The Multi-institution North American Land Data Assimilation System (NLDAS): Utilizing Multiple GCIP Products and Partners in a Continental Distributed Hydrological Modeling System

Kenneth E. Mitchell[1], Dag Lohmann[1], Paul R. Houser[2], Eric F. Wood[3], John C. Schaake[4], Alan Robock[5], Brian A. Cosgrove[2], Justin Sheffield[3], Qingyun Duan[4], Lifeng Luo[5], R. Wayne Higgins[6], Rachel T. Pinker[7], J. Dan Tarpley[8], Dennis P. Lettenmaier[9], Curtis H. Marshall[1,10], Jared K. Entin[2,11], Ming Pan[3], Wei Shi[6], Victor Koren[4], Jesse Meng[1,2], Bruce H. Ramsay[8], Andrew A. Bailey[8]

[1] NCEP Environmental Modeling Center (NOAA/NWS)

[2] NASA/GSFC Hydrological Sciences Branch and Data Assimilation Office

[3] Department of Civil and Environmental Engineering, Princeton University

[4] Office of Hydrologic Development (NOAA/NWS)

[5] Department of Environmental Sciences, Rutgers University

[6] NCEP Climate Prediction Center (NOAA/NWS)

[7] Department of Meteorology, University of Maryland

[8] NESDIS Office of Research and Applications

[9] Department of Civil and Environmental Engineering, University of Washington

Current affiliations

[10] Department of Atmospheric Sciences, Colorado State University

[11]NASA Headquarters

For *Journal of Geophysical Research – Atmospheres*

GCIP3 Special Issue

May 31, 2003

Corresponding Author:

Dr. Kenneth Mitchell

Environmental Modeling Center (W/NP2, WWB)

National Centers for Environmental Prediction (NCEP)

NOAA Science Center (Room 207)

5200 Auth Road

Camp Springs, MD 20746-4304

Email: Kenneth.Mitchell@noaa.gov

Phone: (301)-763-8000 (Ext 7225)

Fax: (301)-763-8545

Abstract

Results are presented from the multi-institution partnership to develop a realtime and retrospective North American Land Data Assimilation System (NLDAS). NLDAS consists of a) four land models executing in parallel in uncoupled mode, b) common hourly land surface forcing and c) common streamflow routing -- all using a 1/8° grid over the continental U.S. (CONUS). The initiative is largely sponsored by the Global Energy and Water Cycle Experiment (GEWEX) Continental-scale International Project (GCIP). As the overview for nine companion NLDAS papers, this paper describes and evaluates the 3-year NLDAS execution of October 1, 1996 to September 30, 1999 – a period rich in observational data for validation. The validation emphasizes 1) the land states, fluxes and input forcing of four land models, 2) the application of new GCIP-sponsored products, and 3) a multi-scale approach. The validation includes a) mesoscale observing networks of land surface forcing, fluxes and states, b) regional snowpack measurements, c) daily streamflow measurements and d) satellite-based retrievals of snow cover, land surface skin temperature (LST) and surface insolation. The results show substantial inter-model differences in soil moisture storage, surface evaporation and runoff (especially over non-sparse vegetation), snowpack and LST. Owing to unexpectedly large inter-model differences in aerodynamic conductance, inter-model differences in mid-day summer LST were distorted relative to those anticipated from the inter-model differences in Bowen ratio. Lastly, anticipating future assimilation of LST, this study assesses geostationary satellite-derived LST, determines the latter to be of good quality, and applies the latter to validate model LST.

1. Introduction

Improving weather and seasonal climate prediction by dynamical models requires multidisciplinary advances in providing reliable initial states for the atmosphere, land and ocean components of the earth system. For two decades, advances in providing atmospheric initial states via 4-dimensional data assimilation (4DDA) have paved the way for the development of counterpart 4DDA systems for the ocean and land. In 4DDA, a geophysical model provides temporally and spatially continuous background states into which temporally and spatially discontinuous observations are assimilated from various observing platforms (in situ, satellite, radar). The backbone then of any atmospheric, ocean or land 4DDA system is the geophysical model whose day-to-day execution provides the continuous timeline of background states.

Present space-based microwave estimates of soil moisture sense only the top 1-5 cm of soil, far short of the depths needed (e.g. root zone) for land state initialization. Hence, a land data assimilation system (LDAS) is needed to blend sparse land observations with the background fields of a reliable land surface model (LSM). The accuracy of the LSM background field (and the attendant surface and sub-surface water/energy fluxes that drive those background fields) is crucial to the viability of an LDAS. The primary objective of the NLDAS study here is to provide and validate -- over multi-year executions with common surface forcing over the CONUS domain -- the background land states and surface fluxes of four LSMs (described in Section 2.3): Noah, Mosaic, VIC, and Sacramento – denoted SAC. (Hereafter, all acronyms are defined in Appendix). Future NLDAS papers will address actual data assimilation experiments using such methods as adjoint models and Kalman filtering. As a preparatory step toward near-future assimilation of satellite LST, this paper includes an assessment of geostationary satellite-derived LST and application of the latter to validate NLDAS LST.

It is instructive to consider the infancy of realtime large-scale land 4DDA, compared to realtime large-scale atmospheric or ocean 4DDA. Global atmospheric 4DDA has been a mainstay of operational NWP centers since the late 1970's. Realtime ocean 4DDA on large-scale ocean basins followed in the late 1980's [*Ji et al.*, 1994] on the heels of the TOGA program, and the extension of such to the global ocean is now a thrust of the CLIVAR program. Yet, until the mid 1990's, initiatives in realtime continental or global land 4DDA were virtually non-existent. The first viable examples of realtime land 4DDA on continental or global scales were the coupled land-atmosphere 4DDA systems at major NWP centers such as NCEP [*Kalnay et al*., 1996] and the European Center for Medium Range Weather Forecasting [*Gibson et al*., 1997]. Such coupled land-atmosphere 4DDA systems (including global reanalysis) often yield significant errors and drift in a) soil moisture and temperature, and b) surface energy and water fluxes, owing to substantial biases in the surface forcing from the parent atmospheric models. To constrain such errors and drift, coupled land-atmosphere 4DDA systems temporally nudge the soil moisture by such means as 1) a climatology of soil moisture [*Kalnay et al*., 1996] or 2) differences between the observed and 4DDA background fields of precipitation [*Kanamitsu et al*., 2002] or screen-level air temperature and dew point [*Douville et al*., 2000]. Such nudging methods, however, do not reduce the main error source, namely large bias in the land surface forcing (especially precipitation and solar insolation) of the parent atmospheric model.

Large biases in atmospheric model surface forcing also plague ocean 4DDA, especially biases in the crucial surface momentum flux (wind vector). To overcome this, "flux corrections" are widely used in ocean 4DDA to correct atmospheric model surface fluxes before they force the ocean model. Motivated by this practice in ocean 4DDA, we pursue a "flux correction" approach here in NLDAS. As a pathfinder in this regard, the GEWEX Global Soil Wetness

Project [GSWP, *Dirmeyer et al*., 1999] demonstrated in retrospective mode the viability of using non-model, observation-based precipitation analyses and non-model, satellite-based surface insolation fields (combined with the remaining near-surface forcing fields from atmospheric 4DDA) to drive distributed, uncoupled, land-surface models over the global domain. However, the GSWP used monthly satellite retrievals of precipitation and insolation, which are not conducive to the prospect of daily updates to initialize operational prediction models. Thus, as an extension to the GSWP paradigm, we set out the key goals to develop and execute the following: 1) the first realtime operational prototype of a continental-scale uncoupled land 4DDA backbone (continuously cycled model-background states) that could be executed at NCEP on a daily update basis using realtime streams of hourly to daily data, and 2) a companion retrospective mode for research use, spanning October 1996 to the present, via GCIP-sponsored archives of NOAA operational data streams. NLDAS thus provides a land 4DDA counterpart from the GEWEX community, to complement the ocean 4DDA thrusts of CLIVAR.

In particular, the GCIP program of GEWEX has stimulated multi-disciplinary research by meteorologists and hydrologists in land data assimilation. Moreover, a core thrust of GCIP is the infusion of GCIP research into NOAA operational practice for the prediction of climate, weather, and water. The hands-on multi-institution collaboration in the NLDAS project here is a flagship of GCIP success in such infusion. The NOAA operational partners in NLDAS include NCEP/EMC and OHD of the NWS and NESDIS/ORA, who have joined with the NLDAS research partners of NASA GSFC, Princeton University, Rutgers University, University of Maryland, University of Oklahoma, and University of Washington. These partners have developed, executed, and evaluated a realtime and retrospective uncoupled NLDAS. The NLDAS generates hourly surface forcing (anchored by model-independent, observation-based

precipitation and solar insolation fields) that drives four LSMs running in parallel to produce hourly output on a common 1/8° grid over a CONUS domain. This paper gives an overview of the methodology and results of the initial development and evaluation of NLDAS, by providing an overview of the following ten papers in this issue by the NLDAS partners:

Cosgrove et al. [2003a]: production of NLDAS realtime and retrospective surface forcing

Luo et al. [2003]: validation of NLDAS surface forcing

Lohmann et al. [2003]: production and validation of NLDAS streamflow and runoff

Robock et al. [2003]: validation of NLDAS soil moisture/temperature and energy fluxes

Schaake et al. [2003]: evaluation of NLDAS soil moisture states and storage changes

Sheffield et al. [2003]: validation of NLDAS simulated snow cover

Pan et al. [2003]: validation of NLDAS simulated snowpack content

Cosgrove et al. [2003b]: evaluation and testing of NLDAS spin-up

Pinker et al. [2003]: production and validation of GOES surface insolation for NLDAS

Shi et al. [2003]: production of the daily precipitation analyses used by NLDAS

To highlight the "NLDAS companion" nature of the above papers when cited herein, the year of their citation is augmented with the label "N". The present and above papers show that a pillar of the NLDAS initiative is the integrated application of a multitude of GCIP-supported products developed by GCIP-supported investigators, as highlighted in Table 1.

Section 2 describes the NLDAS configuration, surface forcing, land models, and streamflow routing. Sections 3 and 4, respectively, present the evaluation of the NLDAS surface water budget and surface energy budget. Section 5 presents conclusions and future plans.

2. NLDAS Configuration

2.1 General Configuration

The initial pilot systems of realtime ocean 4DDA began on a relatively data rich, sub-global domain – namely the large-scale tropical Pacific Ocean, which included the TOGA observing network [*Ji et al.*, 1994]. In choosing the domain for our pilot LDAS demonstration, we too chose a relatively data rich, large scale but sub-global domain. In so doing, we heeded a lesson from GSWP results [*Oki et al*., 1999], namely that land surface simulation over large continental domains is notably less viable over regions lacking moderately dense precipitation gages to anchor the precipitation forcing. Outside such regions, the precipitation analyses in GSWP were dominated by satellite-inferred precipitation, which in the extratropics may only be marginally better then model-based precipitation, or even worse in some cases (e.g., cool season stratiform rainfall, winter snowfall). Hence in the pilot NLDAS here, we limit the domain (shown in Figure 1) essentially to the data-rich CONUS, thereby benefiting not only from relatively dense precipitation gages, but also from the GCIP-supported CONUS-oriented products in Table 1. On the domain in Figure 1, we employ a 1/8° latitude/longitude grid for all the NLDAS LSMs.

In addition to a common domain and grid, the NLDAS applies the following additional shared attributes in executing the four land models: common hourly input surface forcing, land mask and terrain elevation, database of soil texture class and vegetation class, streamflow network and basin boundaries, streamflow routing model, and content, frequency (hourly) and format (GRIB) of model output. The common NLDAS elevation field was computed by averaging, within each 1/8° grid cell, the 30 arc-second (~1km) digital elevation of the global, GTOPO30 database of *Verdin and Greenlee* [1996]. Of the four land models, the VIC model alone also employs sub-grid elevation tiles or bands (see Section 3.4).

The common vegetation classification was derived from the global, 1-km, AVHRR-based, vegetation database of U. Maryland [*Hansen et al*., 2000], which carries 13 vegetation classes. For each 1/8° grid cell, the derived NLDAS vegetation field includes the percent of each class based on its 1-km frequency within the cell. The Mosaic and VIC LSMs employ sub-grid vegetation tiles whose weights within a cell correspond to the percent frequency of occurrence of the 1-km vegetation classes. The Noah model avoids tiling and utilizes the single most predominant vegetation class. The SAC model omits an explicit treatment of vegetation.

The soil texture database over the CONUS was derived from the 1-km STATSGO database of *Miller and White* [1997], which carries 16 texture classes by layer over 11 layers of variable thickness (5-50 cm). For each 1/8° grid cell, the derived NLDAS soil database carries the percent of each class, by layer, based on the corresponding 1-km frequency in the cell. To date, the physical soil parameters utilized in the NLDAS models assume a vertically uniform soil class based on the predominant soil texture of the top 5-cm layer. (The sole exception is the Mosaic porosity, derived for each Mosaic soil layer based on weighted averages from the 11-layer soil textures.) Though the SAC model uses conceptual soil-water storage reservoirs, rather than explicit soil parameters of an explicit soil column like the other LSMs, some a priori parameters for SAC in NLDAS were derived as a function of the same predominant, top-layer soil texture (see Section 2.3). Outside the CONUS, the NLDAS soil database employs the same soil texture classification, but carries only a single, vertically uniform, predominant class derived from the 5-minute ARS FAO global data of *Reynolds et al*. [2000]. Spatial maps depicting the NLDAS vegetation and soils databases, as well as tables defining NLDAS vegetation and soil classes, may be viewed under the NLDAS link maintained by NASA at http://ldas.gsfc.nasa.gov.

Although model control runs in NLDAS employ common fields of vegetation and soil class, we chose NOT to impose any additional commonality, in such areas as 1) physical parameter values, 2) configuration of given vegetation class (e.g. root depth and root density), 3) the seasonal cycle of vegetation, and 4) soil column configuration (e.g. number and thickness of soil layers). The rationale was our wish not to negate the considerable legacy of calibration or tuning invested to various degrees over the past decade in the four land models when executed over large basins or continental scales (e.g. CONUS-wide). No additional calibration or tuning of the LSMs was carried out for the NLDAS control runs. In addition to control runs, the NLDAS project conducts sensitivity tests in which common parameters or treatments are imposed. For example, *Robock et al*. [2003]-N evaluates tests of the use of common physical soil parameters in the NLDAS LSMs, some results of which are discussed in Section 3.3.

The seasonality of vegetation is a notable aspect for which we do not impose a common treatment, as a model's phenology treatment resides at the core of a given model's canopy resistance formulation and hence that model's transpiration response. In broad terms, the Noah, VIC, and Mosaic LSMs apply a satellite-derived, AVHRR-based, monthly seasonal cycle of vegetation. SAC does not apply seasonal phenology, as it lacks explicit vegetation. During execution, Noah, Mosaic and VIC temporally interpolate their respective monthly phenology values to the Julian day of the valid simulation time. Yet, significant nuances exist between these LSMs regarding whether the monthly values are for the given year (as in Mosaic) or from a multi-year climatology (as in Noah and VIC) and whether the seasonality is carried in the LAI (as in VIC), or in the vegetation fraction (as in Noah), or in both (as in Mosaic). A summary for the control runs herein follows.

Noah runs use the global, 0.144° (~15-km), monthly 5-year climatology of the fraction of green vegetation cover (GVF) derived by *Gutman and Ignatov* [1998] of NESDIS from AVHRR-based NDVI. The latter GVF climatology over the NLDAS domain for July is presented in Figure 1a, as a general reference for later discussions of CONUS vegetation cover. For Mosaic runs, NASA obtained monthly 16-km AVHRR-based green LAI fields derived by Boston University for the actual given month (not climatology). NASA then first derived dead LAI (estimated from the difference in green LAI between consecutive months, along with vegetation class-dependent values of minimum dead LAI) and then derived monthly total LAI (sum of green and dead LAI), from which GVF was computed (as green LAI divided by total LAI). VIC runs use a global, AVHRR-based, multi-year monthly climatology of total LAI, which is used in conjunction with a vegetation class-dependent look-up table of fixed annual-maximum vegetation fraction (dead and green).

The NLDAS provides hourly output of a common set of required model fields (about 50). The required set, viewable at http://ldas.gsfc.nasa.gov/LDASnew/LDAS8th/LDAS_output.shtml, closely follows that urged by the GLASS initiative of GEWEX. This set includes all terms of the surface energy and water budgets, all soil and snowpack moisture and temperature states, and various ancillary fields. Looking ahead to infusing NLDAS in NCEP operations, the chosen format (including units, sign convention) of all NLDAS gridded fields (either input or output) is GRIB – the WMO standard used at major NWP centers worldwide. GRIB protocols yield efficient data compression in the output volume by masking out inactive grid points (governed by the land mask file) and retaining only necessary precision (user-specified field by field). For faster execution or reductions of output volume in NLDAS experiments, the NLDAS may be run

in a "limited-domain" mode, easily invoked in model runs via a modified land mask, reduce-able to a handful of points (e.g., nearest neighbors at flux stations), or in the limit, even a single point.

The NLDAS required "cold start" initial values of all LSM state variables for the common start time of 00 UTC on October 1, 1996, after which all state variables were cycled by a given LSM's physics. For all the LSMs, the initial value of sub-surface temperature and moisture states were derived from the soil column states of the NCEP/DOE Global Reanalysis 2 [*Kanamitsu et al.*, 2002] valid at the start time. The moisture state of the latter was provided to each LSM in terms of a vertically uniform percent of saturation. The initial snowpack storage was set to zero, which is reasonable for October 1 at the NLDAS resolution over the NLDAS domain. Canopy interception stores were set to zero. The spin-up behavior from this cold start is examined in detail by *Cosgrove et al*. [2003b]-N, who concluded that the practical drift in land stores and output ceased in the control runs within approximately 1 year. The spin-up from the above Global Reanalysis states was shorter than using idealized wet or dry initial states.

2.2 Surface Forcing

The cornerstone making NLDAS possible is the generation of NLDAS surface forcing. The Noah, Mosaic and VIC LSMs need 7 input forcing fields: wind speed, air temperature, air specific humidity, surface pressure, downward longwave radiation, downward shortwave radiation, and precipitation (the primary forcing in Table 2). The SAC model requires precipitation (*P*), max/min air temperature, and potential evaporation (*PE*). *Cosgrove et al*. [2003a]-N describe how the hourly forcing fields are produced. Here just a short overview and history are given. NCEP originally identified the data streams, assembled the algorithms, and developed the software to produce the realtime surface forcing fields. In mid-April 1999, NCEP commenced and has maintained daily realtime generation and archiving of this forcing.

Additionally, since 1995-1996, GCIP has sponsored the archiving of the NOAA data streams that comprise inputs to NLDAS forcing. NASA GSFC undertook the task of leveraging these archives (at NCAR, with GOES-products at U. Maryland) by adapting NCEP's surface forcing software to produce retrospective surface forcing back to October 1996.

Table 2 summarizes the content and input data sources of both the realtime and retrospective surface forcing data sets, which are both hourly and 1/8° resolution, produced in the same format (GRIB) and featuring the same content of primary and secondary forcing variables. Both sets are produced using the same spatial and temporal interpolation methods and are quality controlled in similar fashions. The baseline source of the surface forcing fields is the set of 3-hourly atmospheric analyses produced by NCEP's Eta model-based Data Assimilation System or EDAS [*Rogers et al*., 1995]. The EDAS includes the assimilation of screen-level atmospheric temperature, dew point, wind velocity, and surface pressure. EDAS analysis fields of the latter, plus EDAS precipitation and surface solar and longwave radiation are provided on a 40-km output grid to the NLDAS forcing software, which interpolates the latter in time and space to hourly 1/8° surface forcing fields. At this point both the retrospective and realtime sets experience the same terrain adjustments [see *Cosgrove et al*., 2003a]-N, namely to the surface pressure, air temperature and humidity, and downward longwave radiation to account for differences in the terrain heights of the NLDAS grid and the Eta/EDAS 40-km output grid that supplies the NLDAS forcing suite.

The aforementioned EDAS precipitation and solar insolation serve only as backups. Normally, these two major components of the forcing are provided by model-independent observations, namely NCEP's 1/4° daily gage-only precipitation analyses of *Shi et al.*, [2003]-N and the 1/2° GOES-based satellite retrieval of solar insolation, produced via the cooperative

venture of NESDIS and U.Maryland, [*Pinker et al.*, 2003]-N. Additionally, the daily precipitation analysis is spatially interpolated to 1/8° and then temporally disaggregated into hourly fields for by deriving hourly disaggregation weights from the hourly, WSR-88D, radar-dominated, 4-km precipitation estimates known as "Stage II/III". These radar-dominated fields do not affect the daily precipitation total in NLDAS, as they are used only to derive the temporal weights for partitioning the gage-based daily total into hourly amounts. Nevertheless, for research purposes, we spatially average the radar-dominated precipitation estimates to the 1/8° NLDAS grid and save them in the NLDAS hourly forcing files.

Figure 2a shows the gage-based retrospective daily precipitation analysis of *Shi et al*. [2003]-N for July 23rd 1998 and the corresponding EDAS 24-hour accumulated precipitation in Figure 2b. Though *Roads et al*. [2003] found that EDAS precipitation has some of the smallest errors on a monthly total basis relative to other assimilation systems like the Global Reanalysis, it can be seen in Figure 2 that on daily time scales, the EDAS precipitation can have substantial errors for individual synoptic events compared to the gage-based analysis. Hence, the gage-based analysis represents a substantial improvement over EDAS in precipitation forcing for NLDAS. *Luo et al* [2003]-N showed that in non-mountainous areas, the NLDAS precipitation verifies well with independent gage observations at time scales of daily or longer, though marginally at the short hourly time scales. Later, Section 3.4 addresses the accuracy of NLDAS precipitation at high mountain elevations in the western U.S.

Figure 3 shows the monthly mean diurnal cycle of the GOES-based and EDAS-based surface solar insolation compared to that obtained by the SURFRAD flux station at Bondville, IL for July (left) and January (right) 1998. Clearly the GOES insolation has far less bias than the EDAS insolation. As concluded from inter-comparisons by *Pinker et al*. [2003]-N and *Luo et al*.

[2003]-N, the satellite insolation estimates were found to verify well against the surface flux stations, though with some deterioration toward positive bias at low sun angles and in the case of snow cover (when cloud detection is difficult). Figure 3b shows that despite some winter deterioration in the bias of the GOES insolation (compared to summer in Figure 3a), the latter is still notably less biased than EDAS. We thus expect the NLDAS simulations to be less affected by radiation biases when GOES-based solar insolation is used.

The basic difference between retrospective and realtime processing is that the retrospective data set is produced over the course of one computer run well after realtime and after the collection in holding directories of the required inputs. The realtime production effort depends upon an extensive and interrelated series of dynamic realtime data archives, C Shell scripts and Fortran programs to produce new forcing data once per day. While the realtime data set relies upon only those data sources that are available in the production time window (e.g. about 6500 gage reports of total precipitation for the 24-hours ending 12 UTC arriving daily in realtime by 19 UTC), the retrospective data set can make use of reprocessed and enhanced sources of data (e.g. about 13000 gage reports of 24-hour precipitation totals, including the above 6500 realtime reports and roughly 6500 non-realtime Cooperative Observer reports received from NCDC after realtime). In addition, the production of near realtime data can be impacted by computing and network conditions, as well as by the day-to-day availability of critical data sets. Errors can surface in the data sets or code used to construct the NLDAS forcing data, and while realtime production, in a process of continual improvement, can quickly adapt to ensure that future forcing files are not affected, any files already produced in the past with the erroneous data set remain tainted. By contrast, when such problems emerge in the

production of the retrospective data set, the forcing production can restarted from the beginning using corrected code or data.

Luo et al. [2003]-N compared the NLDAS retrospective surface forcing with observations from the OU Mesonet and ARM/CART stations during January 1998 to September 1999. They found good agreement between the two data sets for all meteorological variables, though with some caveats for precipitation. Therein, the agreement was only adequate at the hourly time scale, but became rather good at increasing time scales of daily, 5-day, and monthly. They also found a modest low bias in EDAS downward longwave radiation, which partially offsets the small high bias in the GOES-based solar insolation.

Luo et al. [2003]-N also used all four LSMs in a number of limited-domain experimental runs of NLDAS in which locally observed forcing replaced the control NLDAS forcing, to investigate the influence of NLDAS forcing on the resulting energy and water fluxes of the LSMs. When comparing model results using the local and control forcing, they found that the models show relatively small differences in their land surface states and surface fluxes on daily to monthly scales. Although the swapping of surface forcing was significant at certain specific times at individual stations, the overall differences were not significant at longer time scales and the differences were much smaller than the differences among the models and between the models and observations.

The NLDAS hourly forcing files, for convenience and future use, contain a number of additional fields of interest to the land modeling community. These are CAPE and convective precipitation from EDAS (for land models that may use these to infer subgrid distributions of precipitation) and PAR and LST from the GOES-based product suite of NESDIS/U.Maryland.

2.3 The Land Models

The LSMs currently implemented in the NLDAS are Noah, VIC, Mosaic, and SAC. Table 3 lists the main attributes of these four LSMs. From the many available LSM candidates, these four give a solid cross-section of different legacies, including small scale versus large scale, coupled versus uncoupled, physical versus conceptual, distributed versus lumped, with and without explicit vegetation, tiled and non-tiled, and extensively versus minimally calibrated. Mosaic and Noah emerged from the surface-vegetation-atmosphere transfer (SVAT) scheme settings of coupled atmospheric modeling with little calibration. VIC and SAC grew from the hydrological community as uncoupled hydrology models with considerable calibration. Mosaic was developed for coupled use in the NASA global climate model [*Koster and Suarez*, 1994, and 1996; *Koster et al*., 2000], while Noah was developed for coupled use in the NCEP mesoscale NWP Eta model [*Chen et al*. 1997; Betts et al., 1997; *Ek et al.*, 2003, this issue], but has also been used for large scale studies of the hydrological cycle [*Chen and Mitchell*, 1999]. VIC was developed as a macro-scale quasi-distributed uncoupled hydrological model [*Liang et al*., 1994; *Wood et al.*, 1997], while SAC was developed as a lumped conceptual hydrology model [*Burnash et al.*, 1973], highly calibrated for small catchments (50-5000 km^2), and executed operationally in NWS RFCs.

Subsequent to their original heritage, Mosaic, Noah, and VIC have come to be executed extensively in both coupled and uncoupled mode from small to large scales. Thus at present, all 3 models can be considered as both SVATs and semi-distributed hydrological models. All three have undergone testing on local and regional scales in various phases of the PILPS project [*Chen et al*., 1997; Woo*d et al.,* 1998; *Schlosser et al*., 2000; *Bowling et al*., 2003] and on the global

scale in GSWP [*Dirmeyer et al*., 1999]. Seeking similar SAC suitability over a broad range of spatial scales, OHD of NWS recently developed a macro-scale, distributed (non-lumped) version of SAC [*Koren et al*., 2000] with a priori un-calibrated parameters. This SAC version is intended for testing on small to large river basins, or entire continents. SAC executions over the CONUS here in NLDAS are the first continental-scale tests of the distributed version of SAC. We emphasize that these CONUS-wide SAC executions must be viewed as pilot tests, as they lack the legacy of continental-scale testing enjoyed by Noah, Mosaic, and VIC.

