B. Cosgrove and V. Koren, Office of Hydrologic Development, National Weather Service, NOAA, 1325 East West Hwy., Silver Spring, MD 20910, USA.
- Q. Duan, College of Global Change and Earth System Science, Beijing Normal University, Beijing 100875, China.
- M. Ek, J. Meng, K. Mitchell, H. Wei, and Y. Xia, Environmental Modeling Center, National Centers for Environment and Prediction, NOAA, 5200 Auth Rd., Camp Springs, MD 20746, USA. (Youlong.Xia@noaa.gov)
- Y. Fan, Office of Science and Technology, National Weather Service, 1325 East West Hwy., Silver Spring, MD 20910, USA.
- D. Lettenmaier and B. Livneh, Department of Environmental and Civil Engineering, University of Washington, Seattle, WA 98195, USA.
- L. Luo, Department of Geography, Michigan State University, East Lansing, MI 48824, USA.
- K. Mo, Climate Prediction Center, National Centers for Environment and Prediction, NOAA, 5200 Auth Rd., Camp Springs, MD 20746, USA.
- D. Mocko, SAIC at the Hydrological Sciences Laboratory and Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Code 610.1, Greenbelt, MD 20715, USA.
- J. Sheffield and E. Wood, Department of Environmental and Civil Engineering, Princeton University, Princeton, NJ 08544, USA.