The three SVAT models simulate LST, the surface energy and water fluxes of the surface energy and water balance, snowpack water content, and soil moisture in several soil layers, though the number and thickness of the layers differ. Only Noah simulates frozen soil processes and soil ice [following *Koren et al*., 1999]. The snowpack treatments in all four models are described later in Section 3.4. In all three SVATs, the surface infiltration schemes account for sub-grid variability in soil moisture and precipitation, but the approaches differ, as do the drainage approaches. All three SVATs include direct evaporation from the soil, transpiration from the vegetation canopy, evaporation of canopy interception, and snow sublimation. In their transpiration, the three SVATs include canopy resistance responding to soil moisture, air humidity and temperature and solar insolation, though the formulations (see Table 3 for references) and parameters differ across the models, as does the density and greenness of the vegetation cover, and the depth and density profile of the root zone. Finally, the SVAT models treat the aerodynamic conductance with their own schemes, which is a focus of Section 4.2.

In contrast to SVAT models, SAC is a conceptual rainfall-runoff, storage-type model [*Burnash et al*., 1973]. It omits the surface energy budget, treating only the surface water budget. SAC runs with the temperature-index based snowpack model of *Anderson* [1973],

known as SNOW17. Together they represent the operational NWS River Forecast System in the NWS RFCs. Henceforth, the term "SAC" means the SAC-SNOW17 pair. SAC outputs are evapotranspiration (*E*) and runoff. The output *E* is calculated as a fraction of the input *PE* (Section 2.2 discussed the *PE* input for SAC). SAC has a two-"layer" (two reservoir) storage structure, a relatively shallow upper reservoir and a deeper lower reservoir. The fraction applied by SAC to obtain *E* from *PE* is a function of the soil water deficits (from capacity) in these upper and lower storage reservoirs. Each of these reservoirs consists of tension and free water storages. The free water storage of the lower layer is further divided into two sub-storages that control supplemental (fast) and primary (slow) groundwater flows. SAC is controlled by numerous model parameters (16 primary, and 12 more for SNOW-17). Traditionally, in operational practice in the RFCs, these parameters are carefully calibrated for individual catchments. For the NLDAS runs of SAC, all calibration was omitted. Rather, the 12 SNOW-17 parameters were prescribed with a uniform set of parameters for the entire CONUS domain. (Nevertheless, the SAC snowpack simulations in Section 3.4 performed well.) The 16 primary parameters were specified a priori in NLDAS runs to vary spatially across the domain according to *Koren et al*. [2000], who related these parameters to the 1-km STATSGO soils database of Section 2.1 and National Soil Conservation Service "Curve Numbers". The SAC executions here in distributed mode over the entire CONUS leveraged the recent pilot executions of SAC in distributed mode over much smaller domains in central CONUS in the NWS/HL Distributed Model Intercomparison Project (DMIP).

The Mosaic LSM, developed by *Koster and Suarez* [1994, 1996], accounts for sub-grid heterogeneity of vegetation and soil moisture with a “mosaic” or tile approach. Aside from its tiling, Mosaic was originally derived from the SiB model of Sellers et al. (1986), and still

maintains certain SiB formulations, particularly in the fairly complex treatment of canopy conductance. Up to 13 sub-grid vegetation tiles are carried to represent up to the 13 vegetation classes as described in Section 2.1. Each tile features its own energy and water balance calculations as well as its own state variables for soil temperature and soil moisture. Grid-cell mean values of variables are computed as a weighted average of the tiles. Each tile has three soil layers. The first and second layers comprise the root zone, while the third layer acts as a "recharge" reservoir for long-term moisture storage. In NLDAS, Mosaic is configured to support a maximum of 10 tiles per grid cell with a 5% cutoff that ignores vegetation classes covering less than 5% of the grid cell. Additionally in NLDAS, Mosaic has been configured with each tile having 1) the single predominant soil type of Section 2.1 throughout a given grid cell and 2) the three soil layers fixed at thicknesses of 10, 30, and 160 cm (hence a constant rooting depth of 40 cm and constant total column depth of 200 cm). This latter configuration of Mosaic layer thickness, root and total depth and soil type is a departure from the standard Mosaic soil configuration, for the purpose of easier comparison in NLDAS [see *Robock et al*., 2003]-N with the soil moisture observation levels of the OU Mesonet and the soil layers of the VIC and Noah models (e.g. their 10 cm top layer). Although never executed before with fixed layer thickness, Mosaic performed well in the PILPS 2a and 2c experiments when configured in the standard way [*Chen et al., 1997*; *Lohmann et al., 1998*, *Wood et al., 1998*]. As a reference, the standard Mosaic configuration varies the soil type and layer thickness tile by tile according to the vegetation type. (This standard configuration yields first, second, and third layer thickness-ranges of 1-2 cm, 1-150 cm, and 30-200 cm, respectively, total column depth ranges of 32-350 cm, and root depths of 2-49 cm for non-forest and 150 cm for forests.)

The Noah land surface model is targeted to be a computationally efficient model of intermediate complexity for use in operational weather and seasonal prediction models. Thus it is formulated without tiles and with a single-layer snowpack, a simple canopy conductance approach [*Jarvis*, 1976], and a linearized (non-iterative) solution to the surface energy balance. The Noah model has proven quite effective in reproducing observed energy and water budgets without the complexity of tiling. The Noah surface infiltration scheme follows that of *Schaake et al.* [1996] for its treatment of the subgrid variability of precipitation and soil moisture. Since its origins from the Oregon State University (OSU) land model [*Pan and Mahrt*, 1987], the Noah model has benefited from the quasi-regular advancements at NCEP that arise from its operational coupled assessment year-round in the NCEP Eta model over N. America, both by NCEP investigators and external GCIP collaborators [see *Ek et al., 2003*, this issue and references therein]. The Noah version here in NLDAS is that implemented operationally in the NCEP Eta/EDAS suite on June 19, 2002. Virtually this same version is being executed in the 25-year (1979-2004) Eta/EDAS-based Regional Reanalysis now underway at NCEP. The study by *Berbery et al*. [2003, this issue] examines the large-scale hydrology of the coupled Eta/Noah model over the Mississippi Basin for the period June 1995 to May 2002, giving a coupled counterpart to the assessment of uncoupled Noah hydrology here in NLDAS.

The variable infiltration capacity (VIC) model is a semi-distributed grid-based hydrological model developed at the University of Washington and Princeton University (*Liang et al*., 1994, 1996a,b; *Cherkauer et al*., 1999). As compared to other SVATS, VIC's distinguishing hydrologic features is its representation of subgrid variability in soil storage capacity as a spatial probability distribution, to which surface runoff is related (*Zhao et al.,* 1980), and its parameterization of base flow, which is represented in the lower soil moisture zone

as a nonlinear recession. Subgrid-scale variability in soil properties is represented in VIC by a spatially varying infiltration capacity, or infiltration curve. Thus, the spatial variability in soil properties, and topographic effects, at scales smaller than the grid scale are represented statistically, without assigning infiltration parameters to specific subgrid locations. In NLDAS, VIC executes with a one-hour time step and uses three soil layers with a constant top layer thickness of 10 cm, and spatially varying depths for the bottom two layers, partly determined from calibration over different regions. The root zone can span all three soil layers, depending on vegetation class. Like the Noah model, VIC evapotranspiration is calculated using a Penman-Monteith formulation with adjustments to canopy conductance following *Jarvis* [1976]. Like Mosaic, the VIC model carries sub-grid vegetation tiles. Additionally, VIC carries subgrid elevation bands, most notably for its two-layer snowpack model. The elevation band approach (described further in Section 3.4) has proved effective in VIC's reliable simulation of snowpack processes across a wide range of grid resolutions in several high latitude, high elevation model intercomparison studies [*Bowling et al.*, 2003; *Boone et al*. 2003]. The VIC model has been widely applied to large continental river basins, for example the Columbia [*Nijssen et al*., 1997] and the Arkansas-Red [*Abdulla et al*., 1996; *Wood et al*., 1997], as well as at continental scales [*Maurer et al.,* 2002*, Roads et al.,* 2003, this issue], and global scales [*Nijssen et al*., 2001].

2.4 Streamflow routing

The simulation and assessment of NLDAS streamflow on small to large river basins is presented by *Lohmann et al.* [2003]-N. The latter study utilizes measured streamflow for 9 major and 1145 small to medium-sized basins (ranging from 23 km^2 to 10,000 km^2) across the CONUS obtained from the USGS and Army Corp of Engineers. Out of 8053 basins having data during October 1996 - September 1999, 1145 basins were chosen for assessment based on no

missing data, upper bound on basin size (10,000 km^2) and no obvious signatures of reservoir operation (the majority of chosen basins thus being in CONUS-East). The 1145 basins represent about 25% of the CONUS area (over 50% of CONUS-East, see Figure 8). From this set, a smaller subset of 188 basins has been selected for especially close scrutiny of the simulated streamflow. Basins in this subset have area A no smaller than 2000 km^2, have realtime streamflow observations available from the NWIS web site of the USGS, have an adequate number N of precipitation gages (N-1 > 0.6A**0.3), and collectively represent a variety of soil and vegetation types and climate conditions (e.g. P/PE).

The streamflow routing requires both a streamflow-direction mask on the NLDAS grid and a routing model. The source and derivation of these is presented in *Lohmann et al.* [2003]-N. The linear routing model is identical to that in PILPS 2c [*Lohmann et al.*, 1998] and 2e [*Bowling et al.,* 2003; *Nijssen et al.*, 2003]. It can be coupled to a land surface scheme or (as in NLDAS) applied as a "post-processing" step that ingests the gridded runoff of any land model on the NLDAS grid. The routing model was executed in two modes: i) distributed, using a-priori, default, grid-cell specific routing model parameters common to all four models and ii) lumped, in which constant ("lumped") routing parameters for a basin were separately calibrated for each basin and model.

Figure 4 shows a small-basin example of the two-year time series (initial spin-up year not shown) of observed versus simulated daily streamflow of the four NLDAS models for the East Fork of the White River in Indiana. The inset depicts the derived unit hydrograph for this basin for each model from the calibrated lumped mode. The unit hydrograph manifests the differences in runoff generation and timing among the models. It is a measure of the distribution of the residence time of stream-channel water in the catchment after being produced as runoff by the

model. For the basin in Figure 4, Mosaic produces a faster runoff response than the other models, and hence its unit hydrograph needs to keep this runoff longer in the stream-channel of the catchment. The table at the top of Figure 4 shows the bias and correlation of the simulated streamflow of each model with respect to the observed streamflow. In this basin, all the models show relatively high correlation with observed streamflow (with SAC yielding the highest), but with significant bias – namely, large negative bias in Mosaic and SAC, modest positive bias in Noah, and large positive bias in VIC. Section 3 shows this pattern of streamflow bias is typical of CONUS-East and is reflected by annual evaporation signatures in the opposite sense.

3. Assessment of the NLDAS water budget

The hallmark of the assessments in Sections 3 and 4 is their breadth, as they address spatial scales from continental to local, all four seasons and multiple types of validating observations (Table 1, Section C). The assessment focuses mainly on the last two years of the three-year simulation, as the first year is a necessary and (mostly) sufficient spin-up year.

3.1 Annual water budget: partitioning between evaporation and runoff

Over one or more annual periods, the total storage change in the surface water budget is usually negligible compared to the other terms; and hence to good approximation, the total annual precipitation equals the sum of annual evaporation and runoff. We thus examine the mean annual partitioning of input precipitation between evaporation and runoff. Figures 5 and 6, from *Lohmann et al.* [2003]-N, show the mean annual evaporation and runoff of the four LSMs across the CONUS over the last two years of simulation.

Since the models received the same precipitation, the disparity in Figures 5-6 between the models is striking, especially over CONUS-east, which is well vegetated (Figure 1a). The disparity is notably less but significant over the arid CONUS-west. Over CONUS-east, Noah

and VIC have notably lower evaporation and hence higher runoff than the Mosaic and SAC models, which have high evaporation and low runoff. Such large disparity among land models in the runoff-evaporation partitioning was noted in the PILPS and GSWP studies [*Wood et al.*, 1998; *Dirmeyer et al.*, 1999]. Like these studies, we depict this partitioning in the manner of Figure 7a, wherein each diagonal represents the mean-annual area-averaged precipitation (given by the diagonal's x- or y-axis intercept value) of a given CONUS quadrant. Projecting a model symbol from a given diagonal onto the x-axis (y-axis) yields the regionally averaged annual mean runoff (evaporation) for that model. Since each model conserves water, the tiny displacement in Figure 7a of any model symbol from its respective diagonal represents the negligible change in that model's total water storage over the two years.

In the two eastern quadrants in Figure 7a, VIC has the least evaporation and the most runoff, while Mosaic and SAC pair off closely with high evaporation and low runoff, with Noah falling rather evenly between VIC and the Mosaic/SAC pair. The close similarity of Mosaic and SAC is totally unexpected, as Mosaic (like Noah and VIC) includes a detailed treatment of the vegetation canopy its conductance, while SAC includes no explicit treatment of vegetation.

The evaporation-runoff partitioning in Figure 7a was validated by *Lohmann et al.* [2003]-N, and summarized here, using the observed streamflow from the 1145 evaluation basins cited in Section 2.4. For each such basin, observed streamflow (m^3s^{-1}) was converted to total mean-annual discharge (m^3), which was then converted to the area-average mean-annual runoff (mm) of the upstream basin area (Figure 7c). Figure 7b is the 7a counterpart obtained by spatially averaging (in each quadrant) over only the 1145 basins of observed runoff in Figure 7c. The vertical lines in Figure 7b denote the observed runoff and their intersection (bold asterisks) with the diagonals provide the budget-based estimate of the area-average, mean annual evaporation.

The reliability of this estimate of actual evaporation depends on the reliability of both the observed streamflow (high reliability) and the input precipitation forcing.

Over the eastern quadrants, we trust the precipitation analysis (see Section 2.2) and thus the budget-based estimates of evaporation there. In both eastern quadrants, the evaporation and runoff of the Noah model are rather close to the observed, while Mosaic and SAC manifest a large high bias in evaporation and large low bias in runoff, with VIC manifesting the opposite in large bias. Figure 8 from *Lohmann et al.* [2003]-N depicts the spatial variability of the relative bias [(model-observed)/observed] of modeled mean annual runoff over the 1145 evaluation basins. Over CONUS-east, the Mosaic and SAC models have similar patterns of widespread underestimation of runoff, often exceeding 60 % (dark red) in magnitude around the Appalachians and upper Midwest. The VIC model typically overestimates runoff (except for one corridor of near neutral bias west of the Appalachian), often by more than 60 % in the southeast and mid-west (dark green). The Noah model generally has smaller runoff bias, with a more balanced likelihood of positive, negative, or near-neutral bias and fewer basins of large bias.

In the western quadrants, NLDAS precipitation is not reliable, owing to the mountainous terrain, the low density of precipitation gages, and the omission of a PRISM [*Daly et al*., 1994] adjustment (Section 2.2.). Section 3.3 later shows a 50 % low bias in NLDAS precipitation at SNOTEL sites over high mountain elevations of the West. In the northwest quadrant, the low precipitation bias causes a substantially low runoff bias in all four models in Figure 7b. In Figure 8 over the northwest, all four models under-predict the runoff in the northern Rocky Mountains and Cascade Mountains by 20 to 100%, though VIC yields notably smaller low runoff bias in this region, because VIC has less evaporation over the area (Figure 5, 7b).

Lohmann et al. [2003]-N provide additional NLDAS runoff and streamflow results over the small to medium-sized basins of Figure 8, including runoff bias in cold season versus warm season, bias in "time-to-peak" streamflow response, and Nash-Sutcliffe efficiency of the simulated streamflow, as well as river discharge results and validation for nine large river basins. In the latter large-basin results, one feature of the modeled discharge (not shown) in the Columbia and Colorado basins (noted for their winter snowpack over mountains) is the early bias in the spring discharge peak in the Noah model. Later, Section 3.4 shows this arises from an early bias in the timing of springtime snowmelt.

Sections 3.2-3.4 examine the surface water budget on monthly and shorter time scales and address some of the physical causes for the systematic model biases identified above.

3.2 The monthly water budget and soil moisture change

In studying non-annual water budgets, storage changes of soil moisture and snowpack are important. On monthly scales, we apply the full surface water budget equation given by

$$\frac{dS_1}{dt} + \frac{dS_2}{dt} + \frac{dSn}{dt} = P - E - R_1 - R_2 \,. \qquad (1)$$

Each term in (1) is the spatial average (e.g. grid cell, region, basin) of storage change (left-side) and accumulations (right-side) of water mass per unit area (kg/m^2) per month – equivalent to depth of water (mm) per month. S_1 and S_2, respectively, are soil moisture stores of an upper and lower zone (defined later), S_n is snowpack storage, P is monthly precipitation, E is monthly evaporation, R_1 is monthly surface runoff and R_2 is monthly subsurface runoff. On the monthly scale here, canopy interception storage is negligible and omitted in (1). Figure 9 gives the time series of area-average monthly evaporation E for each CONUS quadrant for the full 3-year

execution period, starting October 1, 1996. There is no obvious sign of spin-up during the first-year or thereafter (except possibly in SAC during first spring in SE). The seasonal changes in the evaporation look reasonable in all four models, as do the regional differences, with the SE quadrant yielding the most evaporation and the SW quadrant the least, as expected.

We focus here on the eastern quadrants, where model evaporation in Section 3.1 showed the most disagreement. Though Mosaic and SAC both showed high annual evaporation over CONUS-east in Figure 5, Figures 10b and 10d show that Mosaic and SAC have significant seasonal differences. SAC has the highest CONUS-east evaporation in winter and spring, while Mosaic has the highest in summer and fall (see also Figure 13 later). During mid and late summer, Mosaic has distinctly higher evaporation than all three other models. Noah evaporation generally falls between that of Mosaic and VIC in the warm season, while VIC has the lowest evaporation in virtually every month in the eastern quadrants, consistent with the mean annual results in Figures 7a-b. Figures 10b and 10d strongly suggest that Mosaic and VIC manifest distinctly different transpiration stress (canopy conductance). This is most evident in the SE during 1998, where VIC monthly evaporation tops off at 80 mm during May-August while that of Mosaic rises to and sustains 120-150 mm.

Such large differences in warm season evaporation imply large differences in soil moisture storage change. Figure 10 shows the time series of area-average monthly-mean 1) total soil moisture storage (S_1+S_2) for NW, NE, and SE (Figures 1a-c), and 2) root-zone soil moisture for SE (Figure 1d). The focus below is on years 2-3 of Figure 10, as significant spin-up of the total soil moisture is evident in Figure 10 in all four models during year one, though more so in Noah and VIC. After year one, spin-up is essentially complete for practical purposes, though some small spin-up may still be occurring in the NW in Noah. Again, *Cosgrove et al.* [2003b]-N

examine the NLDAS spin-up character in detail, including over smaller regions. Focusing henceforth on years 2-3, an examination of Figure 10 (with Figure 9) reveals:

a) very different levels of annual mean total soil moisture across the models, ranging in the SE from about 325 mm for VIC and SAC to 550 mm for Mosaic and 650 mm for Noah,

b) more similarity, yet important differences, in annual-cycle amplitude (seasonal change) of total soil moisture among the models, ranging in the SE from about 100-150 mm for Noah, VIC and SAC to a notable high of 230 mm for Mosaic ,

c) larger differences among the models over the wetter eastern quadrants than the drier western quadrants, in both total soil moisture and its seasonal change or range,

d) among the three models with a root zone, the contribution of the sub-root zone to the change in total soil moisture varies widely, e.g. over the SE it is very large in Mosaic (about 180 of 230 mm), moderate in Noah (about 70 of 140 mm), and small in VIC (about 10 of 100 mm),

e) the model with the highest level of total soil moisture is not the model with the largest seasonal change in soil moisture, nor the model with the largest monthly/annual evaporation,

f) a model with high annual evaporation is not necessarily a model having a high annual range of soil moisture storage (as the counterpart of SAC illustrates).

Intriguingly in Figure 10, while VIC and Noah have very different levels of total soil moisture in the SE and NE, they have more similar magnitudes of soil moisture in the root zone and seasonal change of soil moisture in the root zone in the SE (Figure 10d) and NE (not shown). Thus the moisture source for the higher summer evaporation in Noah then VIC is Noah's sub-root zone.

A dramatic feature in Figure 10 is the contrast in Mosaic values between Figure 10c and 10d. A priori to Figure 10d, it is a paradox that Mosaic has the highest warm season evaporation rates and annual range in soil moisture storage over CONUS-east, as Mosaic executions

employed a shallow rooting depth of 0.4 m, compared to 1-2 m in Noah and 1.35 - 3 m in VIC. Comparison of Figures 10c and 10d shows that Mosaic's sub-root zone accounts for the bulk of Mosaic's annual storage range. Later, we show that Mosaic includes vigorous upward diffusion of soil water to its root zone from below during the warm season, when its root zone is typically drier than below. Mosaic's upward diffusion is not dominant in the arid west, because there both root zone and sub-root zone are typically dry simultaneously.

Figure 11 shows the May through September (nominal soil dry-down season) change in total soil moisture for 1999. Positive values denote soil drying. As an aside, the soil moistening (negative change: dark red) of Florida, south Texas, and Arizona/New Mexico is out of phase with the dry-down over the bulk of the CONUS, because their "wet season" is summer. (Also, the Atlantic coastal states show moistening from two coastal hurricanes in September 1999.)

The hallmark feature of Figure 11 is the vivid inter-model differences in the depletion of total soil moisture over the greater central Mississippi basin, which experienced significantly below normal precipitation during this period (not shown). Mosaic shows the largest depletion there and the largest evaporation (Figure 9d), followed by SAC and then Noah, with VIC showing the least depletion and lowest evaporation (Figure 9d). Together with Figures 9d and 10c-d, Figure 11 illustrates that over a vegetated region with a warm-season precipitation shortage, Mosaic taps substantially more from its deep (sub-root) soil moisture to sustain evaporation, though too much so, in the end yielding too much evaporation (Figure 7b).

Figure 11 shows large inter-model variation in the east-west gradient across the CONUS of the seasonal range in total soil moisture storage (S_1+S_2). Figures (not shown) from *Schaake et al.* [2003]-N depict this inter-model difference as a function of basin climatology, in terms of basin-mean *P/PE* ratio (wherein PE denotes potential evapotranspiration) for the 12 CONUS

RFCs. As one can infer from east-west inspection along latitude 37° N in Figure 11, *Schaake et al*. [2003]-N find that VIC's storage range is relatively invariant (showing the least sensitivity) with respect to east-west climatic gradient in *P/PE* between arid and moist regions, while SAC and Noah show more variability, and Mosaic the most variability. *Schaake et al.* [2003]-N emphasize that a given model's dynamic storage range in a given region is not reliably inferred from the model's maximum water holding capacity, but is rather the result of more complex interplay between a region's climatology and major facets of a model's physics. For example, the study by *Koster and Milly* [1997] shows that a model's dynamic range of total soil storage is highly controlled by the interaction of the model's runoff and evaporation formulations and the functional dependence of these formulations on the model's soil moisture.

Figure 12, from *Lohmann et al.* [2003]-N, shows the monthly time series of every term in (1) over the annual cycle. Here storage S_1 is defined as the top two soil layers in Mosaic, Noah, and VIC, and the top storage reservoir in SAC and S_2 represents all remaining soil storage. The difference between the solid black line (total precipitation) and dashed black line (liquid precipitation) is the snowfall; and the red triangle is $P\text{-}dS_n/dt$ [= $(dS_1/dt+dS_2/dt+E+R_1+R_2)$]. Via (1), the red triangle's departure below (above) the solid black line equates to the monthly increase (decrease) in snowpack depth. The red triangle coincides with *P* (solid black line) in the warm season (no snow), or in the cool season when monthly snowfall is balanced by the sum of monthly snowmelt and sublimation, yielding zero net change in monthly snowpack.

In analyzing Figure 12, we first consider the SE region (top row), which has negligible snow and rather small month-to-month changes in total precipitation (in percent terms). Thus the annual cycle of the SE water budget is driven mainly by potential evaporation. Of special interest are the model differences in runoff response. Mosaic's dominance in soil moisture

depletion in the warm season is mirrored by Mosaic having the largest soil moisture recharge during November-February, leaving notably less precipitation during cool months for Mosaic runoff. SAC also has less cool season runoff than Noah or VIC, not so from storage recharge as in Mosaic (except in January), but rather from SAC having the highest cool season evaporation of all four models. Noah and VIC have larger total runoff then Mosaic and SAC in virtually every month, with VIC having the notably largest runoff (mostly subsurface) throughout the fall, winter, and spring, as VIC requires less cool-season soil recharge to replenish its smaller summer depletion. VIC, Noah and Mosaic produce mostly sub-surface runoff, while SAC produces mostly surface runoff.

In the NE in Figure 12, the above tendencies from the SE continue to hold in a broad sense, but other signatures arise from less precipitation in the cool season, greater monthly variability of precipitation in the warm season, and non-negligible snowpack processes in the cool season. The lower precipitation in the cool season compared to the SE results in less cool season runoff in all four models, but in general VIC still produces the most monthly runoff, followed by Noah, then Mosaic and SAC. The lack of SAC runoff all year in the NE is noteworthy, as is the continued high SAC evaporation in spring (March and April).

Given that SAC in NLDAS takes its potential evapotranspiration (*PE*) from Noah, the March-April evaporation excess in SAC versus Noah in both the NE and NW is instructive, as the vegetation greenness and hence transpiration are low then in both quadrants in Noah. Noah, Mosaic and VIC draw soil moisture for direct evaporation only from their topmost and relatively thin soil layer (0.1 m). This top soil layer is likely unable to sustain the same direct evaporation rates as SAC, which can tap into both of its top two soil storages in response to the *PE* demand. We speculate that this structural difference in direct evaporation between SAC and the three

SVATS largely explains the dominance in cool season evaporation in SAC, contributing substantially in turn to the high SAC annual mean evaporation.

Additionally in Figure 12, the high precipitation in June in NE followed by a dramatic drop of precipitation in July yields another vivid example of Mosaic's ability to draw upon its deep soil moisture. In the face of the steep drop in July precipitation, Mosaic is the only model in NE to yield July evaporation larger than June evaporation, and it does so via the largest July soil moisture depletion. Moreover, in the following months of August and September, Mosaic continues to sustain higher evaporation rates than the other models, despite its larger July storage depletion. To further illuminate Mosaic's mechanisms here, its layer 3 storage change (i.e., dS_2/dt) was dissected in detail for July 1998 at a single grid point near 45°N latitude and 92°W longitude, where Mosaic had three vegetation tiles. For this month at this grid point, the drainage (R_2) out the bottom of layer 3 ranged from 4-10 mm across the tiles, while the upward vertical diffusion of water to layer 2 ranged from 116 to 124 mm, clearly demonstrating Mosaic's vigorous supply of sub-root water to the root zone above by diffusion when the root zone becomes depleted. Though Noah and VIC, like Mosaic, physically include vertical diffusion of soil water, the magnitude of this diffusion is much larger in Mosaic.

The contrast between the models in Figure 12 is clearly less in the western quadrants, wherein the warm-season water budget of the three SVAT models is quite similar (while SAC has less warm season evaporation and soil moisture depletion), likely owing to the sparse green vegetation in CONUS-west (Figure 1a), whereby canopy conductance and root-zone processes are not dominant. In the cool season of the western quadrants, VIC still tends to have the most runoff and is still dominated by sub-surface runoff. Interestingly, SAC has virtually no sub-surface runoff in the western quadrants. A notable cool season contrast between the models in

the NW (and NE) is the notably lower snowpack accumulation in Noah during November - February. Specifically, comparison of the January NW water budget components of Noah with the other models reveals inter-model agreement in monthly snowfall amount (difference between solid and dashed lines for total and liquid precipitation), but Noah – unlike the accumulating snowpack in VIC, Mosaic and SAC – is melting and sublimating in the month about as much snow as it receives in snowfall. Noah's low snowpack bias is examined further in Section 3.4.

We conclude this section with some discussion of the causes of the systematic biases noted in the models thus far. The low bias in Noah snowpack development compared to the other models is shown later (Section 3.4) to arise most likely from a low albedo bias over snow in Noah. The high bias in Mosaic evaporation over CONUS-east is most likely a result of the very vigorous upward diffusion of water from the sub-root zone to the root zone in Mosaic. The high bias in SAC evaporation over CONUS-east is most likely a consequence of SAC's use in NLDAS of 1) uncalibrated a-priori parameters (Section 2.3) and 2) *PE* from Noah. As a counterexample, in NWS operations at the RFCs, SAC inputs a NOAA monthly climatology of *PE*, derived from long-term measurements of NOAA evaporation pans. This climatological *PE* is then scaled during SAC runs by a monthly fractional coefficient (range 0-1). The latter coefficient is a key SAC calibration parameter allowed to be moderately larger or smaller than 1.0, as determined from extensive SAC calibration runs over each operational catchment. No such calibrated coefficient has been derived to date for use in SAC large-scale runs over the CONUS-wide domain, either with Noah *PE* or NOAA climatological *PE*, and hence a universal coefficient of 1.0 was used in NLDAS. Moreover, the Noah *PE* is known to be higher than NOAA *PE* climatology. These two factors contribute to high SAC evapotranspiration (*E*) values in NLDAS here, but not always the highest *E* in NLDAS, and typically less than the *E* of Mosaic

in the warm season over non-sparse vegetation. Hence the SAC results in NLDAS here are not isolated outliers and thus they represent fundamentally important "pathfinder" runs of SAC executed in a fully national distributed mode with uncalibrated a-prior parameters. Moreover, Section 3.4 will show that the SAC snowpack simulations in NLDAS performed rather well.

We lastly address the unexpectedly low VIC evaporation bias in the results here over CONUS-east. Two separate but related VIC modeling efforts have been conducted over the NLDAS grid and terrain heights; specifically, the 3-year retrospective runs executed here with 1-hour time steps, and the 50-year retrospective runs reported in *Maurer et al.* [2002], executed with 3-hour time steps and with different sources for the surface forcing. The 3-year VIC runs here use essentially the same parameters as the VIC runs of *Maurer et al.* [2002]. Yet two significant differences were hourly temporal disaggregation of the daily precipitation and subgrid spatial disaggregation within a grid box, both used in the VIC runs here but not in those of *Maurer et al.* Not having the advantage of hourly radar-anchored precipitation analyses, *Maurer et al.* used uniform distribution of the daily precipitation throughout the day and within each 3-hour forcing interval. *Maurer et al.* (Figure 2) analyzed the impact of this uniform distribution versus non-uniform disaggregation to 3-hour time steps and the results showed that the differences for the subregion analyzed (Lower Mississippi basin) were modest. Nonetheless, subsequent comparisons between the retrospective runs of *Maurer et al.* and the 3-year retrospective runs of VIC here show that the combined and interactive impact of the three factors of temporal disaggregation, spatial disaggregation, and 1-hour versus 3-hour time steps can be significantly larger than suggested by the Lower Mississippi tests of the temporal disaggregation alone. The differences (shown for a transect across the eastern and central U.S. at http://www.hydro.washington.edu/Lettenmaier/Models/VIC/VIChome.html) are most evident in

portions of the country with a high fraction of convective precipitation and full canopy cover (e.g., CONUS-east summer). More examination of the differences, and development of parameter transformations to account for both disaggregation factors and the time step differences will be addressed in a future paper.

3.3 Regional validation of soil moisture

This section evaluates NLDAS soil moisture, first over Illinois and then over Oklahoma. The Illinois State Water Survey operates 18 sites across the state that measure soil moisture nearly bimonthly at 11 levels down to a depth of 2 m [*Hollinger and Isard*, 1994]. *Schaake et al.* [2003]-N evaluated NLDAS soil moisture at 17 of the sites. Figure 13 shows the resulting scatter plot of the state-wide average of model versus measured total-column 2 m soil moisture at bimonthly intervals during October 1997 through September 1999. This period omits the prior year of model spin-up. Each plot depicts the best-fit linear line. For VIC results, two best-fit lines for northwest and southwest Illinois were required, because past VIC calibration has yielded rather different storage levels in these regions.

In Figure 13, a best-fit line having slope greater than one indicates a storage range greater than observed. The Mosaic model yields a storage range greater than the other models (as expected from Sections 3.1-3.2) and about 50 % greater than observed. The results from Noah and SAC show a good fit to the observed soil moisture. Hence Noah and SAC give both excellent storage range as well as storage magnitude itself. VIC also shows good storage range over both halves of the state, but VIC absolute storage values are lower than observed.

Over Oklahoma, as described by *Robock et al.* [2003]-N and references therein, in-situ observations of soil moisture have been recently installed and calibrated at 72 Oklahoma Mesonet stations by the Oklahoma Climatological Survey. Figure 14 from *Robock et al.* [2003]-

N shows a 21-month time series during 1998-1999 of daily-mean, model and observed, 0-40 cm total soil moisture averaged over all 72 stations. Figure 14a shows absolute value, while Figure 14b the shows departure from time series mean. *Robock et al.* [2003]-N discuss the nuances of how the modeled and observed values for Figure 14a are derived. As this derivation involves several assumptions, the storage departures from series mean in Figure 14b are viewed as more reliable.

Superimposed on the obvious seasonal cycle in Figure 14 are the shorter temporal variations that are driven by individual precipitation events. In Figure 14a, the Noah model has somewhat higher soil moisture than observed most of the time. VIC agrees more closely with the observations in 16a, but yields too little depletion during the summer drought of 1998. Mosaic again has an overly amplified depletion of warm season soil moisture. Mosaic in the summer of 1998 is much drier then observations and the other models – so much so that Mosaic has difficulty recovering and thus remains too dry throughout the remainder of the time series. Finally in Figure 14, the SAC soil moisture time series shows a marked over-amplification of the short-term response to individual precipitation events.

Additionally, *Robock et al.* [2003]-N present results (not shown) of important soil moisture sensitivity tests with the three SVAT models at all 72 of the soil measuring stations. In these tests, the soil type in the models was set to match the reported soil type at each station and the soil parameter values were unified across the three models for each soil type. These station-matching soil types and unified soil parameters improved a model's performance if the parameters were not incompatible with the model's own calibrated parameters. Since the Noah model is not calibrated, the matching of model and station soil type and the unification of soil parameters did generally improve Noah's soil moisture performance at the stations (for example

reducing the Noah moist bias evident in Figure 14a – not shown). However**,** in the more extensively calibrated VIC model, use of the station-matching soil types showed notably less positive impact. Hence, changing a subset of a model's parameters in isolation without considering the others may degrade a model's performance.

3.4 Validation of snow cover and snowpack content

The studies of *Pan et al.* [2003]-N and *Sheffield et al.* [2003]-N, respectively, perform a large-scale assessment of NLDAS snowpack water equivalent (SWE) and areal fraction of snow cover extent (SCE) over large portions of the CONUS. This section presents key results from these papers, but first we summarize how the four models simulate snowpack.

VIC, Noah and Mosaic model both the snowpack water budget (snowfall, snowmelt, sublimation) and the snowpack energy budget (net radiation, sensible, latent and subsurface heat flux, phase-change heat sources/sinks), though the treatments of such details as snow cover fraction, albedo and retention/refreezing of snowmelt differ substantially between the models. SAC simulates only the snowpack water budget, via the separately executed SNOW17 snowpack model of *Anderson* [1973], which includes snowfall and snowmelt but omits sublimation. SNOW17 determines snowmelt empirically via an index method based on maximum and minimum daily air temperature. Only VIC and Noah explicitly account for snowpack aging and retention of liquid water (snowmelt or rainfall) in the snowpack. Only Noah employs frozen soil physics and the associated soil heat sources and sinks and reduction of surface infiltration. Mosaic and VIC carry explicit sub-grid tiles of vegetation class, but only VIC carries explicit sub-grid tiles for elevation, known as "elevation banding." The latter is cited in several PILPS studies as a key factor in the good snowpack simulations of VIC [*Bowling et al.*, 2003; *Boone et*

al., 2003]. For each of their sub-grid tiles, Mosaic and VIC execute separate water and energy balance computations, and thus they carry distinct tile-space snowpack states for SCE and SWE. Details of model snowpack physics for modeling SWE (and the empirical diagnosis of SCE from SWE in each model) are provided by *Wigmosta et al.* [1994], *Koster and Suarez* [1996], *Koren et al.*, [1999] and *Anderson* [1973] for the VIC, Mosaic, Noah and SAC models, respectively.

There is no distinction between rainfall and snowfall in the NLDAS precipitation forcing, thus requiring a criterion to infer snowfall. Steps were taken to infer the same snowfall across all the models, but some differences were unavoidable owing to model configuration differences. Overriding such configuration differences would have violated the spirit of the intercomparison. Specifically, in all four models, the input precipitation at each model time step was assumed to be all rainfall for surface air temperature > 0°C and all snowfall for temperature ≤ 0°C. This common criterion did not guarantee identical snowfall across the models owing to 1) the different model time steps (15 minutes in Noah and Mosaic and one hour in VIC and SAC) and 2) elevation banding in VIC. Noah and Mosaic interpolate the input hourly air temperatures to their 15-minute model time steps, thus Noah and Mosaic allow both rainfall and snowfall inside one hour. VIC adjusts the input air temperature to the elevations of its sub-grid elevation bands, thus VIC allows both rainfall and snowfall inside one grid cell.

We next present key results from *Pan et al*. [2003]-N, wherein NLDAS SWE simulations are validated against NRCS SNOTEL stations west of 104°W. This region includes the Rockies, Cascades and Sierra Nevada mountains and encompasses a total of 560 SNOTEL stations, which measure SWE, precipitation and air temperature every 15 minutes. The majority of the stations are at elevations above 1000 m, with the mean elevation around 2500 m. (See *Pan et al.* [2003]-N for SNOTEL references.) Since the resolution of NLDAS and its terrain field is 1/8° (about

11-14 km), comparison of model SWE against point-wise SNOTEL measurements is problematic, given the large standard deviation in space of point-wise elevation and surface air temperature in mountainous regions. Therefore *Pan et al*. [2003]-N omit use of SNOTEL stations whose elevation differs from that of the nearest-neighbor NLDAS grid point by more than 50 m. This retains 110 SNOTEL stations, distributed regionally as follows: 3 in Sierra Nevada, 17 in Cascades, 29 in southern Rockies, 61 in northern Rockies.

Figure 15 shows the scatter plot of the 3-year mean-annual maximum SWE for the four model control simulations (and two VIC test simulations discussed later) versus the 110 SNOTEL sites, each identified with one of the aforementioned regions. All the models substantially underestimate maximum SWE over all regions, most prominently over the Sierra Nevada and the Cascade Mountains where some model versus SNOTEL differences approach 1000 mm, and less prominently eastwards into the Rocky Mountains. Noah has the largest low bias, and also the lowest correlation. Mosaic also shows a rather low correlation. The SAC and VIC models have notably smaller low bias and substantially higher correlation. It is noteworthy that the model with elevation tiling (VIC) yields the highest correlation, yet the simplest model (SAC) without elevation tiling or an energy balance treatment is closely competitive.

Given the low SWE bias in all four models, it is not surprising that *Pan et al.* [2003]-N find the NLDAS precipitation forcing to be consistently low when compared to the observed SNOTEL precipitation at all 110 stations (not shown). They find that the stations with the highest observed precipitation are the stations where the NLDAS precipitation forcing has the largest low bias and such stations are those in the Cascade Mountains (region of heavy maritime cool-season upslope precipitation, see Figure 1a), while the sites with the smallest low bias in NLDAS precipitation are typically near the eastern edge of the Northern Rockies (interior

continental, leeside precipitation). Moreover, *Pan et al*. [2003]-N determine the linear regression between the 110 station SNOTEL precipitation and the coincident NLDAS precipitation to be P_{SNOTEL} = 2.1693 P_{NLDAS}, with an R^2 value of 0.64. This shows a factor-of-two underestimation by NLDAS precipitation forcing on average over the SNOTEL sites. The low bias in NLDAS precipitation documented here with SNOTEL is consistent with the low NLDAS precipitation bias suspected from the low bias in annual streamflow in all four models in the NW in Section 3.1 (Figure 8).

Using the model with the highest correlation in the control runs of Figure 15 (VIC), *Pan et al.* [2003]-N executed two sensitivity tests with two different methods of bias-adjusted precipitation. In "Test 1," the VIC model was executed in the NLDAS reduced-grid mode (see Section 2.2) comprised of the grid cells of the 110 SNOTEL stations, using precipitation forcing scaled by a regional factor (for the cited four regions) based on a regional regression fit of the NLDAS and SNOTEL mean annual precipitation. In VIC Test 1 in Figure 15, the simulated SWE bias is dramatically less, with the scatter rather evenly balanced about the 1:1 line. However, the scatter remains substantial and the correlation is only marginally improved, which is not surprising, since the regional scaling does not eliminate site-specific bias. To address the latter, in Test 2 VIC was forced with site-specific adjusted precipitation wherein the NLDAS precipitation was scaled to match the observed annual total precipitation at each SNOTEL site. VIC Test 2 in Figure 15 yields not only a much smaller bias, as did Test 1, but also a substantially reduced scatter and increased correlation of R^2 = 0.82.

The reader is referred to *Pan et* al. [2003]-N for the evaluation of NLDAS air temperature forcing biases at the 110 SNOTEL sites. These biases proved to be rather small over the cool season (with exceptions at some stations), contributing much less to the model *SWE* biases than

the precipitation bias. The high bias in NLDAS solar insolation over snow documented by *Pinker et al.* [2003]-N is another cause of the low bias in simulated *SWE* in the three SVAT models in NLDAS, however the impact of this insolation bias has not yet been quantified by NLDAS investigators. Interestingly though, the SAC model, which omits an energy balance treatment for snowpack in favor of its temperature index method, has virtually the smallest low bias of the four model control runs in Figure 15.

Sheffield et al. [2003]-N, who validated the NLDAS simulations of *SCE* over the CONUS against the NESDIS operational daily snow cover product known as the Interactive Multisensor Snow (IMS) [*Ramsay*, 1998]. The IMS analysis is a spatially complete, N. Hemisphere field of snow cover and sea-ice extent on a 23-km polar stereographic grid. It is derived by NESDIS analysts, who use an interactive workstation to assess snow-cover related visible, infrared, and microwave satellite products, as described by *Ramsay* [1998], as well as in situ snow depth observations. IMS snow cover fields are viewable at http://www.ssd.noaa.gov /PS/SNOW. The IMS flags each land grid cell as either 0 (zero or little snow cover) or 1 (all or substantially snow covered). For comparison with the higher resolution NLDAS, the IMS snow cover was re-sampled to the NLDAS grid (~12km) using a nearest neighbor algorithm.

All four NLDAS models diagnose *SCE* (0-1 fraction) as an empirical function of the model simulated *SWE* state, but the model treatments differ substantially in terms of a) the *SWE* threshold required to reach *SCE* = 1 and b) the form of the function relating *SCE* to *SWE*. In Mosaic, at each subgrid tile, *SCE* is specified by $SCE = SWE / (SWE + SWE_{mid})$, wherein the threshold parameter SWE_{mid} is vegetation-class dependent (0.05 m for forest and tall shrubs, 0.002 m otherwise) and denotes value of *SWE* at which *SCE* = 0.5. Interestingly, *SCE* in Mosaic never fully reaches *SCE* = 1 (e.g., *SCE* = 0.9 for $SWE = 9 \times SWE_{mid}$). By contrast, at each

subgrid tile in VIC, *SCE* is a step function: *SCE* = 1 for any nonzero value of *SWE*, otherwise *SCE*=0. In both Mosaic and VIC, the final grid-cell mean *SCE* is the weighted average of the tile values. In VIC, nonzero values of *SCE* are by far dominated by *SCE*=1 values given VIC's tiny *SWE* threshold. Noah and SAC both use nonlinear (but different) functions (not shown) to specify *SCE* according to *SWE*. In SAC, the *SWE* threshold for *SCE* = 1 is 0.09 m while in Noah this threshold varies by vegetation class (0.08 m for forest classes, 0.040 m otherwise, except 0.025 m for bare soil). Inspection of the *SWE* thresholds for large *SCE* in the four models reveals that SAC and Noah require comparatively deep snowpack for high *SCE* values, while Mosaic requires substantially less, and VIC requires very little. Thus for a given nontrivial but non-deep *SWE* value, VIC will generally yield the highest snow cover, followed by Mosaic, then Noah, and finally SAC. In the case of the three SVATs for example, assuming a non-forest and non-bare vegetation type, a *SWE* value of 0.004 m yields *SCE* = 1.0, 0.67, and 0.24 in VIC, Mosaic, and Noah, respectively. The low *SCE* tendency of Noah is manifest later in notably low snow albedo compared to Mosaic (SAC does not require albedo), and the high *SCE* tendency of VIC produces higher snow albedo than Mosaic.

To compare model simulated *SCE* values with the binary "0/1" format of the IMS product, model *SCE* values were converted to 0/1 snow cover flags using an *SCE* threshold of 0.1. This *SCE* threshold was adopted among others after much testing and for the reasons described in *Sheffield et al.* [2003]-N. The validation of NLDAS *SCE* was then carried out separately over the 8 of the 12 RFC regions in the CONUS that experience substantial winter snow cover. Sample results for 4 of these 8 RFCs are provided in Figure 16, which shows the time series of daily basin-average snow cover area for each model and the IMS analysis for the period February 1997 (beginning of IMS archive) to September 1999. In general, the results

indicate broadly reasonable agreement between the modeled and observed basin-mean *SCE*, although there are systematic biases for all models and regionally and seasonally dependent differences in how well the models perform. Viewing all regions and seasons overall, the VIC model predicts the highest agreement and Noah the lowest, with SAC a close second to VIC and Mosaic in between SAC and Noah. The more mountainous regions (Northwest, Colorado, California, Nevada) appear to show the largest differences with the IMS observations and between models, where VIC notably overestimates *SCE* and Noah notably underestimates *SCE*. In the Northwest RFC (Figure 16, top panel), an early bias in Noah's spring snowmelt is especially evident. On the other hand, VIC's high *SCE* bias in the two western RFCs in Figure 16 is surprising, since the control runs of all four models, including VIC, had notably low bias in modeled *SWE* in Figure 15, which reflects the western states of the SNOTEL network. This paradox between VIC low *SWE* and high *SCE* bias, we surmise, results from VIC assigning *SCE* = 1 over any tile with any nonzero *SWE*, however small.

For the Northwest RFC domain, where the inter-model differences in *SCE* in Figure 16 were some of the largest, Figure 17 from *Sheffield et al*. [2003]-N shows the corresponding 3-year time series of monthly and domain mean i) snowmelt, ii) snow sublimation and iii) albedo from the four models (only snowmelt in SAC, as SAC excludes sublimation and albedo). There are large differences in snow albedo among the three SVATS, with Noah having very low albedo (0.2-0.3), Mosaic having moderate albedo (0.3-0.5), and VIC having high albedo (0.5-0.65) – a relative ordering consistent with the aforementioned ordering of Noah, Mosaic, and VIC having low, moderate, and high snow cover fractions for given moderate depths of *SWE*. The substantial differences in snow albedo translate directly into model differences in net solar insolation and hence into large differences in available energy for melting and sublimation. The

Noah snow albedo appears troublingly low, and the effect of this low bias on net solar insolation is exacerbated by the high bias in the NLDAS solar insolation forcing over snow. Not surprisingly then, Noah yields notably higher early and mid-winter snowmelt and sublimation, while VIC yields very little sublimation, with Mosaic in between. These larger Noah mid-winter snowpack sinks are consistent with Noah having the largest low bias in annual maximum *SWE* in Figure 15. By spring, Noah's snowmelt and sublimation reduce to almost zero as much of Noah's snowpack has already melted or sublimated earlier in the winter, while VIC and SAC have the largest spring snowmelt volumes, in part because they had the smallest sublimation sinks during early and mid-winter. SAC and Mosaic tend to have higher melt in the spring than the winter months while VIC melts at a more quasi-steady rate throughout the winter and spring. The high snow albedo in VIC may be too high, but this served to offset the high incoming insolation bias in NLDAS over snow.

The detrimental effect of low albedo or low snow cover on the energy balance of the snow pack is amplified by a positive feedback mechanism. Specifically, a low bias in snow cover contributes to a low bias in albedo, which contributes to a high bias in net insolation, which melts more snow and reduces the snow cover, leading to still lower albedo and so on. This feedback is of course accelerated and amplified by any high bias in the incoming solar insolation, such as the case in NLDAS. High albedo and high snow cover, such as in VIC, is also vulnerable to positive feedback in the opposite direction, but the high snow albedo in VIC here likely acted to offset the high bias in NLDAS incoming insolation over snow.

A counterpart to such feedback risks is the relative simplicity of snow processes in the SAC/SNOW17 model. Its snowpack predictions appear to perform just as well, if not better, than the three SVAT models when assessed at the large regional scales of NLDAS. One reason

for SAC's success may be its simple empirical treatment of snow melt using a temperature index approach that avoids the feedback loops that can plague energy balance models over snowpack.

4. Validation of the NLDAS surface energy fluxes and land surface skin temperature

This section examines the surface energy fluxes and land surface skin temperature (*LST*) of the three NLDAS models (Noah, VIC, Mosaic) that simulate the surface energy budget, which does not include the SAC model.

4.1 In situ validation of surface energy fluxes over the Southern Great Plains (SGP)

This section presents key results from *Robock et al.* [2003]-N. This study validates NLDAS surface energy fluxes during January 1998 - September 1999 using the 24 "extended facility" (EF) flux stations of the ARM/CART network, which spans portions of Oklahoma and Kansas. As *Robock et al.* [2003]-N discuss, it is difficult to assess NLDAS surface fluxes on a single station basis, because the NLDAS fluxes reflect the grid scale (about 11 x 14 km in the SGP) whereas single stations reflect the local plot scale – which can exhibit wide variations in soil texture, vegetation state, and soil moisture. Thus, multi-station spatial averaging and hourly temporal averaging is used here to reduce the influence of the scale differences. As cited in *Robock et al. [2003]-N*, such averaging is common in similar studies for similar reasons.

We strove to use as many stations in the averaging as possible; hence, radiation fluxes were averaged over the 22 of 24 EF stations operating Solar and Infrared Radiation Station instruments (SIRS), while heat fluxes were averaged over the 14 of 24 EF stations operating Energy Balance Bowen Ratio systems (EBBR). Since the number of SIRS and EBBR stations differ, the energy budget is not precisely closed in the averaging, but the discrepancy is less then 20 W m^{-2} in most months. Figure 18 shows the time series of monthly-mean model and

observed surface energy fluxes, averaged over the cited EF stations for nearly two annual cycles during 1998-1999, including net radiation (*R*), latent heat flux (*LE*), sensible heat flux (*H*), and ground heat flux (*G*). Columns 1-2 of Figure 19 provides the corresponding monthly-mean diurnal cycles for July and April 1999. (See *Robock et al*. [2003]-N for September 1999). Columns 3-4 of Figure 19 depict experiments described later.

In Figure 18, there is rather good agreement in monthly mean *R* between all three models and observations in virtually all months. In Figure 19, there is a one-hour phase lag in the diurnal cycle of daytime *R* in VIC and Noah, which as shown in *Robock et al*. [2003]-N arises from non-optimum techniques used in VIC and Noah to project the hourly input solar forcing onto the model physical time step (which will be rectified in future integrations).

Of more interest in Figures 18-19 are the results for *LE*, *H*, and *G*, which exhibit several situations of substantial bias. Figure 18 shows that in spring and summer, Mosaic has a substantial high bias in monthly mean *LE* and, correspondingly, a substantial low bias in *H*. In contrast in Figure 18, VIC has a substantial low bias in *LE* and high bias in *H* throughout most of the year (except spring), while Noah falls in between with notably smaller bias in monthly mean *LE* (low in warm season) and modest bias in *H* (high in warm season). The *LE* results here for Mosaic, VIC, and Noah agree with the sign and relative magnitude of the evaporation biases inferred from the water balance over non-sparse vegetation in Section 3. Additionally, we recall from Section 3 that the evaporation biases were reflected in counterpart runoff biases of opposite sign. The analogy in the surface energy context here is counterpart biases in sensible heat flux *H* of opposite sign to those of *LE*.

The warm season *LE* and *H* biases of Mosaic, VIC, and Noah are highlighted further in the July 1999 daytime biases depicted in Figure 19, again showing Mosaic with a significant

positive bias in daytime latent heat flux and substantial negative bias in daytime sensible heat flux, thus very low Bowen ratio (*BR*=*S*/*LE*, not shown), with VIC showing the opposite, and Noah showing bias similar to VIC in sign but notably smaller in magnitude. For April 1999, Mosaic retains its high *LE* bias and low *H* bias, while VIC and Noah show little bias in *LE* or *H*.

A second situation of large bias is that of *G* in VIC and especially Mosaic. Mosaic has a large bias in *G* during most months of Figure 18 and most hours in Figure 19, though the sign changes depending on the time in the diurnal/annual cycle. VIC has large hourly biases in *G* for most hours in Figure 19, but small monthly biases in Figure 18, because VIC's large daytime and nighttime biases in *G* are temporally symmetric and of opposite sign and thus they cancel for the most part on a daily or monthly mean basis. Noah shows little monthly bias in *G* in Figure 18, and a rather modest hourly bias, due mostly to a 2-hour phase error in the daytime maximum *G*. It is unique here to Mosaic that the very high daytime biases that occur simultaneously in *G* and *LE* in Mosaic during the warm season of April - July conspire to yield dramatically low Mosaic sensible heat flux *H* during these four months. In particular, the Mosaic values of *H* during May-June are less than the winter minimums in observed *H*.

It is noteworthy here that Mosaic and VIC are the two NLDAS land models that employ sub-grid vegetation tiling and hence considerably more computational load and complexity, yet they showed no advantage here over the SGP area in the quality of their simulated surface energy fluxes compared to the non-tiled Noah model. Admittedly the NLDAS grid is a rather high-resolution grid compared with coupled climate models and the latter result may or may not hold on the coarse grids of coupled global or regional climate models.

As the NLDAS project proceeds, more effort will be devoted to model experiments that seek to reduce the spread between the fluxes and states among the models and between models

and observations. One such experiment in ground flux is summarized here from *Robock et al*. [2003]-N. The Mosaic model utilizes the force-restore or "slab" treatment to predict its ground heat flux and surface radiative temperature. This method assumes that the surface slab has a non-trivial heat-storage capacity. The VIC and Noah models use a surface energy balance approach for the surface radiative temperature, calculated in VIC for a thin but non-vanishing skin layer that has a nontrivial heat storage capacity, while Noah assumes an infinitesimally thin skin layer with negligible (zero) heat storage. As a sensitivity experiment, VIC was re-executed by imposing zero heat storage in its skin layer, analogous to Noah. In this test, the VIC ground heat flux was dramatically improved, as demonstrated for July and April 1999 in the lowermost panels of the two right columns of Figure 19. *Robock et al.* [2003]-N show that this dramatic improvement in *G* in VIC holds for all months of the year. However, the corresponding response in VIC's high Bowen-ratio bias was disappointing. Specifically, columns 3-4 of Figure 19 show that in the VIC experiment, the reduction of VIC's high daytime bias in *G* is balanced for the most part by a further increase of the already high VIC bias in *H*, while *LE* and *R* showed negligible change. The lack of change in *LE* suggests an overly high canopy resistance in VIC. Sensitivity tests of canopy resistance will be a focus in all three SVAT models in NLDAS follow-on studies.

4.2 In situ validation of land surface skin temperature over the SGP

Our ultimate interest for future studies is to improve NLDAS soil moisture, and in turn surface fluxes, by assimilation of satellite-derived daytime *LST*. The prospects for positive impact from assimilation of satellite *LST* are greatest if errors in modeled *LST* stem primarily from errors in the assimilating model's Bowen ratio arising from errors in model soil moisture states. The prospects are far less favorable if *LST* errors arise from Bowen ratio errors caused

not by soil moisture errors, but by errors as the model's treatment of 1) vegetation cover and non-soil moisture attributes of its canopy conductance, 2) the surface air layer and its aerodynamic conductance (surface-layer turbulence), 3) albedo and thus net solar insolation, or 4) ground heat flux (G) and its impact on the available energy (R - G). Section 4.1 uncovered substantial daytime errors in G. This section uncovers significant impact on model *LST* from inter-model differences in the treatment of aerodynamic conductance.

At the top of each hour, all three models output an instantaneous, grid-cell mean, radiometric surface temperature, referred to here as the land surface skin temperature, *LST*. To obtain *LST*, each model applies the Stephan-Boltzmann Law given by $L = \varepsilon \sigma LST^4$, in which ε is the surface emissivity (= 1 in all three models), σ is the Stephan-Boltzmann constant, *LST* is the skin temperature (K), and L is the upwelling longwave radiation (W m^{-2}). In Noah, which is non-tiled, the surface energy budget is solved once for each grid cell to obtain *LST* and L. In Mosaic and VIC, *LST* and L are obtained for each tile from a tile-specific energy budget, then the tile-weighted mean L over the grid cell is obtained, from which the grid-cell mean *LST* is derived, again using $L = \varepsilon \sigma LST^4$. Finally, at each of the 22 SIRS stations, the observed L is time averaged to the top of the hour, and *LST* is obtained also from $L = \varepsilon \sigma LST^4$ using $\varepsilon = 1$.

Figure 20 shows the monthly-mean diurnal cycle of SIRS-observed and co-located model *LST*, averaged over all 22 SIRS sites, for July and April 1999. (The dashed line in Figure 20 is a Noah test described later.) Mosaic has a mid-day cool bias in both months, as expected given its high bias in *LE* and *G* and low bias in *H* in Figure 19 (columns1-2). VIC and Noah have a mid-day warm bias in July, and both have smaller April than July *LST* bias, also as expected from Figure 19, given their low *LE* and high *H* bias in July and small *LE* and *H* bias in April.

While the sign of the models' mid-day *LST* bias in Figure 20 is generally as expected overall, the comparative magnitude of the *LST* bias between the models is totally unexpected, especially in July, given the *LE* and *H* fluxes in Figure 19. Specifically, the VIC mid-day (19-20 UTC) warm bias in July (about +2 K) is about half as large as Noah (about +4 K) and about equal to the magnitude of the Mosaic mid-day cool bias (about -2 K), despite VIC's Bowen ratio (*BR* = 2.91) at this time being much higher than Noah's (*BR* = 0.70) and the observed (*BR* = 0.38) and notably more disparate from the observed than Mosaic's (*BR* = 0.14).

Thus, VIC does not yield the largest July *LST* bias, despite having the largest *BR* bias. One explanation is the large bias in mid-day *G* noted in VIC and Mosaic in Figure 19 (columns 1-2, bottom frames). This large *G* bias provides an inflated heat storage, which in VIC reduces the mid-day warm bias induced by VIC's high *BR*, while in Mosaic it acts to amplify the mid-day cool bias induced by Mosaic's low *BR*.

We next show that a second explanation (especially relevant in the Noah case, as Noah had much smaller biases in *G*) lies in significant inter-model differences in the aerodynamic conductance of the surface layer. In the three models, the sensible heat flux *H* (W m^{-2}) is computed from the typical bulk transfer formulation given by

$$H = -\rho\, c_p\, C_h\, |V|\, (T_a - LST) \qquad (2)$$

where ρ is the air density (kg m^{-3}), c_p the heat capacity for air (1004.5 J kg^{-1} K^{-1}), $|V|$ the wind speed (m s^{-1}), T_a the air temperature (K), and C_h the surface turbulent exchange coefficient for heat (m s^{-1}) – also known as the aerodynamic conductance ($1/C_h$ is the aerodynamic resistance). C_h typically manifests a strong diurnal cycle with larger values during daytime heating. In (2),

positive H means a heat source to the atmosphere and heat sink to the land surface for daytime *LST* exceeding T_a. The models get common surface forcing values of ρ, $|V|$, and T_a. Only C_h and *LST* in (2) are computed uniquely in each model. Therefore Noah can have higher mid-day values of *LST* than VIC simultaneously with lower mid-day values of H than VIC if and only if Noah has lower daytime values of C_h. Figure 21 depicts the July 1998 monthly-mean diurnal cycle of C_h for each model, averaged across the 14 EBBR stations. The dashed line in Figure 21 is a Noah test case discussed later. Indeed, Noah has substantially smaller daytime values of C_h than Mosaic, and far smaller daytime values than VIC. It is not straightforward to judge which control-run C_h time series in Figure 21 is the most reasonable in terms of magnitude. Future NLDAS research will seek to derive C_h explicitly from the EBBR station observations.

In diurnal character, Noah C_h values vary smoothly over the diurnal cycle and peak in the early afternoon, as expected, as do Mosaic values for the most part (though Mosaic C_h manifests an odd jump behavior in early morning, which deserves future investigation). C_h in VIC peaks surprisingly late, in the early evening, and then drops rather slowly during the night, partially explaining the phase-lag error in the nighttime *LST* decrease in VIC in Figure 20.

The low daytime C_h values for Noah in Figure 21 motivated a Noah sensitivity run. The formulation for C_h in the Noah model was the subject of the NCEP study by *Chen et al.* [1997], which included tests of three separate schemes for C_h (for details, see *Chen et al.*). The three schemes gave very similar values for C_h and the resulting surface heat fluxes, if the same treatment was used for the roughness length for heat, z_{0t}, and in particular, the ratio z_{0t}/z_{0m}, where z_{0m} is the roughness length for momentum (which is fixed at each grid point in NLDAS from the assigned vegetation type). After a suite of tests, *Chen et al.* recommended using the Paulson scheme for C_h in the Noah model. More importantly, they recommended the z_{0t}/z_{0m} formulation

of *Zilitinkevich* [1995]. While the reader is referred to *Chen et al.* for details of this formulation, it is flow-dependent and based on the roughness Reynolds number (which includes the surface friction velocity u^*) and an adjustable empirical parameter, denoted here C_z, in the range 0-1. The Noah control run here used $C_z = 0.2$. In the limit $C_z \rightarrow 0$, $z_{0t}/z_{0m} = 1$. Throughout the C_z range, decreasing C_z increases z_{0t}, which increases both C_h (i.e. aerodynamic conductance) and the land/atmosphere coupling, thereby decreasing daytime *LST*.

This is demonstrated here by a Noah sensitivity test, depicted by the dashed curve in Figure 21, which used $C_z = 0.05$ in place of $C_z = 0.2$ in the Noah C_h formulation. The July monthly-mean, mid-day C_h values in the Noah test exceed the Noah control values by nearly 70%. The *LST* and surface fluxes of this Noah test are depicted, respectively, by the blue dashed line in Figure 20 and columns 3-4 in Figure 19. Figure 20 shows a pleasing 2-3 K decrease in Noah peak daytime *LST* in both July and April, cutting the July warm bias in half and virtually eliminating the April warm bias. Inspection of the numbers associated with the Noah test and control surface fluxes in Figure 19 (columns 3-4) reveals small mid-day changes in *R* (increase) and *G* (decrease) of about 10-20 W m^{-2} each, as expected from the lower mid-day *LST*. This improved the already small biases in Noah *R* and *G* and yielded a roughly 20-40 W m^{-2} increase in mid-day "available energy" (*R* - *G*) for sensible and latent heat fluxes. Interestingly, with regard to *LE* changes in the test, though the mid-day aerodynamic conductance increased about 65-70 percent compared to the control, both in July (Figure 21) and April (not shown) of 1999, the *LE* change in the test in both months was negligible, especially so in July, hence the *R* - *G* increase was realized almost entirely by an increase in *H* in the July case and by a more even-split of a tiny increase in *H* and *LE* in the April case (maintaining the good Noah *BR* in April). The lack of change in *LE* in the July test (which has drier soil than April, see Figure 14) strongly

indicates that the evaporative resistances to canopy transpiration (canopy resistance) and bare soil evaporation are much larger than, and thus dominant over, the aerodynamic resistance in influencing *LE* in this situation. Figure 1a shows that the vegetation cover over the SGP ARM region in July is of order 50 percent. The analysis in Vogel et al. [1995] shows that even over an irrigated midlatitude wheat field in June, the LE change from a 20 percent change in aerodynamic resistance yielded only a 2 percent change in LE --that is, canopy resistances over non-sparse vegetation in mid-latitude summer are typically much larger than, and dominant over, aerodynamic resistances, even when the soil is wet and not contributing to a rise in canopy conductance. These results strongly suggest that the remaining July bias of +2 K in the Noah test in Figure 20 is caused by an overly high canopy resistance (e.g. Sensitivity tests of aerodynamic resistance and canopy resistance will be major focus in follow-on NLDAS studies.

4.3 Satellite-based validation of land-surface skin temperature.

This section demonstrates the use of hourly GOES-based *LST* retrievals to validate NLDAS *LST* over large regions of the CONUS. First, we validate the GOES *LST* against the ARM *LST* observations from Section 4.2. Then we use GOES *LST* to assess NLDAS *LST* over north central CONUS. This material is not presented in the other NLDAS papers of this issue.

The study here is limited to CONUS regions east of the Rocky and west of the Appalachian Mountains, to avoid mountain-shadowing effects on the GOES retrievals. The retrievals are obtained from GOES-East (GOES-8) and they provide gridded fields of hourly *LST* at 0.5° spatial resolution in cloud-free conditions during daytime throughout the year. The GOES-East LST fields span 25°-53°N and 65°-125°W. The underlying target consists of 9 x 8 arrays of 4 km center-to-center pixels (at nadir), where the visible data have been averaged up to 4 km resolution coincident with the 4 km resolution infrared pixels. The resultant nominal target

area is approximately 50 km on a side, or roughly 0.5°. The *LST* retrieval provides a single "aggregate" *LST* for the entire 50-km scene.

In NLDAS, we bilinearly interpolate the 0.5° *LST* fields to the 1/8° NLDAS grid. The GOES *LST* fields are produced by the GCIP/GAPP partnership in GOES land-surface retrievals between NESDIS and the University of Maryland [*Pinker et al.*, 2003]-N. The GOES *LST* is retrieved only at 0.5° targets deemed 100% cloud-free at the retrieval time. The cloud detection algorithm is based on that of earlier GOES insolation-retrieval studies such as *Tarpley* [1979], as refined in later studies such as *Pinker et al.* [2003]-N. Despite the 100% cloud-free screening criteria, clouds may still be present in the scene owing to 1) optically thin cirrus, 2) sub-resolution or "sub-pixel" cloud (e.g., fair weather cumulus), and 3) difficulty of cloud detection over snow cover.

GOES *LST* is retrieved by the so-called "split-window" technique of *Wu et al.* [1999], in which *LST* is obtained from a linear regression of the GOES brightness temperatures in the 11 μm and 12 μm bands. The regression coefficients were derived by *Wu et al*. from GOES brightness temperatures coincident with sea surface temperature (*SST*) observations from an ocean buoy under cloud-free conditions. Though the coefficients were derived from *SST* observations, we find the land-based validation here to be excellent. More specifically, though the regression coefficients were derived assuming a surface emissivity of $\varepsilon = 1$ valid over sea, this emissivity assumption is also valid over land surfaces of non-sparse vegetation or snowpack, but less valid over rather bare soils (wherein $\varepsilon = 0.91$-0.97). Uncertainty from emissivity issues is avoided in this study by staying over non-spare vegetation and by our universal application of $\varepsilon = 1$ in 1) the NLDAS models, 2) the in situ ARM/SIRS sites and 3) the GOES retrievals.

We proceed next to assess GOES LST against the in situ *LST* observations of the 22 ARM/CART SIRS sites. We limit the assessment to the warm season, as our future assimilation of GOES *LST* will generally be confined to this period of notably stronger coupling between *LST* and soil moisture. Figure 22 presents the monthly and multi-station mean of the daytime hourly diurnal cycle of GOES *LST* (dashed) and ARM *LST* (solid) for April and July in 1998 and 1999. The data samples for Figure 22 (and Figure 23) represent only locations and times when the GOES cloud screening detected zero cloud. In Figure 22, the GOES *LST* demonstrates a remarkable ability to match the station-observed mean diurnal cycle, though it does show a small cool bias (likely from undetected clouds) of order 0-1.5 K before 18 UTC (local noon) and 1-2.5 K thereafter. The smaller cool bias in the morning is likely from less prevalent cloud cover then (especially sub-pixel cumulus). In future data assimilation, one may mitigate this cool bias by assimilating the 3-hour increase of GOES *LST* before noon (about 15-18 UTC here), rather than *LST* itself. *Tarpley* [1994] applied the morning rise of GOES *LST* to infer a monthly mean surface moisture availability over Kansas. Like the ARM *LST*, the 18 UTC GOES *LST* in Figure 22 is warmer (3-4 K) and its preceding 3-hour morning rise is larger (by about 1K) in July 1998 than July1999. This interannual *LST* variability reflects the drought that occurred over the ARM region in July 1998 (see 1998 versus 1999 July soil moisture in Figure 14), thus conditions were warmer and drier than in 1999 (and likely less cloudy, hence the smaller GOES *LST* afternoon cool bias in July 1998 versus 1999).

Our goal is to use GOES *LST* retrievals to assess NLDAS *LST* over large regions that lack in situ observations of *LST* or surface fluxes. As a benchmark for that goal, Figure 23 illustrates, at the ARM SIRS sites during July and April 1999 for 18 UTC (near local noon), a pleasing similarity between GOES-based and ARM-based site-by-site match-ups with model

LST. Moreover, all three models show good skill in either the GOES or ARM validation setting by yielding rather tight clusters close to the diagonal (and hence high correlations, shown later). In each month, the separate GOES and ARM match-ups use the same sample of instances where the GOES deemed the site to be cloud free. In Figure 23, the sample size of 198 in April (out of a possible 660 = 30 days x 22 stations) is notably smaller than that of 334 (out of 682) in July, as the GOES cloud screening detects cloud more often in the spring. One would expect this from the natural trend of decreasing cloud cover from spring to summer and the greater likelihood in July of shallow, sub-resolvable cumulus. Indeed in July, the GOES LST in Figure 23 manifests a small (but non-negligible) leftward-pointing "cold tail" of outlier values that are not present in either the ARM observations or the models and thus likely represent GOES cloud detection failures. Similarly, the ARM observations in April 1999 show several warm outlier values (near 315 K), not present in either the GOES or model *LST*, likely representing bad ARM values.

Most importantly, as desired, the GOES versus model match-ups yield the same sense of model mid-day *LST* bias as we derived from ARM data alone in the prior section. Table 4 compares the GOES-based versus ARM-based model bias, error standard deviation and correlation obtained from the Figure 23 match-up and listed top-down from warmest to coldest model bias. The table shows good agreement between the sign and magnitude of the GOES-based and ARM-based model bias. The GOES-based model bias is order 1 K warmer than the ARM-based model bias, owing to the aforementioned GOES *LST* cool bias of order 1 K versus ARM *LST*. The GOES-based model *LST* bias essentially reproduced the ARM-based model bias, both in 1) the absolute sense of correct sign and reasonably good magnitude and 2) the relative sense between models and between spring and summer season. Specifically, the GOES-based model bias (in agreement with Figure 20) shows that in summer 1) Noah has the largest

warm bias, which becomes substantially smaller and modest in spring, 2) VIC has a smaller and modest warm bias, which becomes virtually zero in spring, and 3) Mosaic has a modest cool bias, which becomes larger in spring. Similarly, the GOES-based and ARM-based standard deviations in Table 4 from Figure 23 are in reasonable agreement—both in the range of 3-4 K. Finally, the GOES-based correlations with model LST in Table 4 are very encouraging, ranging between 0.66 and 0.78 – with five of six values of 0.70 or more, all without any screening of the GOES cold *LST* outliers. The correlations of model *LST* with GOES *LST* are consistently higher than the ARM-based correlations, likely from the better match of the spatial scale of the GOES footprint and the NLDAS grid-mesh size, versus the point scale of the ARM values.

Encouraged by the results above, we evaluated model *LST* against GOES *LST* in Figure 24 across a large region of northern Midwest, bounded by latitudes 39°N and 53°N and longitudes 82°W and 98°W (straddling 90°W longitude, representing strict local noon at 18 UTC). This region is chosen for its 1) spatial separation from SGP, 2) non-sparse green vegetation in summer (Figure 1a), and 3) vivid model differences in evaporation in Section 3. GOES versus model *LST* over this region at 18 UTC for July and April 1999 are presented in Figure 24. The so-called "screened" results in Figure 24 are described later.

The sample counts in Figure 24 are order 70,000-100,000 (2-3 orders larger than Figure 23, owing to the larger region). To depict data density in Figure 24, we binned the data into 1 K intervals (for display only, kept full precision in statistics) and used colors for four orders of data counts: 0-10 (red), 10-99 (yellow), 100-999 (light green), and 1000-9999 (dark green). The green shades depict the vast majority of the sample and they manifest rather well behaved, elongated data clusters lying near and parallel to the diagonal, as desired.

Yet, the red and yellow left-pointing "tails" of cold GOES *LST* in the unscreened panels of Figure 24 depict a non-trivial number of data points that likely represent failures of the GOES cloud detection, similar to the cold tails in the July GOES panels of Figure 23. Hence, we used the model simulated *LST* to screen out these tails of cold GOES *LST* in columns 2 and 4 in Figure 23. Specifically, we rejected a GOES *LST* data value if model-minus-GOES *LST* exceeded the unscreened, sample-wide model-minus-GOES *LST* bias by more than two times the model-minus-GOES *LST* standard deviation of the unscreened sample. Figure 24 shows the screening preserves the high-density core region of the original data cluster, while eliminating the cold tails. The data counts of the three panels in any column of Figure 24 are identical in the unscreened case, but differ slightly (less than 1.2%) in the screened case, as the screening for each panel uses the given model's *LST*.

Table 5 shows the GOES-based model *LST* bias, error standard deviation and correlation obtained from Figure 24. In Table 5, the screened model bias results are warmer, as expected, but by a modest 0.5 to 1.0 K. More notably in the screened case, the standard deviation of the model errors are significantly less by around 1.0-1.5 K and the already high correlations increase by around 0.1 to 0.76 or higher for almost all entries. In July, the correlation of Mosaic and Noah LST with GOES reaches impressively high values of 0.82 in the screened case, with that of VIC close by at 0.76. In April, Mosaic and VIC also achieve high correlations in the range 0.76-0.81, while that of Noah is less at 0.63. It is instructive to compare the unscreened results from the north central CONUS in Table 5 with the SGP results in Table 4 (all unscreened). Table 5 preserves the relative nature and order noted in the biases in Table 4 – namely, Noah is warmest and Mosaic is coolest in July, with Noah notably less warm and Mosaic notably more cool in April, while VIC remains in-between both Noah and Mosaic in both months. Yet, the

unscreened biases in Table 5 are consistently 1-3 K warmer than those in Table 4, reflecting that either the models are warmer relative to GOES *LST* in this region than over the ARM SGP region, or the GOES *LST* has a larger cool bias (of order 1-3 K) than the GOES cool bias over the ARM region. We speculate that it is the latter, owing to more cloud contamination, since the given region is both further east (later in the solar day at 18 UTC) and generally more humid than the ARM region, increasing the likelihood of sub-resolution cumulus clouds.

The final feature we point out is in the April results of Figure 24, wherein we note a "lower lobe" of cold model LST in both the case of Mosaic and especially VIC, which extends down to a sharp cutoff near the freezing point. This indicates that Mosaic and VIC possibly maintain remnants of melting snow pack somewhat too late into the spring in this region, in agreement with a similar conclusion in *Sheffield et al*. [2003]-N.

Given the encouraging GOES *LST* results of Sections 4.2-4.3 and Figures 22-24, we look forward to use of GOES *LST* in future *LST* assimilation studies and to assess NLDAS land surface model improvements in the warm season between the Rocky and Appalachian Mountains. The validation and utility of GOES *LST* in mountainous regions awaits future study. Lastly, efforts continue for the improvement of GOES cloud detection.

5. Summary and Conclusions

A multi-institution partnership, with GCIP support, has developed and evaluated the backbone for a North American Land Data Assimilation System (NLDAS). The partners have integrated a diverse set of GCIP products and other data sources into robust forcing data sets and multi-scale validation tools. NLDAS employs an uncoupled approach featuring non-model (less biased) sources of precipitation and solar insolation and the four LSMs of Noah, VIC, Mosaic, and Sacramento (SAC) executing on a high resolution (1/8°) CONUS domain. On this domain,

the NLDAS provides land-state background fields over a wide variety of climatological regions for future data assimilation experiments. The NLDAS infrastructure, which includes streamflow routing, supports both realtime and retrospective executions. The hourly NLDAS forcing database, which now spans nearly seven years from October 1996, represents an important NLDAS by-product that should prove widely useful to many LSM modelers.

The NLDAS surface forcing compared well against mesonet observations over the SGP region. In NLDAS test runs using local flux-station observed forcing in place of traditional NLDAS forcing, the test versus control differences in the LSMs soil moisture and temperature were small, and much smaller than the differences between the models and between the models and observations. Over high elevations in mountains of western CONUS, NLDAS precipitation forcing suffers a large low annual bias of order 50 percent. To increase the precipitation over high mountainous, NLDAS collaborators have developed and implemented a PRISM-based technique [*Daly et al*., 1994] into the realtime NLDAS forcing, but have not yet had sufficient time to regenerate the NLDAS retrospective forcing with this change.

The assessment here focused on the retrospective NLDAS execution for the 3-years beginning October 1, 1996. Gage-based analyses of observed precipitation and a large set of streamflow measurements were used in a continental assessment of the annual-mean water budget of the simulated evaporation and runoff of the four NLDAS LSMs. This assessment uncovered substantial model biases and inter-model disparities, especially over the non-sparsely vegetated region east of 98° W longitude (CONUS East). Subsequent evaluation of the monthly energy budget over the ARM SGP region, confirmed the sense of the biases obtained over CONUS-East from the annual water budget. Specifically, the Noah LSM showed the smallest systematic bias in evaporation and hence also the smallest bias in runoff and sensible heat flux,

while VIC showed a rather large low bias in evaporation and hence high bias in runoff and sensible heat flux. SAC and Mosaic showed a rather high bias in evaporation and hence large low bias in runoff and sensible heat flux (Mosaic). The use of vegetation tiles in Mosaic and VIC did not aid visibly aid their performance relative to the Noah model, which omits tiling.

The large disparity between the LSMs in evaporation over non-sparse vegetation is the foremost issue to be addressed in follow-on NLDAS studies. The three SVAT-type models, though they treat vegetation cover explicitly, nevertheless yield strikingly different warm season evaporation totals over vegetation. Such disparity has also been noted between SVATs in various phases of PILPS [*Chen et al.*, 1997; *Wood et al*., 1998]. In contrast to PILPS however, the continuing NLDAS collaboration affords the partners an essential opportunity to pursue a host of follow-on sensitivity experiments and diagnostic studies to dissect and thoroughly characterize the various model treatments of individual transpiration-related components, such as root depth and density, vegetation cover and its seasonal cycle, canopy interception and re-evaporation, soil moisture stress thresholds (and other canopy resistance factors), and the uptake of water from below the root zone. For example, quite surprisingly, despite Mosaic having the shallowest root zone by far of the three SVATs in NLDAS, it had the highest warm season evaporation rates (and largest high bias) and hence the largest warm season storage change in soil moisture, because Mosaic was shown to have the most vigorous upward transport of soil water from below the root zone via vertical diffusion.

We emphasize that unlike the Noah model, the Mosaic, VIC, and SAC LSM configurations in NLDAS differed in important aspects, albeit by intent by their overseers, from their traditional configurations that had yielded rather good performance in other settings. Follow-on NLDAS investigations are now underway to quantify how much impact these

configuration differences had on NLDAS LSM performance. For example, Mosaic was executed for the first time with fixed rather than variable soil-layer thickness and root depth. VIC was executed with one-hour rather than its typical three-hour time steps and with hourly rather than daily uniform disaggregation of precipitation.

The SAC configuration in NLDAS in particular was non-standard, yet represented a major milestone. The SAC runs in NLDAS, by design, were the first executions of the newly available distributed version of SAC over a large-scale continental domain and with a priori non-calibrated parameters. Though SAC showed a high (low) bias in evaporation (runoff) over the CONUS-East, SAC was not an isolated outlier in this regard, as it agreed closely with the Mosaic annual water budget. Thus the NLDAS executions of SAC provide an essential initial benchmark for future experiments of CONUS-wide executions of SAC in distributed mode. Future executions of SAC over the NLDAS domain need some modest SAC calibration efforts, in particular the calibration of the fractional multiplier of potential evaporation in SAC to serve as a poorman canopy resistance.

Aerodynamic conductance (C_h) emerged as a second area of large disparity. We found that overly large or small mid-day C_h values may substantially distort the expected correlation between LST bias and Bowen ratio bias. For example, the Noah model had the largest summer mid-day warm bias over the SGP reason, despite having the best Bowen ratio, because Noah's aerodynamic conductance was substantially low compared to the other two SVATs. The aerodynamic conductance issue has strong implications for viable land assimilation of LST, as the latter depends on robust correlation between LST, Bowen ratio, and soil moisture.

Soil moisture storage arose as a third area of large disparity. The present and companion papers highlighted the wide differences in LSM soil moisture and soil water storage, both

between the models themselves and between some models and observations. This too has been demonstrated in previous PILPS studies. In this study, there were cases of a given LSM's mean area soil moisture comparing reasonably well with observations over Oklahoma, but not with soil moisture over Illinois, and vice versa. In coupled modeling, where surface fluxes are of paramount importance, it is sufficient to get the temporal changes in soil moisture correct, as good temporal changes can yield good surface energy and water fluxes. However, land assimilation of satellite radiances imposes a more stringent requirement for good absolute states of soil moisture, because the forward radiative transfer models that drive assimilation of land-related satellite radiances are sensitive to the absolute states.

Snowpack states, snow cover fraction, and snow albedo represented a fourth area of large disparity among the NLDAS LSMs. The VIC and SAC models provided the best simulations of snowpack water content compared to SNOTEL observations. The elevation tiling unique to VIC appears to provide a definite advantage compared to Mosaic and Noah. Snowpack simulations emerged as the leading weakness of the Noah LSM, which shows a large early bias in timing of seasonal snowmelt. The main culprit appears to be rather low albedo over snowpack in Noah – much lower than in Mosaic and especially VIC. The SAC snowpack simulations performed rather well, as its simple temperature-index based snowpack model bypasses energy balance and albedo and thereby bypasses positive feedback loops that haunt snowpack simulation in models with surface energy balance. The low bias in the NLDAS precipitation forcing at high mountain elevations (e.g. SNOTEL stations) and the high bias in NLDAS solar insolation forcing over snowpack complicated the attribution of cause and effect in poor LSM simulations of snowpack.

Lastly, this paper concluded with assessments and application of satellite-derived, GOES-based, hourly 1/2° LST obtained from the split-window technique. The diurnal cycle of the

GOES-based LST performed well when verified against spring and summer flux-station LST observations over the SGP ARM network. Follow-on use of GOES LST to validate LSM LST over the ARM network yielded LST validation scores very similar to those obtained from ARM LST observations themselves. Subsequently, we applied GOES LST to validate LSM LST over the northern Midwest. We found LSM LST biases that were consistent with those found over the SGP area, indicating that GOES LST provides a powerful large-scale LSM validation tool.

We are now assembling the tools to perform actual land data assimilation experiments. For this purpose, we will be adding one or two forward radiative transfer models into our common NLDAS infrastructure. Additionally, we will pursue adjoint model and ensemble Kalman filter methodologies. As a preview, we have thus far developed an adjoint of the Noah land model and have used ii to date to conduct identical-twin LST assimilation experiments.

The NLDAS initiative here formed an important pathfinder for a companion extension to a Global Land Data Assimilation System (GLDAS) by the NASA and NCEP partners of NLDAS. The GLDAS is described by *Rodell et al*.[2003]. Additionally, the NLDAS initiative here was the forerunner to companion initiatives in 50+ year retrospective executions of VIC on the NLDAS grid by *Maurer et al*. [2002] and of Noah on the NLDAS grid by H. Van den Dool of NCEP (private communication). Aside from the gage-based daily precipitation analyses, these multi-decade NLDAS executions necessarily use rather different data sources for the land surface forcing (e.g. global reanalysis). We seek to compare our NLDAS states and fluxes with 1) the aforementioned 50+ year retrospectives, 2) operational global and regional coupled 4DDA, and 3) global and regional coupled Reanalysis counterparts. Above all, we strive to provide NLDAS initial conditions (soil moisture and snowpack) for land-memory predictability studies and realtime weather and climate model forecasts on daily to seasonal time scales.

Appendix: List of Acronyms

4DDA	4-dimensional data assimilation
ARM/CART	Atmospheric Radiation Measurement/Cloud and Radiation Testbed (DOE)
ARS FAO	Agricultural Research Service Food and Agriculture Organization
AVHRR	Advanced Very High Resolution Radiometer on NOAA polar satellites
CAPE	convective available potential energy
CLIVAR	Climate Variability and Predictability
CONUS	Continental United States
CPC	Climate Prediction Center
DMIP	Distributed Model Intercomparison Project
DOE	Department of Energy
EDAS	NCEP Eta model-based 4-D Data Assimilation System
EMC	Environmental Modeling Center of NCEP
EBBR	Energy Balance Bowen Ratio flux stations in ARM/CART network
EF	Extended Facility flux stations in ARM/CART network
GAPP	GEWEX America Prediction Project
GCIP	GEWEX Continental-Scale International Project
GEWEX	Global Energy and Water Cycle Experiment
GLASS	GEWEX Land Atmosphere Systems Study
GOES	Geosynchronous Operational Environmental Satellite (USA)
GRIB	Gridded Binary data file format (WMO standard)
GSFC	Goddard Space Flight Center (NASA)
GSWP	Global Soil Wetness Project

GTOPO30	Global (30 arc seconds) digital elevation database
GVF	green vegetation-cover fraction
IMS	Interactive Multi-Sensor Snow (NESDIS)
LAI	leaf area index
LDAS	land data assimilation system
LSM	land surface model
LST	land-surface skin temperature
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NCDC	National Climatic Data Center
NCEP	National Centers for Environmental Prediction
NDVI	Normalized Difference Vegetation Index
NESDIS	National Environmental Satellite, Data, and Information Service
NLDAS	North American LDAS
NOAA	National Oceanic and Atmospheric Administration
NRCS	National Resources Conservation Service (USDA)
NWIS	National Water Information System
NWP	numerical weather prediction
NWS	National Weather Service (NOAA)
OGP	Office of Global Programs (NOAA)
OHD	Office of Hydrologic Development (NWS, formerly Office of Hydrology)
ORA	Office of Research and Applications of NESDIS
PAR	photosynthetically active radiation

PILPS	Project for Intercomparison of Land-surface Parameterization Schemes
PRISM	Parameter-elevation Regressions on Independent Slopes Model
RFC	River Forecast Center (NWS)
SAC	Sacramento model (Sacramento Soil Water Accounting Model)
SCE	Snow Cover Extent
SGP	Southern Great Plains (field program)
SIRS	Solar and Infrared Radiation Station
SNOTEL	Snowpack Telemetry network of the NRSC
SNOW-17	Snow accumulation and ablation model (NWS/OHD)
SST	sea surface temperature
STATSGO	State Soil Geographic Database
SURFRAD	Surface Radiation Budget Network (NOAA/OGP)
SVAT	Surface-Vegetation-Atmosphere Transfer (model)
SWE	Snowpack Water Equivalent
TOGA	Tropical Ocean Global Atmosphere
USGS	US Geological Survey
VIC	Variable Infiltration Capacity LSM
WMO	World Meteorological Organization
WSR-88D	Weather Service Radar - Doppler

Acknowledgements

The work by NCEP/EMC, NWS/OHD, and NESDIS/ORA was supported by the NOAA OGP grant for the NOAA Core Project for GCIP/GAPP (K. Mitchell, J. Schaake, J. Tarpley, Co-PIs). The work by NASA/GSFC was supported by NASA's Terrestrial Hydrology Program (P. Houser, PI). The work by Rutgers University was supported by NOAA OGP GAPP grant GC99-443b (A. Robock, PI), the Cook College Center for Environmental Prediction, and the New Jersey Agricultural Experiment Station. The work by Princeton was supported by NOAA OGP GAPP grant NA86GPO258 (E. Wood, PI). The work by NCEP/CPC was supported by NOAA/NASA GAPP Project 8R1DA114 (R. Higgins, PI). The work by University of Maryland was supported by grants NA56GPO233, NA86GPO202, and NA06GPO404 from NOAA OGP and by NOAA grant NA57WC0340 to University of Maryland's Cooperative Institute for Climate Studies (R. Pinker, PI). Some figures were drawn with GrADS, created by Brian Doty. We thank DOE for the ARM/CART meteorological and heat flux data that were provided to the project at no cost, and the NOAA Office of Global Programs and NASA Land Surface Hydrology Program for their purchase of the Oklahoma Mesonet meteorological and soil moisture and temperature data for their funded investigators.

References

Abdulla, F. A., D. P. Lettenmaier, E. F. Wood, and J. A. Smith, Application of a macroscale hydrologic model to estimate the water balance of the Arkansas-Red river basin, *J. Geophys. Res.*, *101*, 7449-7459, 1996.

Anderson, E. A., National Weather Service River Forecast System - Snow Accumulation and Ablation Model, NOAA Technical Memo. NWS Hydro-17, Silver Spring, MD, 1973.

Berbery. E., Y. Luo, K. Mitchell, A. Betts, Eta model estimated land surface processes and the hydrological cycle of the Mississippi Basin, *J. Geophys. Res.*, (submitted) 2003.

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

Boone, A., F. Habets, J. Noilhan, E. Blyth, D. Clark, P. Dirmeyer, Y. Gusev,I. Haddeland, R. Koster, D. Lohmann, S. Mahanama, K. Mitchell, O. Nasonova, G.-Y. Niu, A. Pitman, J. Polcher, A. B. Shmakin, K. Tanaka, B. Van Den Hurk, S. Verant, D. Verseghy, P. Viterbo, The Rhone-Aggregation Land Surface Scheme Intercomparison Project: An Overview of Results, *J.Climate*, accepted 2003.

Bowling, L.C., D.P. Lettenmaier, B. Nijssen, L.P. Graham, D.B. Clark, M.E.Maayar, R. Essery, S. Goers, Y.M. Gusev, F. Habets, B. van den Hurk, J. Jin, D. Kahan, D. Lohmann, X. Ma, S. Mahanama, D. Mocko, O. Nasonova, G.-Y. Niu, P. Samuelsson, A.B. Shmakin, K. Takata, D. Verseghy, P. Viterbo, Y. Xia, Y. Xue, Z.-L. Yang, Simulation of high latitude hydrological processes in the Torne-Kalix basin: PILPS Phase 2(e), 1: Experiment description and summary intercomparisons, *Glob. Plan. Change*, (in press) 2003.

Burnash, R. J. C., R. L. Ferral, and R. A. McGuire, A generalized streamflow simulation system: conceptual models for digital computers, Joint Fed.-State River Forecast Center, Sacramento, CA, 1973.

Chen, F., Z. Janjic, and K. Mitchell, Impact of atmospheric surface-layer parameterizations in the new land-surface scheme of the NCEP mesoscale Eta model, *Boundary-Layer Meteorol.*, *85*, 391-421, 1997.

Chen, F., and K. Mitchell, Using the GEWEX/ISLSCP forcing data to simulate global soil moisture fields and hydrological cycle for 1987-1988, *J. Meteor. Soc. Japan*, *77*, 167-182, 1999.

Chen, F., K. Mitchell, J. Schaake, Y. Xue, H.-L. Pan, V. Koren, Q. Y. Duan, M. Ek, and A. Betts, Modeling of land-surface evaporation by four schemes and comparison with FIFE observations, *J. Geophys. Res.*, *101*, 7251-7268, 1996.

Chen, T.H., A. Henderson-Sellers, P.C.D. Milly, A. J. Pitman, A.C.M Beljaars, and 38 others, Cabauw experimental results from the Project for Intercomparison of Landsurface Parameterization Schemes (PILPS), J. Climate, 10, 1194-1215, 1997.

Cherkauer, K. A., and D. P. Lettenmaier, Hydrologic effects of frozen soils in the upper Mississippi River basin, *J. Geophys. Res.*, *104*, 19,599-19,610, 1999.

Cosgrove, B. A., D. Lohmann, C. H. Marshall, K. E. Mitchell, P. R. Houser, E. F. Wood, J. C. Schaake, A. Robock, J. Sheffield, Q. Duan, L. Luo, R. W. Higgins, R. T. Pinker, J. D. Tarpley, J. Meng, Realtime and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project, *J. Geophys. Res.*, (submitted) 2003a.

Cosgrove, B. A., K. E. Mitchell, P. R. Houser, E. F. Wood, J. C. Schaake, A. Robock, D. Lohmann, J. Sheffield, Q. Duan, L. Luo, R. W. Higgins, R. T. Pinker, J. D. Tarpley, Land surface model spin-up behavior in the North American Land Data Assimilation System (NLDAS), *J. Geophys. Res.*, (submitted) 2003b.

Daly, C., R. P. Neilson, and D. L. Phillips: A Statistical-Topographic Model for Mapping Climatological Precipitation over Mountainous Terrain, *J. Appl. Meteor., 33*, 140-158, 1994.

Dirmeyer, P., A. Dolman, and N. Sato, The pilot phase of the Global Soil Wetness Project, *Bull. Amer. Meteor. Soc.*, *80*, 851-878, 1999.

Douville, H., P. Viterbo, J.-F. Mahfouf, and A. C. M. Beljaars, Evaluation of the optimum interpolation and nudging techniques for soil moisture analysis using FIFE data, *Mon. Wea. Rev.*, *128,* 1733–1756, 2000.

Ek, M. B., K. E. Mitchell, Y. Lin, P. Grunmann, E. Rogers, G. Gayno, V. Koren, Implementation of the upgraded Noah land-surface model in the NCEP operational mesoscale Eta model, *J. Geophys. Res.*, (submitted) 2003.

Gibson, J. K., P. Kallberg, S. Uppala, A. Hernandez, A. Nomura, and E. Serrano, *ERA description*, ECMWF Reanal. Proj. Rep. Ser., 72 pp., Eur. Cent. for Medium-Range Weather Forecasts, Reading, UK, 1997.

Gutman, G., and A. Ignatov, The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models, *Int. J. Remote Sensing*, *19*, 1533-1543, 1998.

Hansen, M. C., R. S. DeFries, J. R. G. Townshend, and R. Sohlberg, Global land cover classification at 1 km spatial resolution using a classification tree approach, *International Journal of Remote Sensing*, *21*, 1331-1364, 2000.

Hollinger, S. E., and S. A. Isard, A soil moisture climatology of Illinois, *J. Clim.*, *7*, 822-833, 1994.

Jarvis, P. G., The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field, *Phil. Trans. Roy. Soc. Lond. B.*, *273*, 593-610, 1976.

Ji, M., A. Kumar, and A. Leetmaa, A multiseason climate forecast system at the National Meteorological Center, *Bull. Amer. Meteor. Soc., 75*, 569-577, 1994.

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, B. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, R. Jenne, D. Joseph, The NCEP/NCAR 40-year Reanalysis Project, *Bull. of the Amer. Meteor. Soc., 77*, 437–472, 1996.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. Hnilo, M. Fiorino, and G. Potter, NCEP-DOE AMIP-II Reanalysis (R-2), Bull. Amer. Meteor. Soc., 83, 1631-1643, 2002.

Koren, V., J. Schaake, K. Mitchell, Q. Duan, F. Chen, and J. Baker, A parameterization of snowpack and frozen ground intended for NCEP weather and climate models, *J. Geophys. Res*., *104*, 19,569-19,585, 1999.

Koren, V.I., M. Smith, D. Wang, and Z. Zhang, Use of soil property data in the derivation of conceptual rainfall-runoff model parameters, Preprints, 15th Conference on Hydrology, Long Beach, CA, Amer. Meteor. Soc., Paper 2.16, 10-14 January 2000.

Koster, R., and P. Milly, The interplay between transpiration and run-off formulations in land-surface schemes used with atmospheric models, *J. Clim.*, *10*, 1578-1591, 1997.

Koster, R., and M. Suarez, The components of a SVAT scheme and their effects on a GCM's hydrological cycle, *Adv. Water Resour.*, *17*, 61-78, 1994.

Koster, R., and M. Suarez, Energy and water balance calculations in the Mosaic LSM, NASA Technical Memorandum 104606, Vol. 9, 60pp, 1996.

Koster, R., M. Suarez, and M. Heiser, Variance and predictability of precipitation at seasonal-to-interannual timescales. *J. Hydrometeorology*, *1*, 26-46, 2000.

Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges, A simple hydrologically based model of land surface water and energy fluxes for GCMs, *J. Geophys Res.*, *99*, 14,415-14,428, 1994.

Liang, X., D. P. Lettenmaier, and E. F. Wood, One-dimensional statistical dynamic representation of subgrid spatial variability of precipitation in the two-layer variable infiltration capacity model, *J. Geophys. Res.*, *101*, 21,403-21,422, 1996a.

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

Lohmann, D., D. P. Lettenmaier, X. Liang, E. F. Wood, A. Boone, S. Chang, F. Chen, Y. Dai, C. Desborough, R. E. Dickinson, Q. Duan, M. Ek, Y. M. Gusev, F. Habets, P. Irannejad, R. Koster, K. E. Mitchell, O. N. Nasonova, J. Noilhan, J. Schaake, A. Schlosser, Y. Shao, A. B. Shmakin, D. Verseghy, K. Warrach, P. Wetzel, Y. Xue, Z.-L. Yang, Q.-C. Zeng, The project for intercomparison of land-surface parameterization schemes (PILPS) phase 2(c) Red--Arkansas River basin experiment: 3. Spatial and temporal analysis of water fluxes, *Glob. Planet. Change*, *19*, 161-179, 1998.

Lohmann, D., K. E. Mitchell, P. R. Houser, E. F. Wood, J. C. Schaake, A. Robock, B. A. Cosgrove, J. Sheffield, Q. Duan, L. Luo, R. W. Higgins, R. T. Pinker, J. D. Tarpley, Streamflow and water balance intercomparison of four land-surface models in the North American Land Data Assimilation System (NLDAS), *J. Geophys. Res.*, (submitted) 2003.

Luo, L., A. Robock, K. E. Mitchell, P. R. Houser, E. F. Wood, J. C. Schaake, D. Lohmann, B. A. Cosgrove, J. Sheffield, Q. Duan, R. W. Higgins, R. T. Pinker, J. D. Tarpley, Validation of the North American Land Data Assimilation System (NLDAS) retrospective forcing over the Southern Great Plains, *J. Geophys. Res.*, (submitted) 2003.

Maurer, E. P., A. W. Wood, J. C. Adam, D. P. Lettenmaier, B. Nijssen, A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States, *J. Climate*, *15*, 3237–3251, 2002.

Miller, D.A. and R. A. White, A conterminous United States multi-layer soil characteristics data set for regional climate and hydrology modeling, *Earth Interactions*, 2. Web-based publication, 1998. (Available at http://EarthInteractions.org)

Nijssen, B., D. P. Lettenmaier, X. Liang, S. W. Wetzel, and E. F. Wood, Streamflow simulation for continental-scale river basins, *Water Resour. Res.*, *4*, 711-724, 1997.

Nijssen, B., R. Schnur, and D. P. Lettenmaier, Global retrospective estimation of soil moisture using the variable infiltration capacity land surface model, 1980-93, *J. Clim.*, *14*, 1790-1808, 2001.

Nijssen, B., L.C. Bowling, D.P. Lettenmaier, D. Clark, M.E. Maayar, R. Essery, S. Goers, F. Habets, B. van der Hurk, J. Jin, D. Kahan, D. Lohmann, S. Mahanama, D. Mocko, O. Nasonova, G.-Y. Niu, P. Samuelsson, A.B. Shmakin, K. Takata, D. Verseghy, P. Viterbo, X. Ma, Y. Xia, Y. Xue, Z.-L. Yang, Simulation of high latitude hydrological processes in the Torne-Kalix basin: PILPS Phase 2e. 2: Comparison of model results with observations, *Global and Planetary Change* (in press*)*, 2003.

Oki, T., T. Nishimura, and P. Dirmeyer, Assessment of annual runoff from land surface models using total runoff integrating pathways (TRIP), *J. Meteor. Soc. Japan*, *77*, 69-89, 1999.

Pan, H.-L., and L. Mahrt, Interaction between soil hydrology and boundary layer development, *Boundary-Layer Meteorol.*, *38*, 185-202, 1987.

Pan, M., J. Sheffield, E. F. Wood, K. E. Mitchell, P. R. Houser, J. C. Schaake, A. Robock, D. Lohmann, B. A. Cosgrove, Q. Duan, L. Luo, R. W. Higgins, R. T. Pinker, J. D. Tarpley, Snow process modeling in the North American Land Data Assimilation System (NLDAS). Part II: Evaluation of model simulated snow water equivalent, *J. Geophys. Res.*, (submitted) 2003.

Pinker, R. T., J. D. Tarpley, I. Laszlo, K. E. Mitchell, P. R. Houser, E. F. Wood, J. C. Schaake, A. Robock, D. Lohmann, B. A. Cosgrove, J. Sheffield, Q. Duan, L. Luo, R. W. Higgins, 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.*, (submitted) 2003.

Ramsay, B. H., The interactive multisensor snow and ice mapping system, *Hydrol. Process., 12*, 1537-1546, 1998.

Reynolds, C. A., T. J. Jackson, and W. J. Rawls, Estimating soil water-holding capacities by linking the Food and Agriculture Organization soil map of the world with global pedon databases and continuous pedotransfer functions, *Water Resources Research, 36(12),* 3653-3662, 2000.

Roads, J., and co-authors, the GCIP Water and Energy Budget Synthesis (WEBS), submitted to *J. Geophys. Res*., (in press) 2003.

Robock, A., L. Luo, E. F. Wood, F. Wen, K. E. Mitchell, P. R. Houser, J. C. Schaake, D. Lohmann, B. A. Cosgrove, J. Sheffield, Q. Duan, R. W. Higgins, R. T. Pinker, J. D. Tarpley, J. B. Basara, and K. C. Crawford, Evaluation of the North American Land Data Assimilation (NLDAS) over the Southern Great Plains during the warm season, *J. Geophys. Res.*, (submitted) 2003.

Rodell, M., P. R. Houser, U. Jambor, J. Gottschalck, K. Mitchell, C.-J. Meng, K. Arsenault, K. Arsenault, B. Cosgrove, J. Radakovich, M. Bosilovich, J. K. Entin, J. P. Walker, D. Lohmann, and D. Toll, The global land data assimilation System, Bull. Amer. Meteor. Soc., (in press) 2003.

Rogers, Eric, Deaven, Dennis G., Dimego, Geoffrey S., The Regional Analysis System for the operational "Early" Eta Model: Original 80-km configuration and recent changes. *Wea. and Forecasting*, 10, 810–825, 1995.

Schaake, J., V. Koren, Q. Duan, K. Mitchell, and F. Chen, Simple water balance model for estimating runoff at different spatial and temporal scale, *J. Geophys. Res*., *101*, 7461-7475, 1996.

Schaake, J. C., Q. Duan, K. E. Mitchell, P. R. Houser, E. F. Wood, A. Robock, D. Lohmann, B. A. Cosgrove, J. Sheffield, L. Luo, R. W. Higgins, R. T. Pinker, J. D. Tarpley, An intercomparison of soil moisture fields in the North American Land Data Assimilation System (NLDAS), *J. Geophys. Res.*, (submitted) 2003.

Schlosser, C. A., A. G. Slater, A. Robock, A. J. Pitman, J., K. Y. Vinnikov, A. Henderson-Sellers, N. A. Speranskaya, K. Mitchell, PILPS 2(D) Contributors, Simulations of a boreal grassland hydrology at Valdai, Russia: PILPS Phase 2(d). *Mon. Wea. Rev*., 128, 301–321, 2000.

Sellers, P., Y. Mintz, Y. Sud, and A. Dalcher, A simple biosphere model (SiB) for use within general circulation models, *J. Atmos. Sci.*, *43*, 505-531, 1986.

Sheffield, J., M. Pan, E. F. Wood, K. E. Mitchell, P. R. Houser, J. C. Schaake, A. Robock, D. Lohmann, B. A. Cosgrove, Q. Duan, L. Luo, R. W. Higgins, R. T. Pinker, J. D. Tarpley, B. H. Ramsay, Snow process modeling in the North American Land Data Assimilation System (NLDAS). Part I: Evaluation of model simulated snow cover extent, *J. Geophys. Res.*, (submitted) 2003.

Shi, W., R. W. Higgins, E. Yarosh, A unified Raingauge dataset and multi-year daily precipitation reanalysis for the United States, *J. Geophys. Res.*, (submitted) 2003.

Tarpley, J., Estimating incident solar radiation at the surface from geostationary satellite data, *J. Appl. Meteorol.*, *18*, 1172-1181, 1979.

Tarpley, J. D., Monthly evapotranspiration from satellite and conventional meteorological observations, *J. Climate, 7*, 704-713, 1994.

Verdin, K. L., and S. K. Greenlee, Development of continental scale digital elevation models and extraction of hydrographic features, *Proceedings, Third International Conference/Workshop on Integrating GIS and Environmental Modeling,* Santa Fe, NM, January 21-26, National Center for Geographic Information and Analysis, Santa Barbara, CA, 1996.

Vogel, Christoph A., Baldocchi, Dennis D., Luhar, Ashok K., Shankar Rao, K., A comparison of a hierarchy of models for determining energy balance components over vegetation canopies. J. of Appl. Meteor., 34, 10, 2182–2196, 1995.

Wigmosta, M. S., L. W. Vail, and D. P. Lettenmaier, A distributed hydrology – vegetation model for complex terrain, *Water Resour. Res.*, *30*, 1665-1679, 1994.

Wood, E. F., D. P. Lettenmaier, X. Liang, B. Nijssen, and S. W. Wetzel, Hydrological modeling of continental-scale basins, *Ann. Rev. Earth Planet Sci.*, *25*, 279-300, 1997.

Wood, E., D. Lettenmaier, X . Liang, D. Lohmann, A. Boone, S. Chang, F. Chen, Y. Dai, R. Dickinson, Q. Duan, M. Ek, Y. Gusev, F. Habets, P. Irannejad, R. Koster, K. Mitchell, O. Nasonova, J. Noilhan, J. Schaake, A. Schlosser, Y. Shao, A. Shmakin, D. Verseghy, K. Warrach, P. Wetzel, Y. Xue, Z. Yang, and Q. Zeng, The Project for Intercomparison of Land-surface Parameterization Schemes (PILPS) Phase 2(c) Red-Arkansas River basin experiment: 1. Experimental description and summary intercomparisons, *Glob. Planet. Change*, *19*, 115-136, 1998.

Wu, X., W. P. Menzel, and G. S. Wade, Estimation of sea surface temperatures using GOES-8/9 radiance measurements. *Bull. Amer. Meteor. Soc.*, *80*, 1127-1138, 1999.

Zhao, R.-J., L.-R. Fang, X.-R. Liu, and Q.-S. Zhang, The Xinanjiang model, in *Hydrological Forecasting Proc.*, Oxford Symposium, IASH Pub. 129, 351-356, 1980.

Zilitinkevich, S. S., Non-local turbulent transport: Pollution dispersion aspects of coherent structure of convective flows, in H. Power, N. Moussiopoulos, and C. A. Brebbia (Eds.), *Air Pollution III – Volume I. Air Pollution Theory and Simulation*, Computational Mechanics Publications, Southampton, Boston, 52-60, 1995.

Figure Captions

Figure 1. Depiction of NLDAS domain, showing (a) mean July green vegetation fraction from *Gutman and Ignatov* [1998] and (b) mean annual NLDAS precipitation (mm) for the period October 1, 1997 to September 30, 1999.

Figure 2. Daily precipitation (mm) in (a) NLDAS and (b) EDAS for the 24-hour period ending 12 UTC on July 23, 1998.

Figure 3. Monthly mean diurnal cycle of downward surface solar insolation (Wm^{-2}) from EDAS (solid circles), GOES-based retrieval (solid circles) and SURFRAD flux station observations (solid line) for July (left) and January (right) 1998 at Bondville, Illinois.

Figure 4. a) Time series of daily streamflow (m^3s^{-1}) from observations (black) and the Noah (blue), VIC (green), Mosaic (red), and SAC (gold) models for East Fork basin of White River at Columbus, IN during October 1, 1997 to September 30, 1999. b) Unit hydrograph (see text) of each model. Upper left: bias (m^3s^{-1}) and correlation of model streamflow relative to observed.

Figure 5. Mean annual evaporation (mm/year) over the NLDAS domain from a) Noah, b) VIC, c) Mosaic and d) SAC for the period October 1, 1997 to September 30, 1999.

Figure 6. Mean annual runoff (mm/year) over the NLDAS domain from a) Noah, b) VIC, c) Mosaic and d) SAC for the period of October 1, 1997 to September 30, 1999.

Figure 7. a) Partitioning of area-mean mean annual precipitation (diagonal, mm/year) into area-mean mean annual runoff (x-axis, mm/year) and evaporation (y-axis, mm/year) for the CONUS quadrants (inset) of SE (green), NE (red), NW (black) and SW (blue) by Noah (N), VIC (V), Mosaic (M), and SAC (S) for the period October 1, 1997 to September 30, 1999. Model symbols below the diagonal indicate positive storage change for the time period. b) As in a), except area-mean for each quadrant is for quadrant sub-area represented by basins depicted in the inset and the crosses depict the observed area-mean mean annual runoff from basins in c) for the same sub-area. c) Observed area-mean mean annual runoff (mm/year) for 1145 small basins in the NLDAS domain for the period in a), calculated from observed streamflow provided by USGS (via http://www.usgs.gov).

Figure 8. Relative bias [(model-observed)/observed] of mean annual model runoff for selected basins for the period October 1,1997 to September 30, 1999 for a) Noah, b) VIC, c) Mosaic, and d) SAC.

Figure 9. Time series of area-averaged monthly evaporation (mm/month) of Noah (blue), VIC (green) Mosaic (red), and SAC (gold) over the four CONUS quadrants of NW (a), NE (b), SW (c) and SE (d) for the period October 1996 to September 1999.

Figure 10. Time series of area-averaged monthly mean TOTAL COLUMN soil moisture (mm) Noah (blue), VIC (green), Mosaic (red), and SAC (gold) over the three CONUS quadrants of NW (a), NE (b), and SE (c) for the period October 1996 to September 1999. Panel d also for SE as in Panel c, except for ROOT ZONE soil moisture.

Figure 11. Difference in total column soil moisture (mm) of 30 April minus 30 September of 1999 at 23 UTC for Noah (a), VIC (b), Mosaic (c), Sacramento (d). All colors (except dark red) are positive and denote a net drying during the period.

Figure 12. Monthly water budget for October 1997 to September 1998. In order, columns 1-4 are Noah, VIC, Mosaic, and SAC, and rows 1-4 are quadrants SE, NE, NW, and SW. Colors depict terms (mm/month) in (1): dS_1/dt (orange), dS_2/dt (red), R_1 (light blue), R_2 (dark blue), E (green). Solid black line = total precipitation P (mm/month) and dotted black line = liquid precipitation (mm/month). See text for definition of red triangles.

Figure 13. Comparison of model versus observed bi-monthly soil moisture (mm) in top 2 m, averaged over 17 sites throughout Illinois for the time period October 1997 to September 1999 for Noah (a), VIC (b), Mosaic (c), and Sacramento (d).

Figure 14. a) For the top 0-40 cm soil layer during the period January 1, 1998 to September 30, 1999, the time series of daily multi-station mean volumetric soil moisture of observations (black) and Noah (blue), VIC (green), Mosaic (red) and SAC (gold) model and over the Oklahoma Mesonet, for a) absolute value and b) departure from time average (over entire period) of corresponding time series in a).

Figure 15. Comparison of mean annual maximum snow water-equivalent, *SWE* (mm), during the period October 1996 to September 1999 between observations (x-axis) and model simulations (y-axis) at 110 SNOTEL sites for the control runs of SAC (a), Noah (b), VIC (c) and

Mosaic (d), and two VIC tests runs forced with regionally adjusted precipitation (e) and locally adjusted precipitation (f).

Figure 16. Time series of daily area mean snow cover extent (*SCE*) for IMS observed data (black) and the Mosaic (red), Noah (blue), VIC (green) and SAC (gold) models for the Northwest (a), North-central (b), Northeast (c), and Colorado (d) RFC regions for the time period February 1, 1997 to September 30, 1999.

Figure 17. Time series of monthly area mean a) snowmelt (mm/month), b) snow sublimation (mm/month) and c) albedo for Noah (blue), VIC (green), Mosaic (red), and SAC (gold) for the time period October 1996 to September 1999 for the Northwest RFC. (Note different y-axis range in top panels.)

Figure 18. Time series of monthly mean surface energy fluxes (Wm^{-2}) of (a) net radiation *R*, (b) latent heat *LE*, (c) ground heat *G* and (d) sensible heat *H* averaged over the ARM/CART sites during January 1998 to September 1999 from observations (black) and Noah (blue), VIC (green), and Mosaic (red). Note different y-axis range among panels. For *R*, *LE* and *H*, a vertically upward flux is positive, representing a heat source to the atmosphere and sink to the surface. For G, a positive flux is a heat source to the surface.

Figure 19. Monthly mean diurnal cycle in Wm^{-2} of net radiation *R* (row 1), latent heat *LE* (row 2), sensible heat *H* (row 3) and ground heat *G* (row 4) for July (columns 1,3) and April (columns 2,4) of 1999. Columns 1-2 show observations (black) and control runs for Noah (blue), VIC

(green) and Mosaic (red). Columns 3-4 give control run (solid line) versus test run (dashed line) for Noah (blue) and VIC (green). Main text describes VIC and Noah tests. Note different y-axis range between rows. Sign convention as in Figure 18 (except for *G*).

Figure 20. Monthly mean diurnal cycle of *LST* (K) averaged over all ARM/CART SIRS sites for July (left) and April (right) 1999 from the observations (solid black), the control runs (solid colors) for Noah (blue), VIC (green) and Mosaic (red), and a Noah test run (dashed blue) using a modified treatment for roughness length for heat in Noah's aerodynamic resistance.

Figure 21. Monthly mean diurnal cycle of aerodynamic conductance (ms^{-1}) averaged over all ARM/CART SIRS sites for July 1998 for the control runs (solid colors) of Noah (blue), VIC (green) and Mosaic (red), and a test run of Noah (dashed black) using a modified treatment for the roughness length for heat.

Figure 22. Monthly mean diurnal cycle of *LST* (K) averaged over all ARM/CART SIRS sites for July (left) and April (right) during 1999 (top) and 1998 (bottom) from SIRS observations (solid) and GOES-East retrieval (dashed).

Figure 23. Comparison of model (y-axis) versus observed (x-axis) *LST* (K) at 18 UTC over all ARM/CART SIRS sites for July (columns 1-2) and April (columns 3-4) 1999 for Mosaic (top), Noah (middle) and VIC (bottom) versus SIRS observations (columns 1,3) and GOES-East observations (columns 2,4). Match-up point included only if GOES *LST* available (cloud free), yielding sample size of 334 in columns 1-2 and 198 in columns 3-4.

Figure 24. Comparison of model (y-axis) versus GOES-East (x-axis) *LST* (K) at 18 UTC over the northern Midwest during July (columns 1-2) and April (columns 3-4) 1999 for Mosaic (top), Noah (middle) and VIC (bottom) versus unscreened (columns 1, 3) and screened (columns 2,4) GOES *LST*. See text for color scale definition and latitude/longitude range of region.

Table 1. GCIP-supported products applied by the NLDAS project.

A) <u>FOR PRODUCING NLDAS SURFACE FORCING</u>

1) Daily, gage-only 1/8° CONUS precipitation analysis by NCEP
2) Hourly, radar-dominated, 4-km CONUS "Stage IV" precipitation analysis by OHD and NCEP
3) Hourly, GOES 1/2° CONUS surface insolation by NESDIS/U. Maryland
4) 3-hourly 40-km Eta-based 4DDA analyses of near-surface meteorology by NCEP, and upgrades therein to land-atmosphere interactions and coupling via Noah LSM

B) <u>FOR VALIDATING NLDAS SURFACE FORCING</u>

5) OU Mesonet surface meteorology observations
6) SURFRAD network surface solar insolation observations (receives support from GCIP sister program in NOAA/OGP)

C) <u>FOR VALIDATING NLDAS LAND MODEL OUTPUT</u> (states/fluxes)

7) Oklahoma/Kansas ARM/CART surface flux stations (DOE)
8) Oklahoma Mesonet soil moisture/temperature observations (OU Climate Survey)
9) CONUS-wide GOES-based satellite skin temperatures (NESDIS and U. Maryland)
10) N. Hemisphere 23-km IMS daily snow cover analysis by NESDIS
11) Illinois Water Survey network of 18 soil moisture measuring stations
12) Western U.S. network of SNOTEL observations (not GCIP supported)
13) USGS streamflow observations (not GCIP supported)

D) <u>FOR NLDAS LAND SURFACE CHARACTERISTICS</u>

14) 1-km CONUS soil texture database by Pennsylvania State U.
15) NESDIS 0.144-degree global monthly NDVI-based vegetation greenness by NESDIS

E) <u>IMPROVED LAND/HYDROLOGY MODELS</u>

16) Noah model improvements by NCEP and OHD (and collaborators)
17) VIC model improvements by Princeton U. and U. Washington (and collaborators)
18) SAC model improvements by OHD and NCEP (and collaborators)
19) Mosaic model improvements by NASA GSFC
20) Streamflow connectivity network and routing model by NCEP, U. Washington, Princeton U., and OHD

Table 2. Content of hourly land surface forcing files in NLDAS.

Content						
	EDAS	GOES	Gage	Radar	realtime	retrospective
Primary forcing						
2-m air temperature (K)	X				X	X
2-m air specific humidity (kg/kg)	X				X	X
10-m u-wind component (m/s)	X				X	X
10-m v-wind component (m/s)	X				X	X
surface pressure (mb)	X				X	X
downward longwave radiation (W/m^2)	X				X	X
downward shortwave radiation (W/m^2)		X			X	X
total precipitation (kg/m^2)			X[1]	X[1]	X	X
Backup forcing						
downward shortwave radiation (W/m^2)	X				X	X
total precipitation (kg/m^2)	X				X	X
Auxiliary forcing						
total precipitation: WSR-88D (kg/m^2)			X[2]	X[2]		
photosynthetically active rad. (W/m^2)		X[3]			X	X
convective precipitation (kg/m^2)	X				X	X
CAPE	X				X	
For validation (+future assimilation)						
LST: land surface skin temperature (K)		X			X	X

1 Daily total is gage-only. Radar estimate used only to temporally partition daily into hourly.
2 Radar-dominated estimate (some gage data used), known as "Stage II/III" (see text).
3 Often denoted as PAR. Subtracting PAR from total shortwave radiation gives diffuse radiation.

Table 3. Primary attributes of the four NLDAS land models.

	Mosaic	Noah	VIC	SAC
Full domain runs	Yes	Yes	Yes	Yes
Limited domain runs	Yes	Yes	Yes	Yes
Input surface forcing	Precipitation, downward solar and longwave radiation, 10-m wind speed, surface pressure, 2-m air temperature, 2-m air humidity			precipitation, Noah *PE*, 2-m air temp.
Energy balance	Yes	Yes	Yes	N/A
Water balance	Yes	Yes	Yes	Yes
Model time step	15 min	15 min	1 hour	1 hour
Model soil layers Model geometry	3 10, 30, 160cm	4 10, 30, 60, 100cm	3 10 cm, variable	2 equivalent (6 storages)
Tiling: Vegetation Tiling: Elevation	Yes No	No No	Yes Yes	No No
Snow model layers	1	1	2	1
Frozen soil-thermal -hydraulics	No Simplified	Yes Yes	Disabled Disabled	N/A
Soil thermodynamics	Force-restore	Heat conduction Eqn	Heat conduction Eqn. modified	No
Soil temperature profile	No	Yes	Yes	No
Soil water -drainage - vertical diffusion	Yes Yes	Yes Yes	Yes No	Yes No
Snow-free Albedo	Vary wrt LAI, GVF, biome	Monthly input background field	Vary wrt LAI and biome	N/A
Diurnal Albedo	Yes	No	No	N/A
Explicit vegetation	Yes	Yes	Yes	No
Canopy resistance	*Sellers* [1986]	*Jarvis* [1976]	*Jarvis* [1976]	N/A
Rooting depth	0.4 m	Variable (1 or 2 m)	Variable (1.35 – 3 m)	N/A
Root density profile	Constant	Constant	Exponential	N/A
Canopy capacity	0-1.6 mm	0.5 mm	0.1-1.0 mm	N/A

Table 4. Bias (K), error standard deviation (STDE) (K), and correlation (CORR) of model LST versus both ARM LST and unscreened GOES LST over the SGP from Figure 23.

Model	July 1999			April 1999		
	GOES / ARM			GOES / ARM		
	BIAS (K)	STDE (K)	CORR	BIAS (K)	STDE (K)	CORR
Noah	+6.5/+5.4	3.8/3.5	0.70/0.64	+2.4/+1.7	2.9/4.1	0.78/0.61
VIC	+2.8/+1.7	3.4/3.3	0.76/0.68	+0.3/-0.5	3.9/4.3	0.66/0.57
Mosaic	-1.2/-2.3	3.7/3.0	0.70/0.72	-4.7/-5.4	3.0/4.0	0.77/0.62

Table 5. Bias (K), error standard deviation (STDE) (K) and correlation (CORR) of model LST versus screened and unscreened GOES LST for north central CONUS from Figure 24.

Model	July 1999			April 1999		
	GOES: Unscreened / Screened			GOES: Unscreened / Screened		
	BIAS (K)	STDE (K)	CORR	BIAS (K)	STDE (K)	CORR
Noah	+9.3/+8.8	3.6/2.6	0.73/.082	+4.5/+3.7	6.5/4.9	0.60/0.63
VIC	+3.9/+3.3	4.2/3.0	0.64/0.76	-0.4/-0.7	6.9/5.3	0.65/0.76
Mosaic	+1.6/+1.1	3.7/2.4	0.69/0.82	-3.9/-4.7	5.8/3.7	0.73/0.81

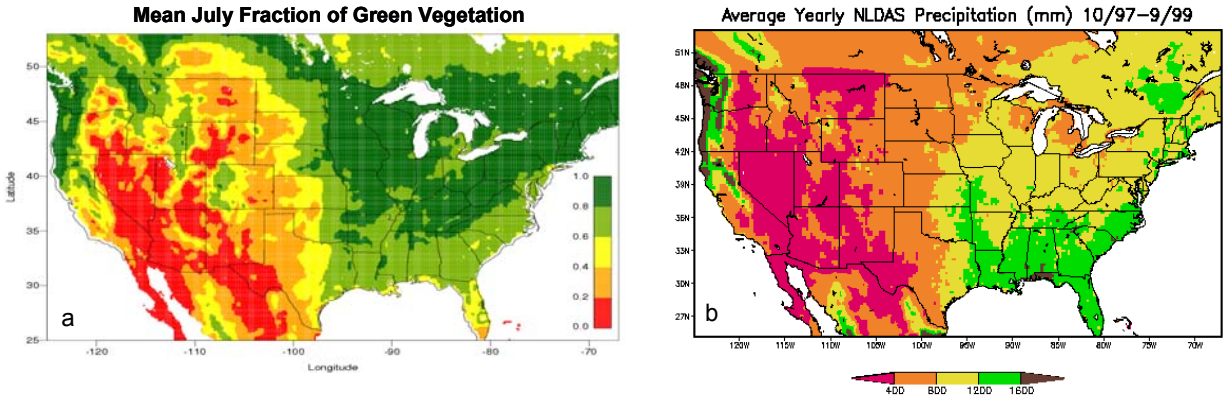

Figure 1. Depiction of NLDAS domain, showing (a) mean July green vegetation fraction from *Gutman and Ignatov* [1998] and (b) mean annual NLDAS precipitation (mm) for the period October 1, 1997 to September 30, 1999.

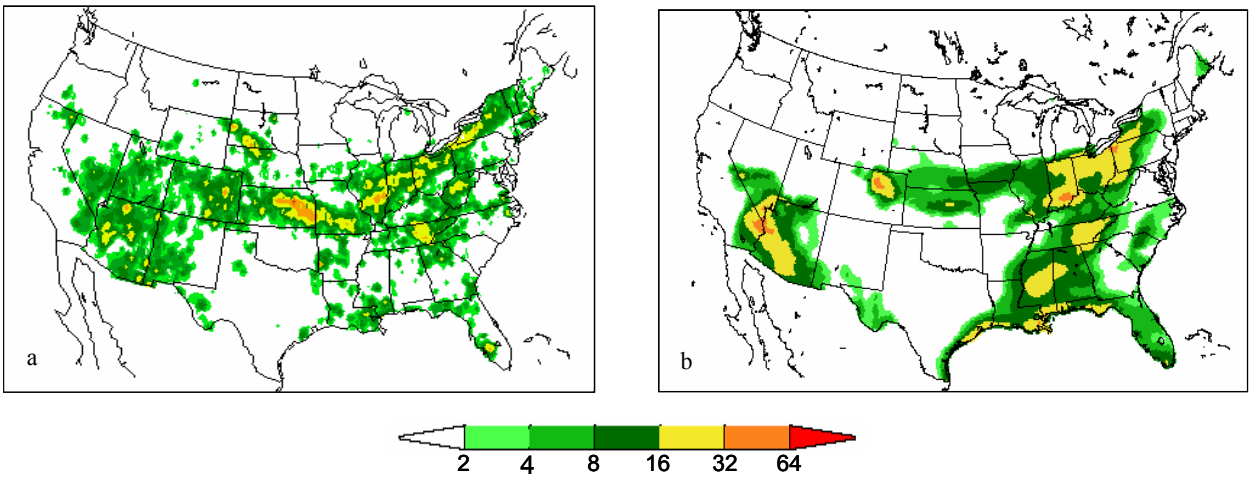

Figure 2. Daily precipitation (mm) in (a) NLDAS and (b) EDAS for the 24-hour period ending 12 UTC on July 23, 1998.

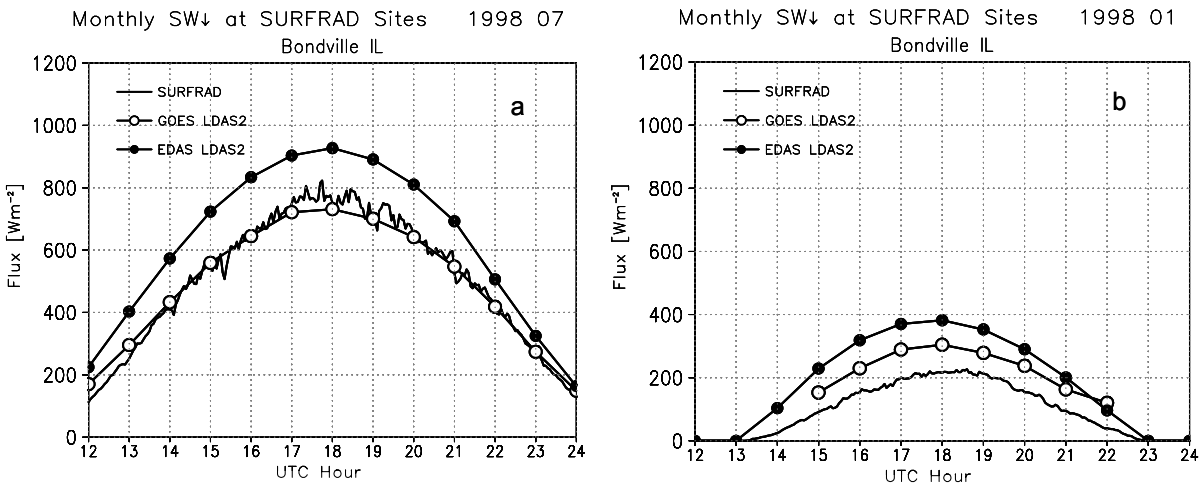

Figure 3. Monthly mean diurnal cycle of downward surface solar insolation (Wm^{-2}) from EDAS (solid circles), GOES-based retrieval (solid circles) and SURFRAD flux station observations (solid line) for July (left) and January (right) 1998 at Bondville, Illinois.

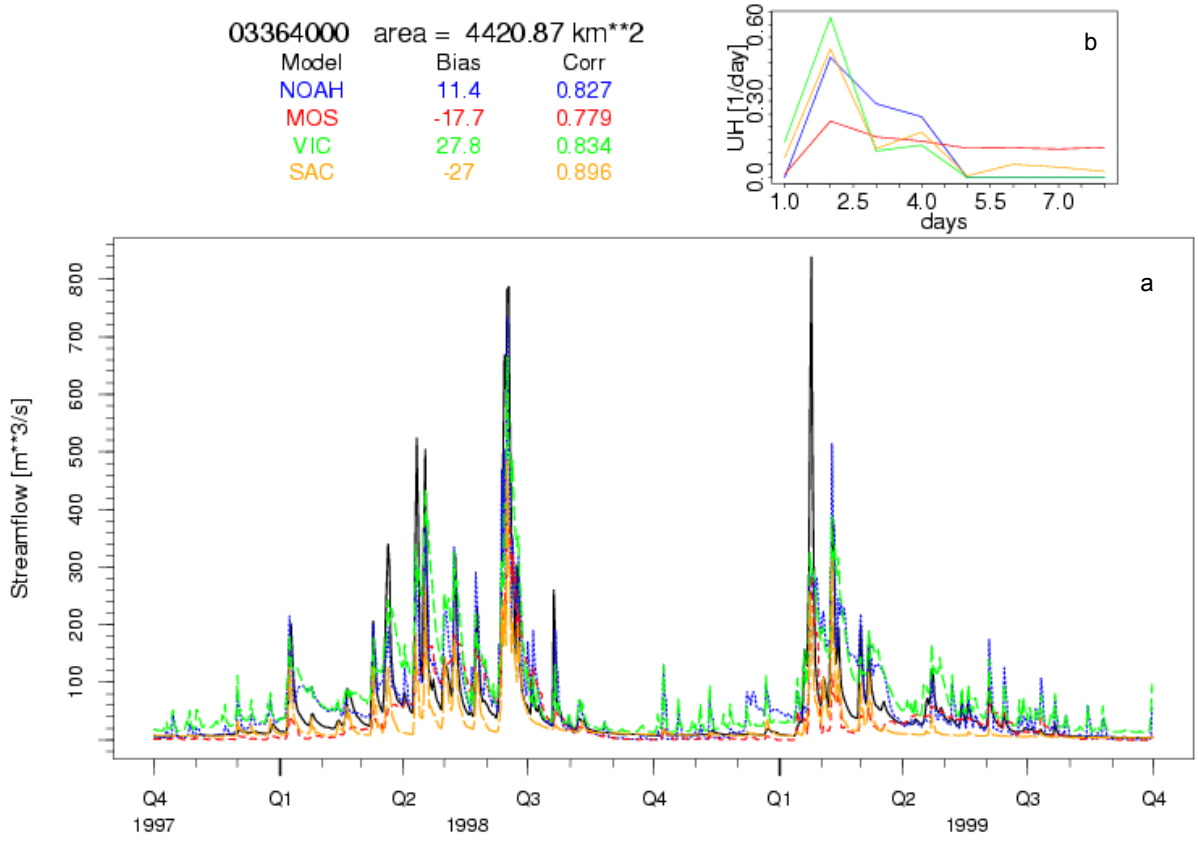

Figure 4. a) Time series of daily streamflow (m^3s^{-1}) from observations (black) and the Noah (blue), VIC (green), Mosaic (red), and SAC (gold) models for East Fork basin of White River at Columbus, IN during October 1, 1997 to September 30, 1999. b) Unit hydrograph (see text) of each model. Upper left: bias (m^3s^{-1}) and correlation of model streamflow relative to observed.

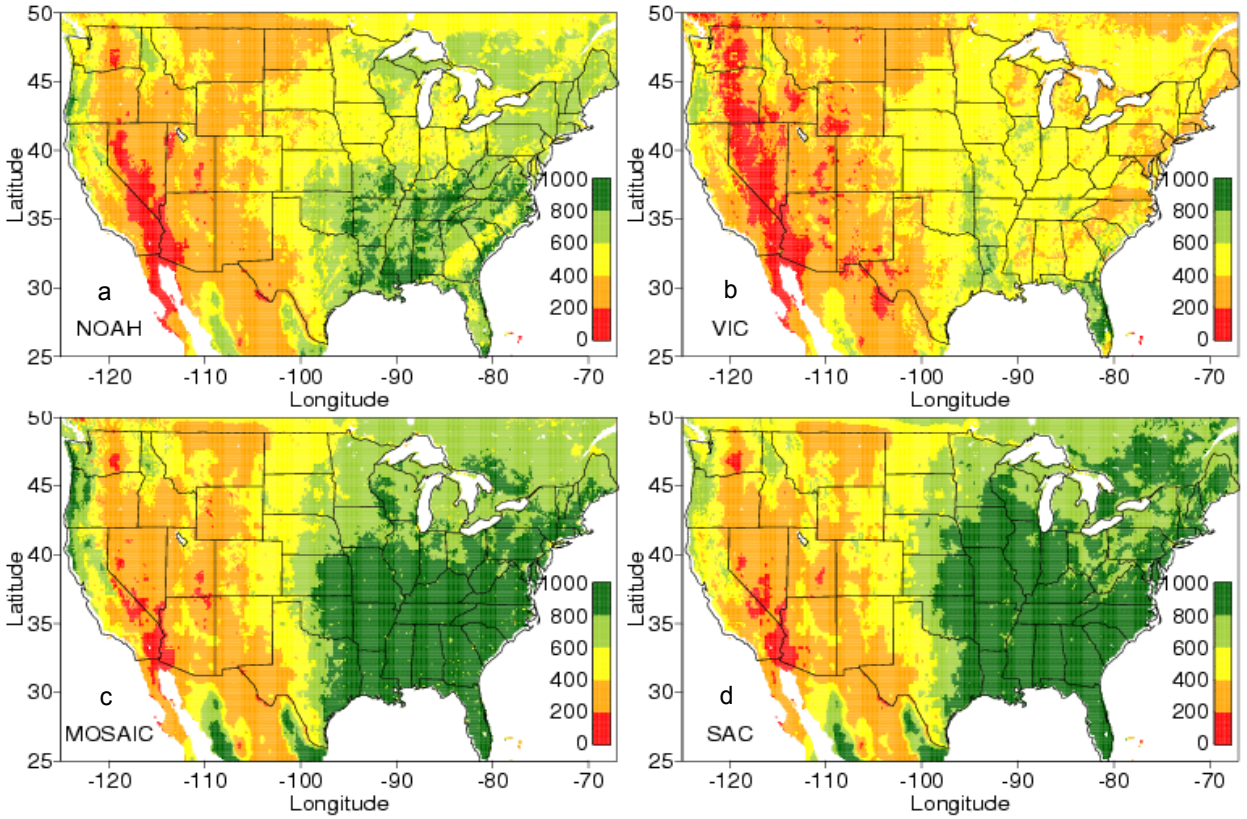

Figure 5. Mean annual evaporation (mm/year) over the NLDAS domain from a) Noah, b) VIC, c) Mosaic and d) SAC for the period October 1, 1997 to September 30, 1999.

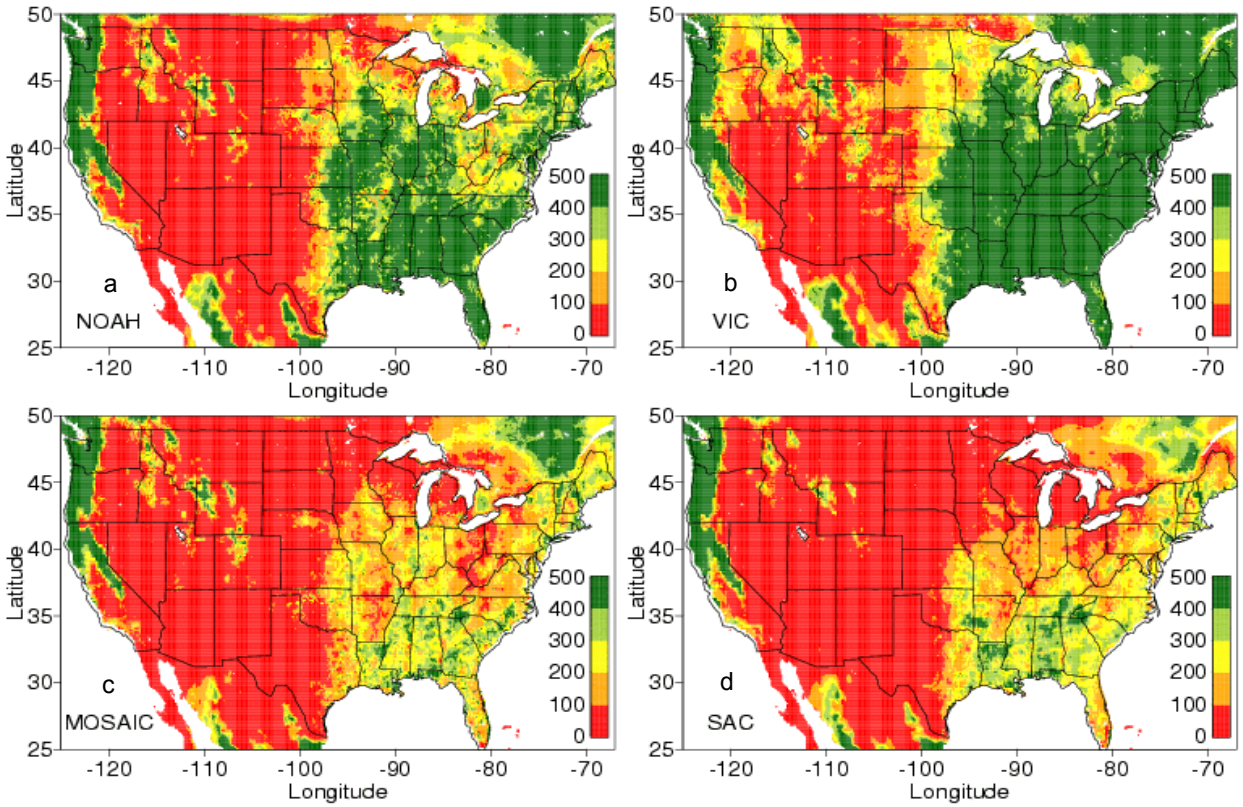

Figure 6. Mean annual runoff (mm/year) over the NLDAS domain from a) Noah, b) VIC, c) Mosaic and d) SAC for the period of October 1, 1997 to September 30, 1999.

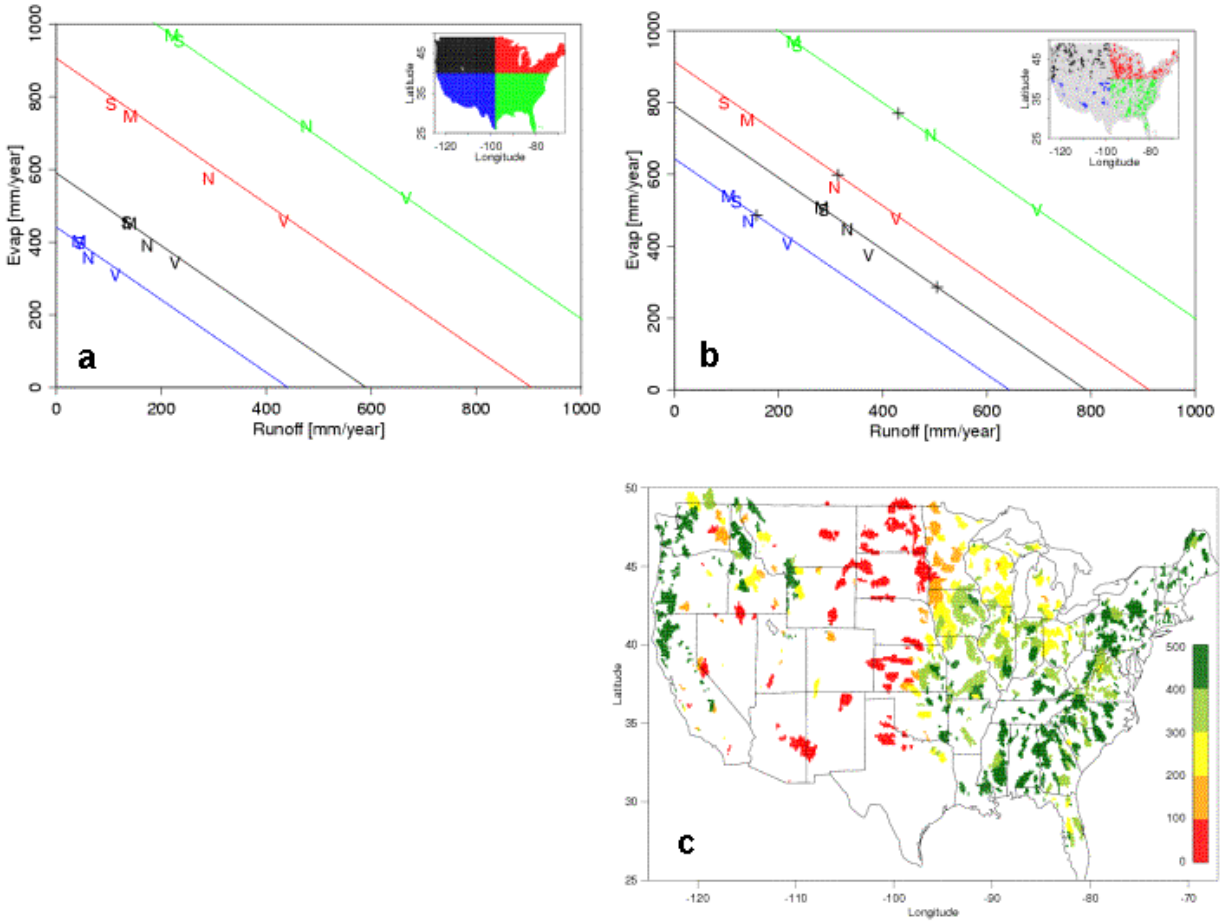

Figure 7. a) Partitioning of area-mean mean annual precipitation (diagonal, mm/year) into area-mean mean annual runoff (x-axis, mm/year) and evaporation (y-axis, mm/year) for the CONUS quadrants (inset) of SE (green), NE (red), NW (black) and SW (blue) by Noah (N), VIC (V), Mosaic (M), and SAC (S) for the period October 1, 1997 to September 30, 1999. Model symbols below the diagonal indicate positive storage change for the time period. b) As in a), except area-mean for each quadrant is for quadrant sub-area represented by basins depicted in the inset and the crosses depict the observed area-mean mean annual runoff from basins in c) for the same sub-area. c) Observed area-mean mean annual runoff (mm/year) for 1145 small basins in the NLDAS domain for the period in a), calculated from observed streamflow provided by USGS (via http://www.usgs.gov).

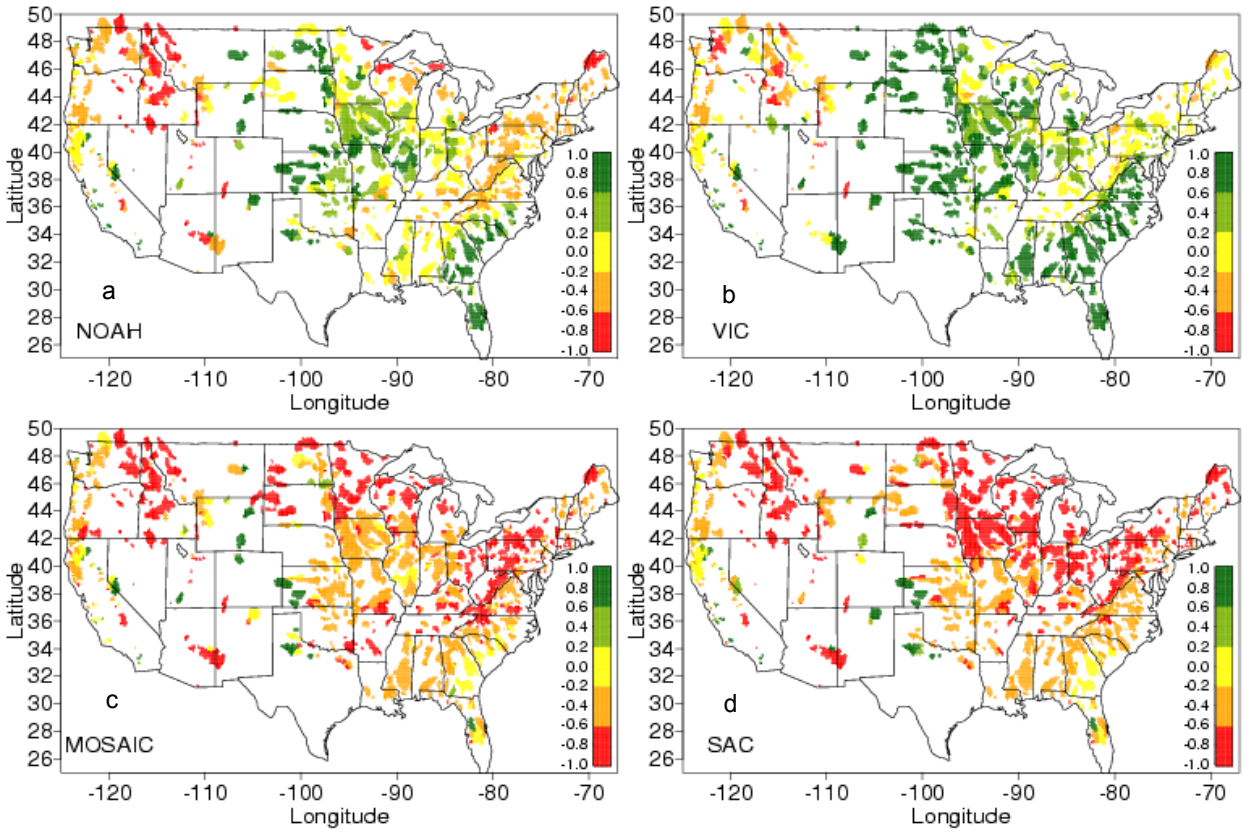

Figure 8. Relative bias [(model-observed)/observed] of mean annual model runoff for selected basins for the period October 1,1997 to September 30, 1999 for a) Noah, b) VIC, c) Mosaic, and d) SAC.

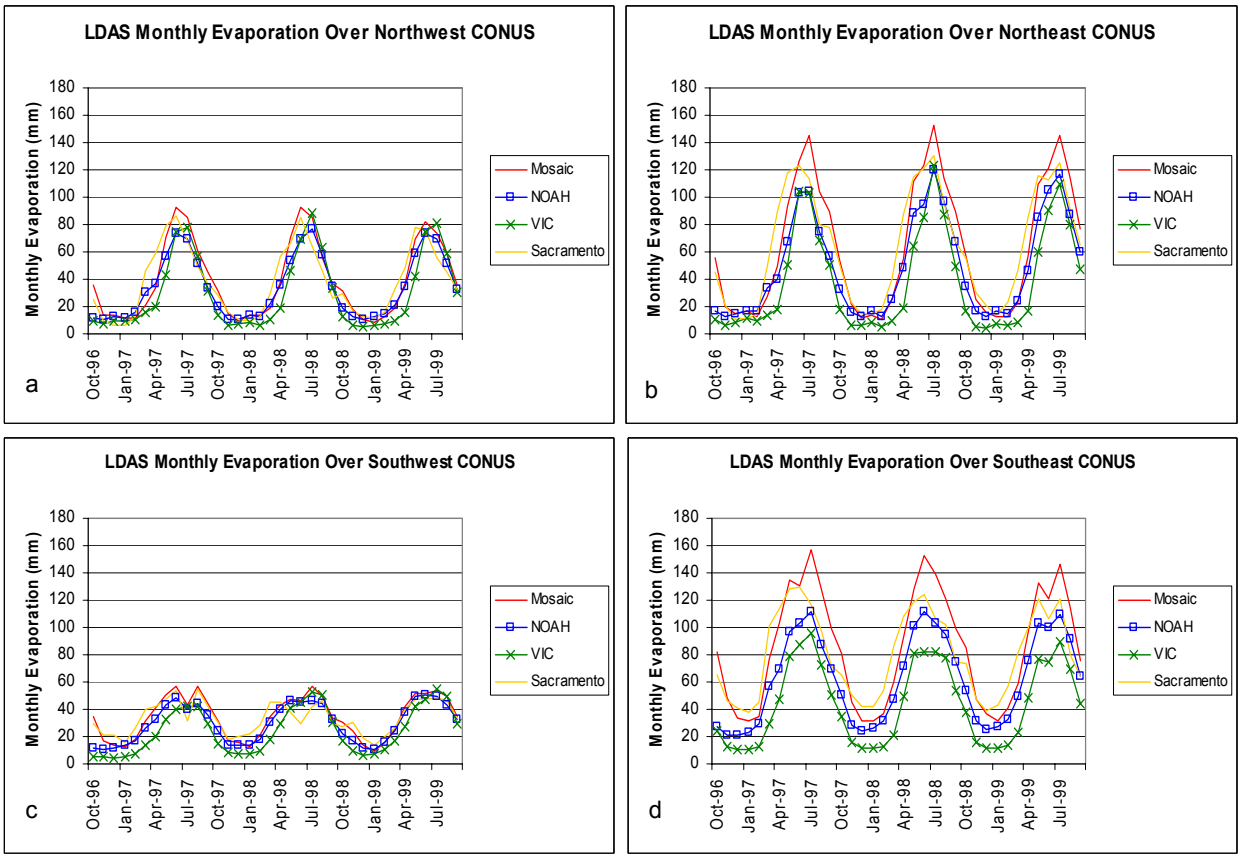

Figure 9. Time series of area-averaged monthly evaporation (mm/month) of Noah (blue), VIC (green) Mosaic (red), and SAC (gold) over the four CONUS quadrants of NW (a), NE (b), SW (c) and SE (d) for the period October 1996 to September 1999.

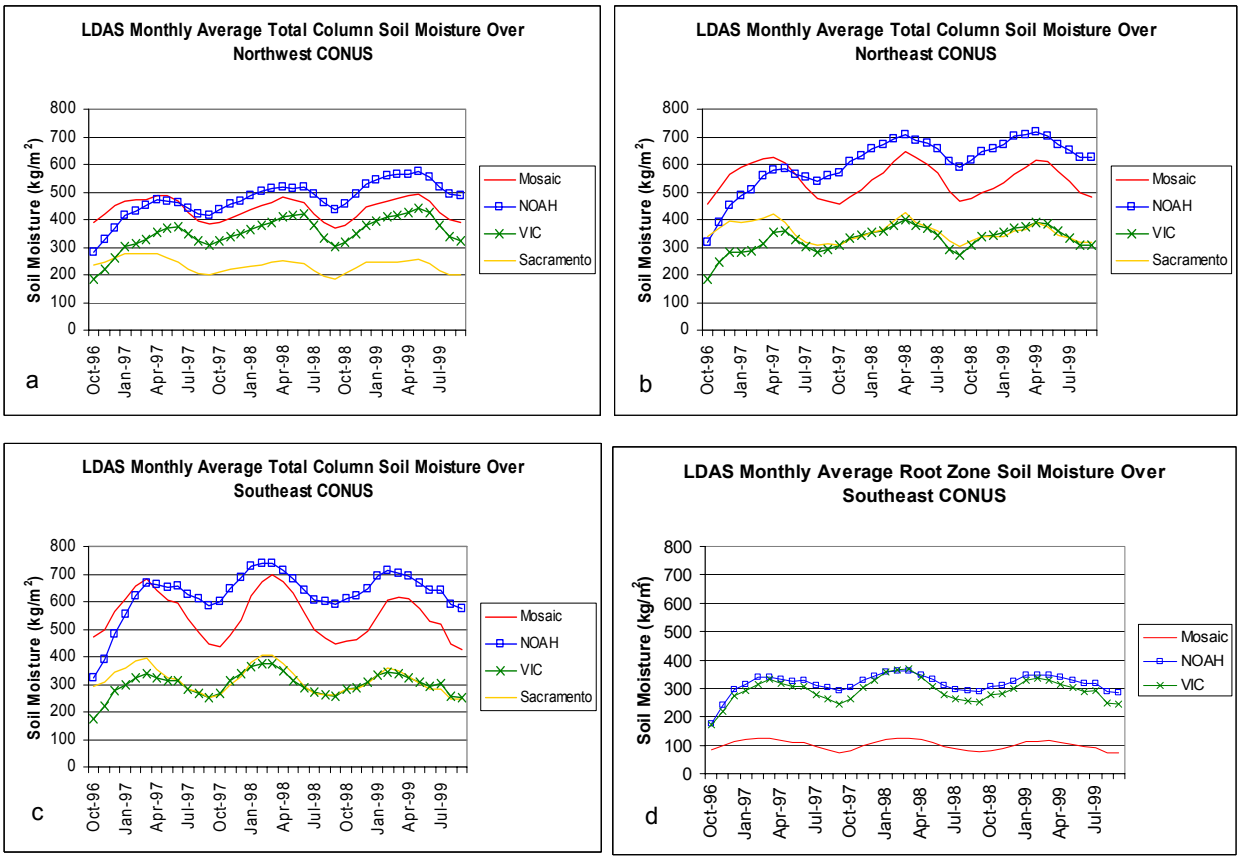

Figure 10. Time series of area-averaged monthly mean TOTAL COLUMN soil moisture (mm) Noah (blue), VIC (green), Mosaic (red), and SAC (gold) over the three CONUS quadrants of NW (a), NE (b), and SE (c) for the period October 1996 to September 1999. Panel d also for SE as in Panel c, except for ROOT ZONE soil moisture.

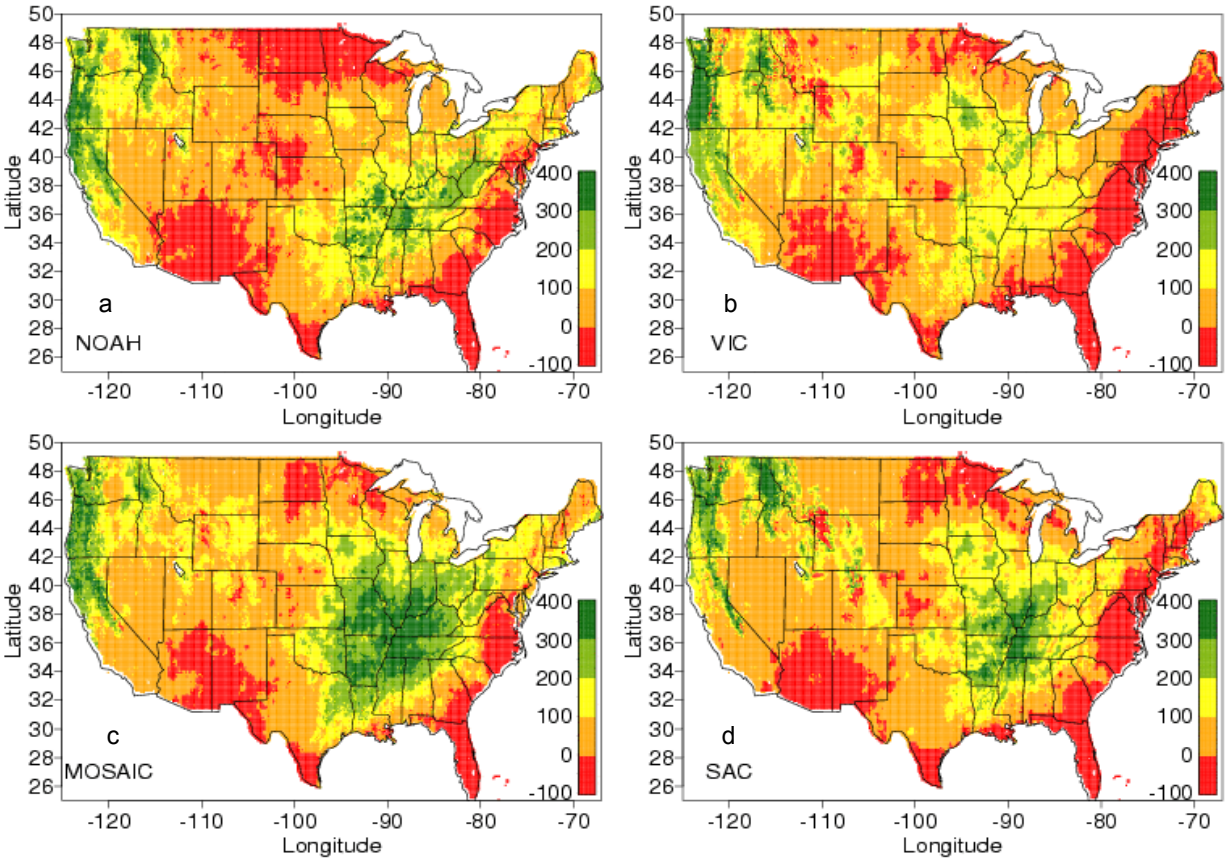

Figure 11. Difference in total column soil moisture (mm) of 30 April minus 30 September of 1999 at 23 UTC for Noah (a), VIC (b), Mosaic (c), Sacramento (d). All colors (except dark red) are positive and denote a net drying during the period.

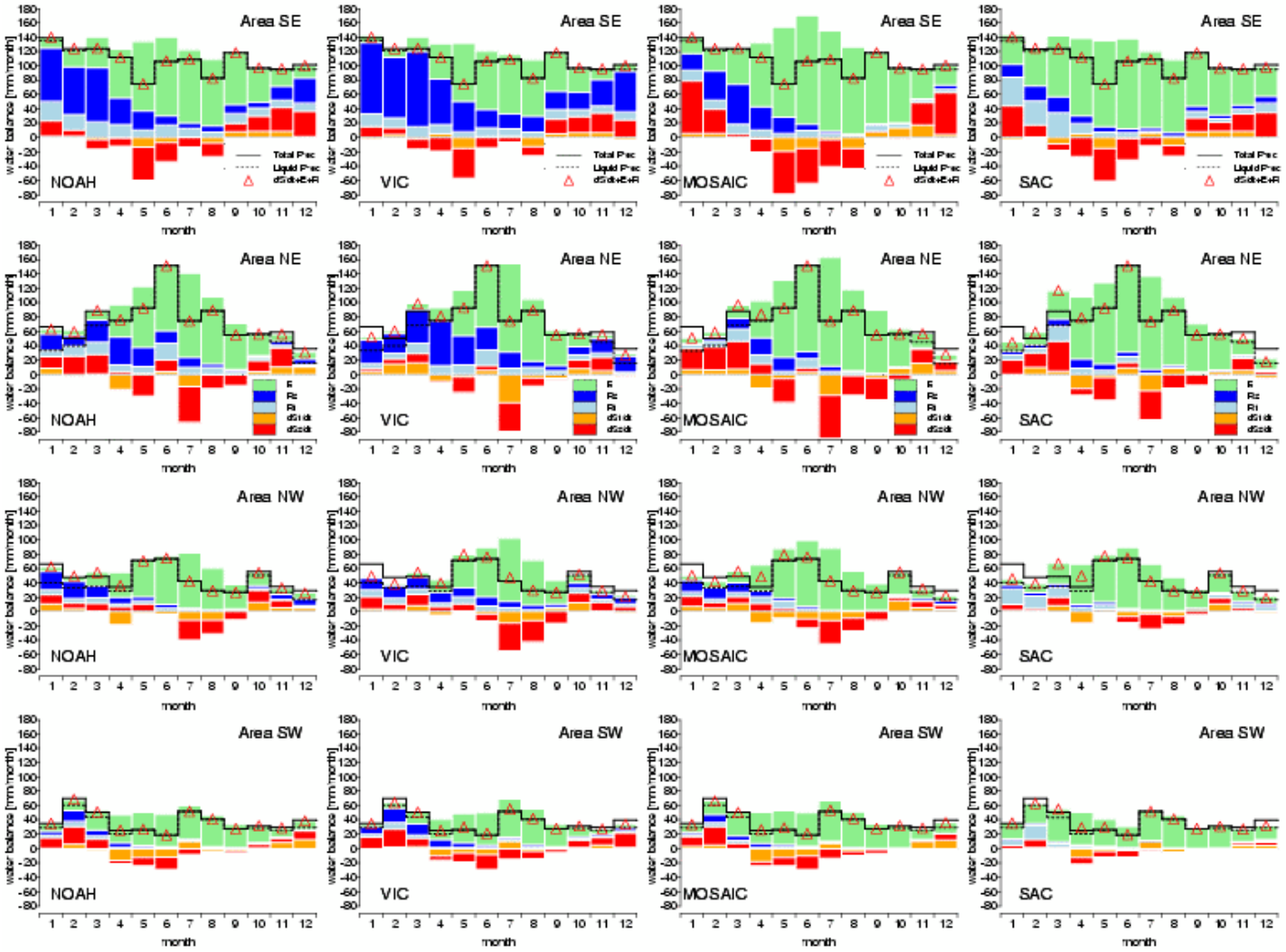

Figure 12. Monthly water budget for October 1997 to September 1998. In order, columns 1-4 are Noah, VIC, Mosaic, and SAC, and rows 1-4 are quadrants SE, NE, NW, and SW. Colors depict terms (mm/month) in (1): dS_1/dt (orange), dS_2/dt (red), R_1 (light blue), R_2 (dark blue), E (green). Solid black line = total precipitation P (mm/month) and dotted black line = liquid precipitation (mm/month). See text for definition of red triangles.

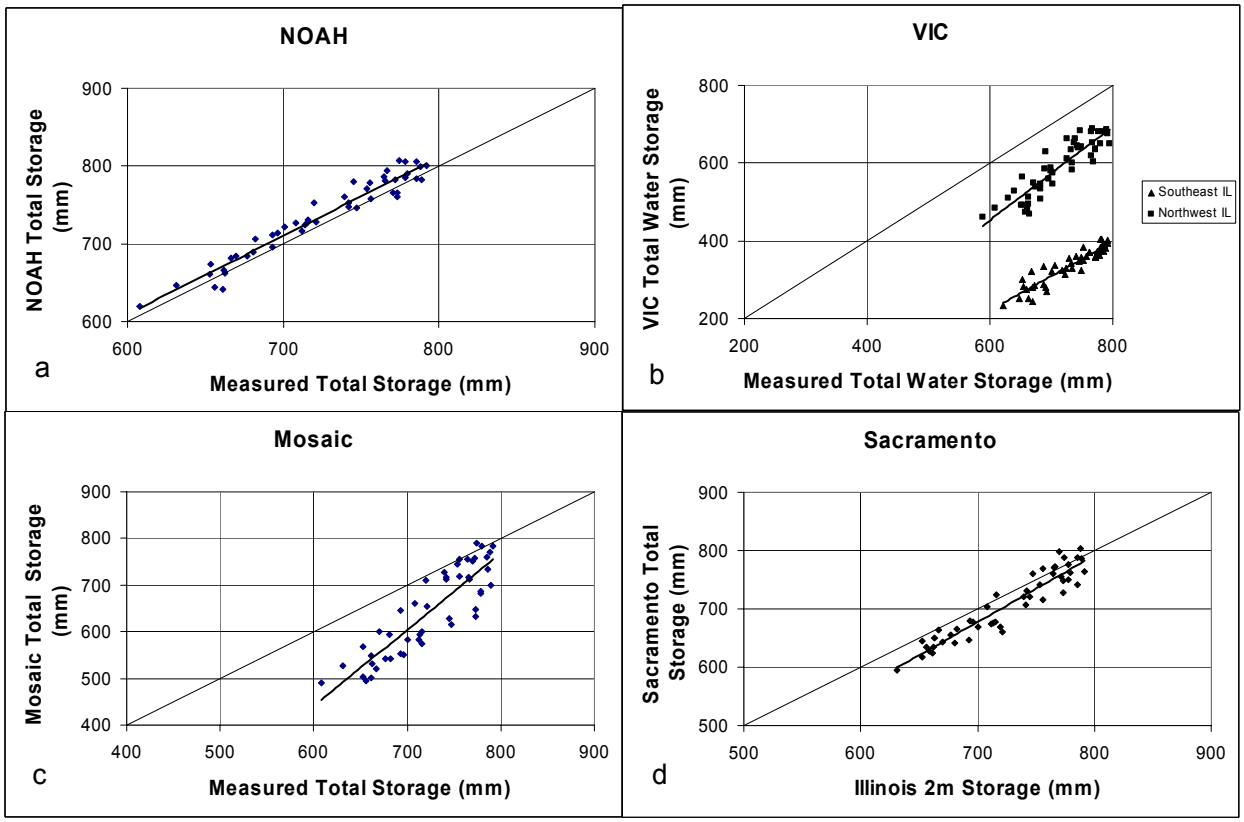

Figure 13. Comparison of model versus observed bi-monthly soil moisture (mm) in top 2 m, averaged over 17 sites throughout Illinois for the time period October 1997 to September 1999 for Noah (a), VIC (b), Mosaic (c), and Sacramento (d).

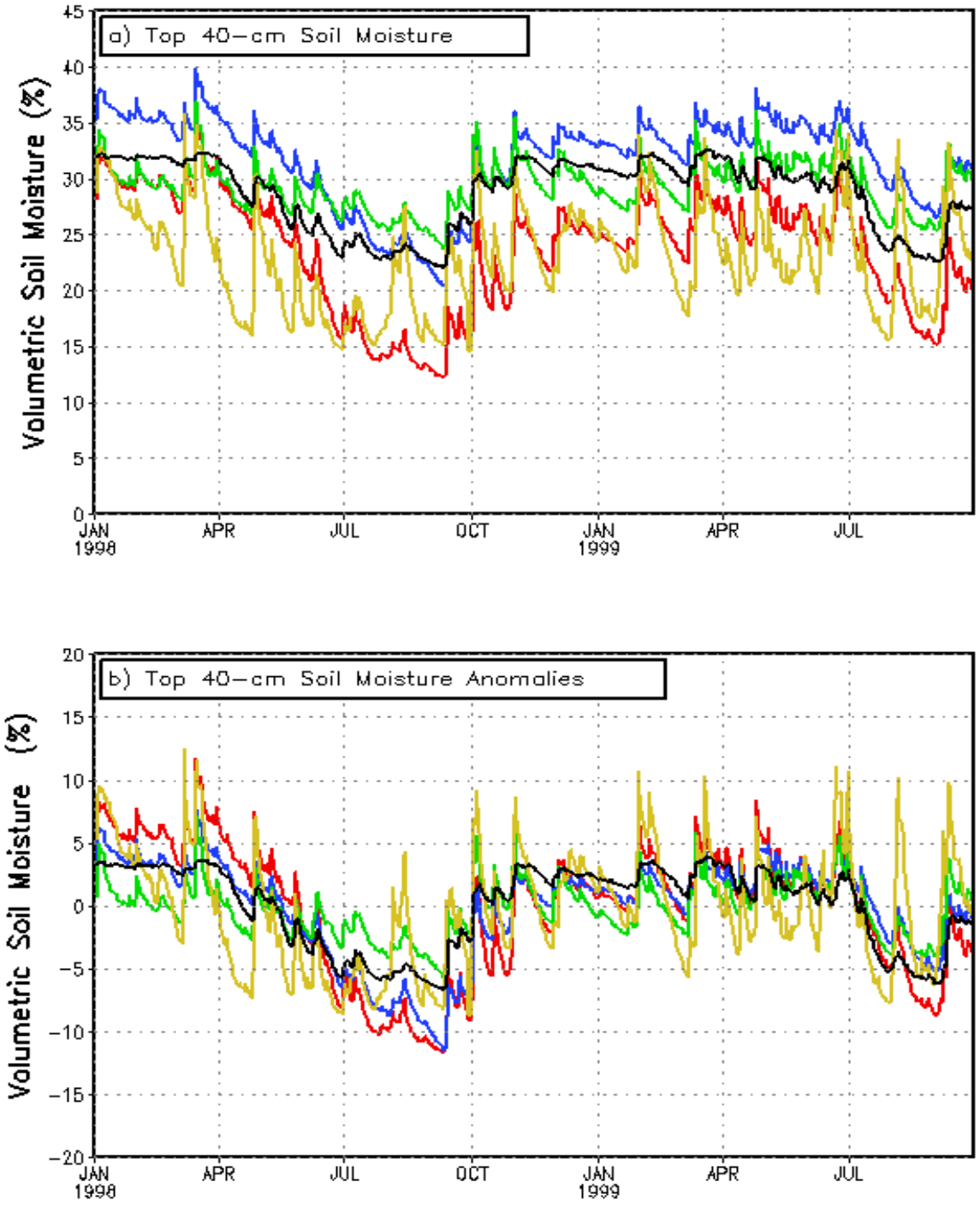

Figure 14. a) For the top 0-40 cm soil layer during the period January 1, 1998 to September 30, 1999, the time series of daily multi-station mean volumetric soil moisture of observations (black) and Noah (blue), VIC (green), Mosaic (red) and SAC (gold) model and over the Oklahoma Mesonet, for a) absolute value and b) departure from time average (over entire period) of corresponding time series in a).

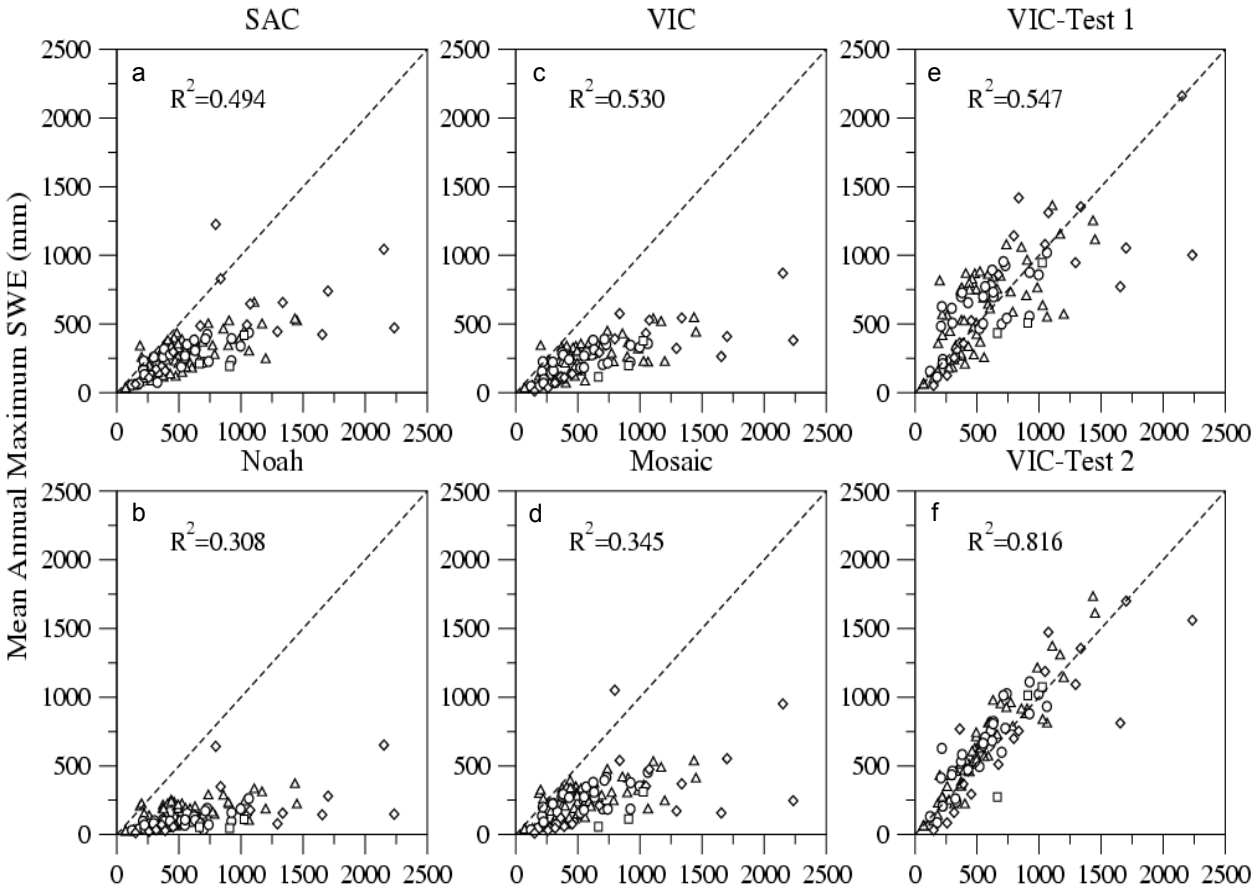

Figure 15. Comparison of mean annual maximum snow water-equivalent (*SWE*) during the period October 1996 to September 1999 between observations (x-axis) and model simulations (y-axis) at 110 SNOTEL sites for the control runs of SAC (a), Noah (b), VIC (c) and Mosaic (d), and two VIC tests runs forced with regionally adjusted precipitation (e) and locally adjusted precipitation (f).

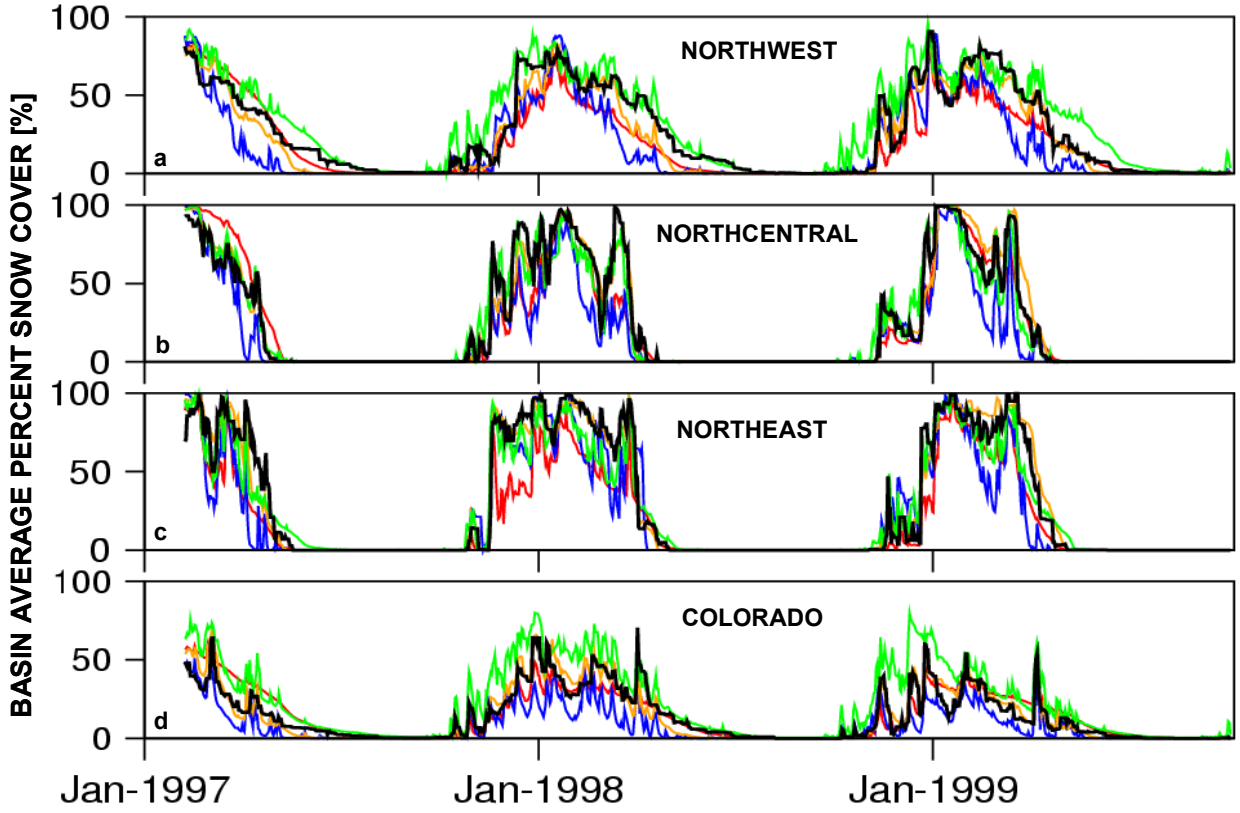

Figure 16. Time series of daily area mean snow cover extent (*SCE*) for IMS observed data (black) and the Mosaic (red), Noah (blue), VIC (green) and SAC (gold) models for the Northwest (a), North-central (b), Northeast (c), and Colorado (d) RFC regions for the time period February 1, 1997 to September 30, 1999.

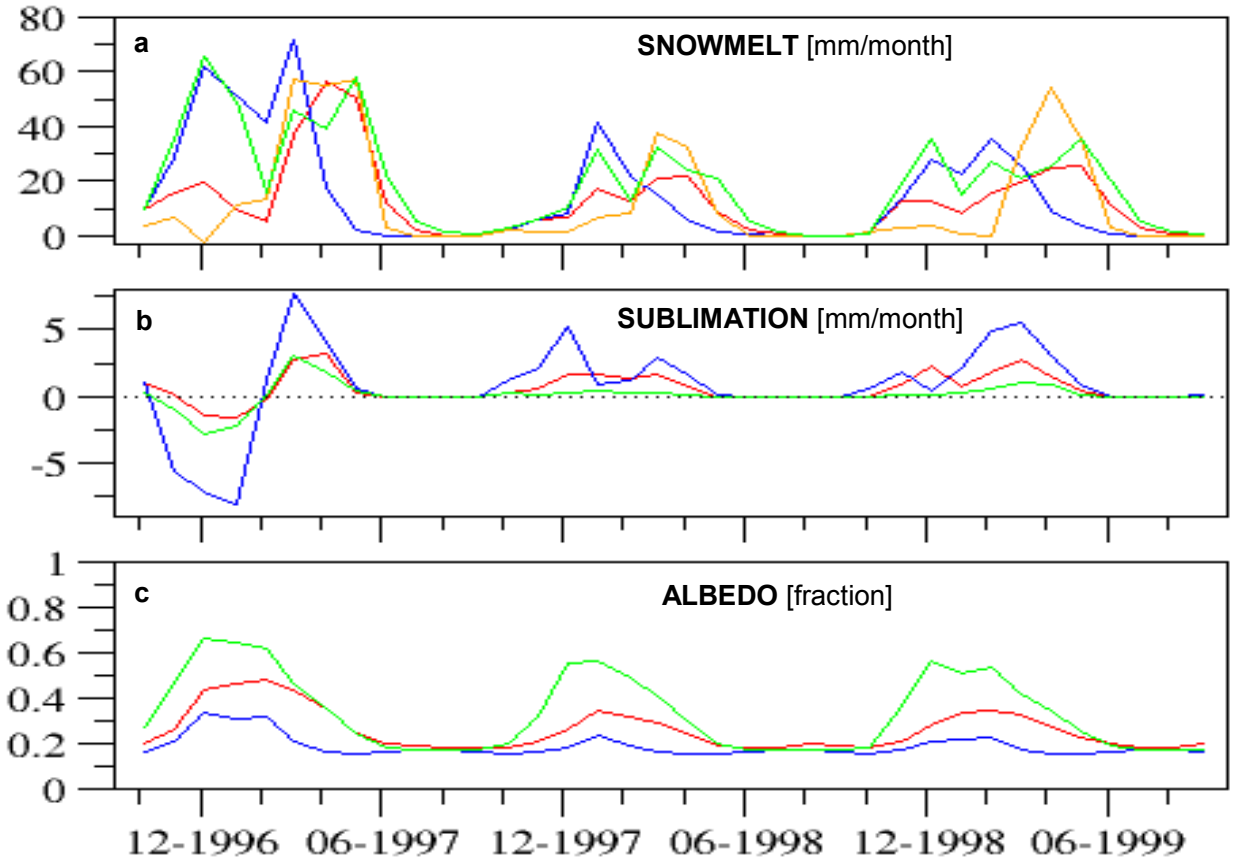

Figure 17. Time series of monthly area mean a) snowmelt (mm/month), b) snow sublimation (mm/month) and c) albedo for Noah (blue), VIC (green), Mosaic (red), and SAC (gold) for the time period October 1996 to September 1999 for the Northwest RFC. (Note different y-axis range in top panels.)

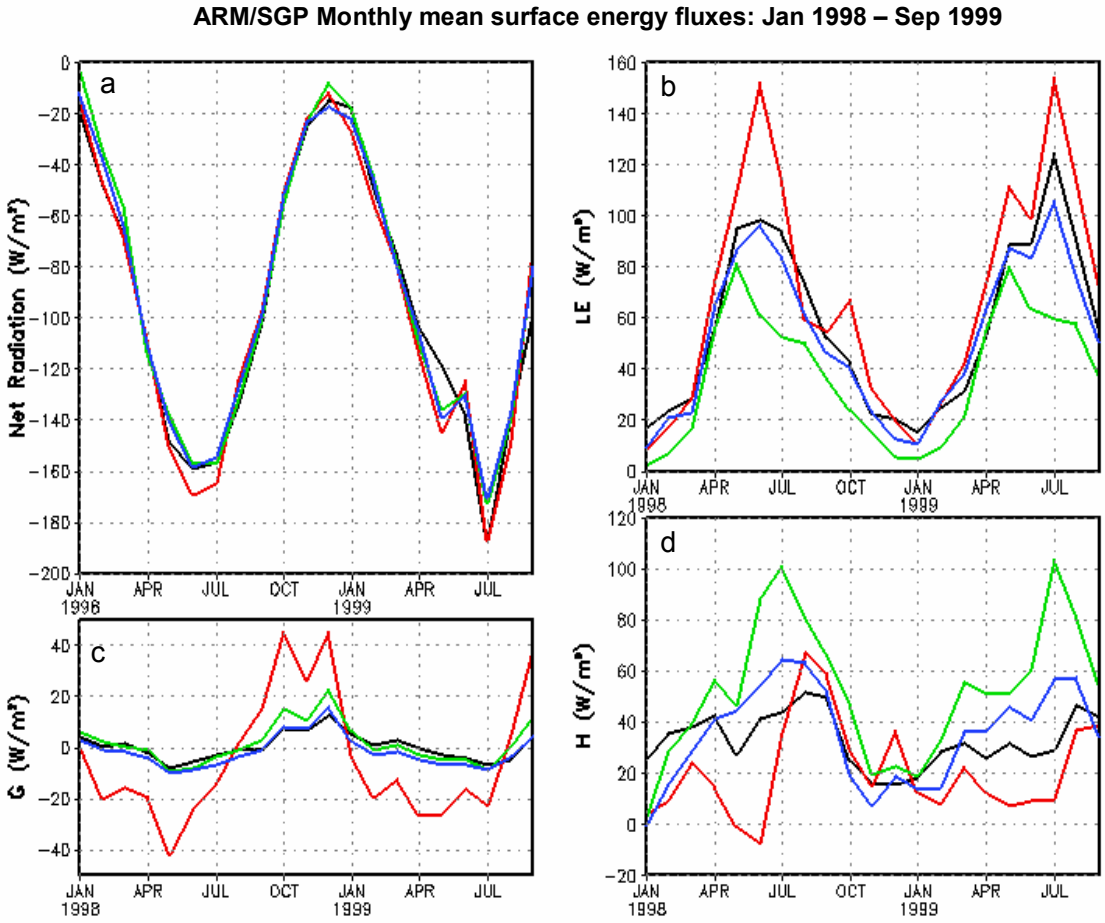

Figure 18. Time series of monthly mean surface energy fluxes (Wm^{-2}) of (a) net radiation *R*, (b) latent heat *LE*, (c) ground heat *G* and (d) sensible heat *H* averaged over the ARM/CART sites during January 1998 to September 1999 from observations (black) and Noah (blue), VIC (green), and Mosaic (red). Note different y-axis range among panels. For *R*, *LE* and *H*, a vertically upward flux is positive, representing a heat source to the atmosphere and sink to the surface. For G, a positive flux is a heat source to the surface.

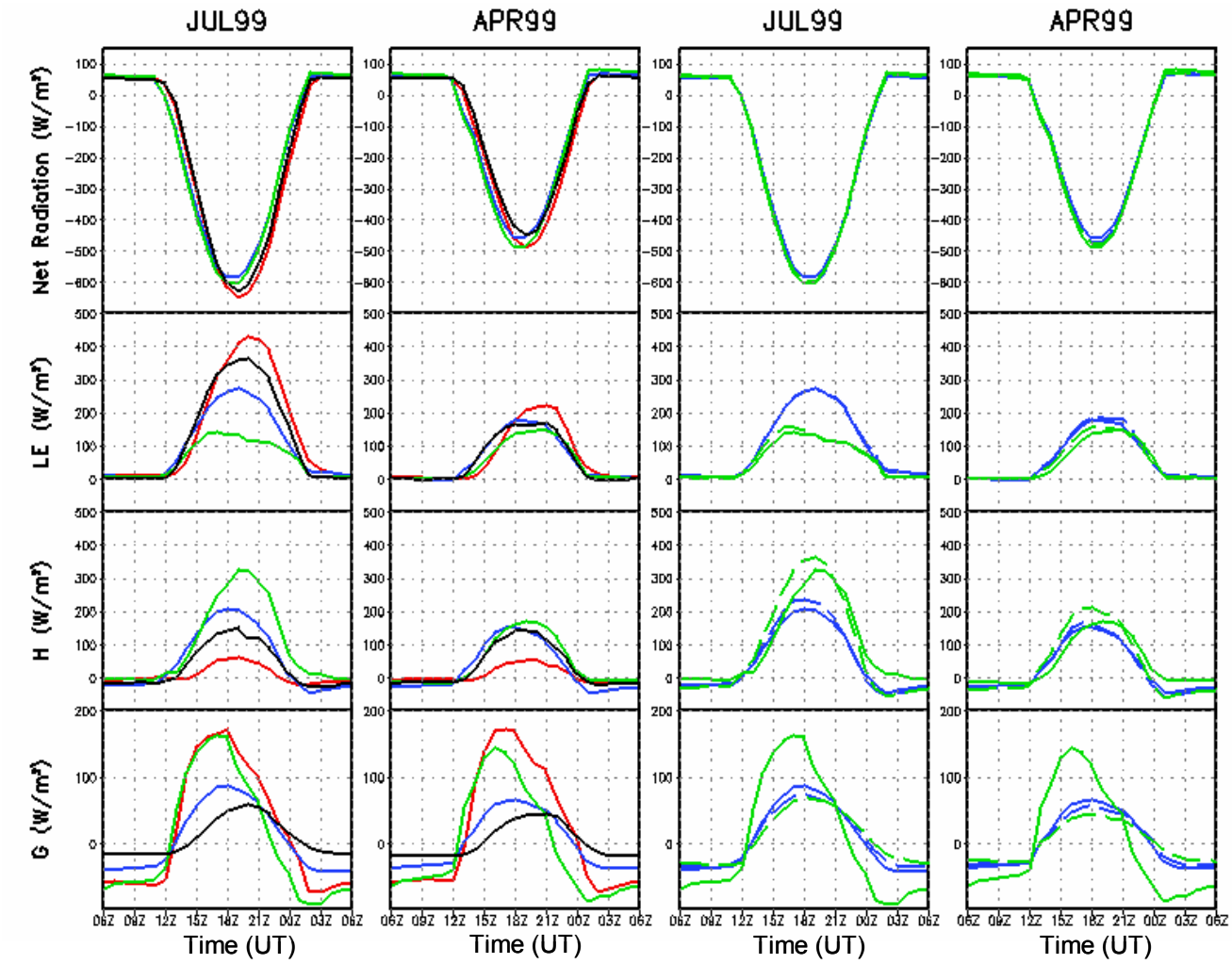

Figure 19. Monthly mean diurnal cycle in Wm^{-2} of net radiation *R* (row 1), latent heat *LE* (row 2), sensible heat *H* (row 3) and ground heat *G* (row 4) for July (columns 1,3) and April (columns 2,4) of 1999. Columns 1-2 show observations (black) and control runs for Noah (blue), VIC (green) and Mosaic (red). Columns 3-4 give control run (solid line) versus test run (dashed line) for Noah (blue) and VIC (green). Main text describes VIC and Noah tests. Note different y-axis range between rows. Sign convention as in Figure 18 (except for *G*).

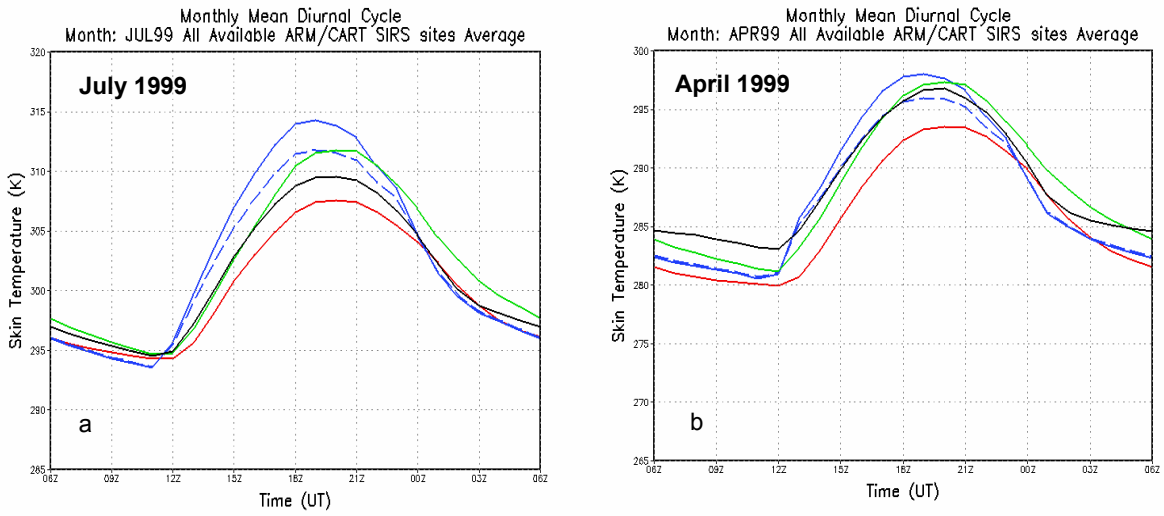

Figure 20. Monthly mean diurnal cycle of *LST* (K) averaged over all ARM/CART SIRS sites for July (left) and April (right) 1999 from the observations (solid black), the control runs (solid colors) for Noah (blue), VIC (green) and Mosaic (red), and a Noah test run (dashed blue) using a modified treatment for roughness length for heat in Noah's aerodynamic resistance.

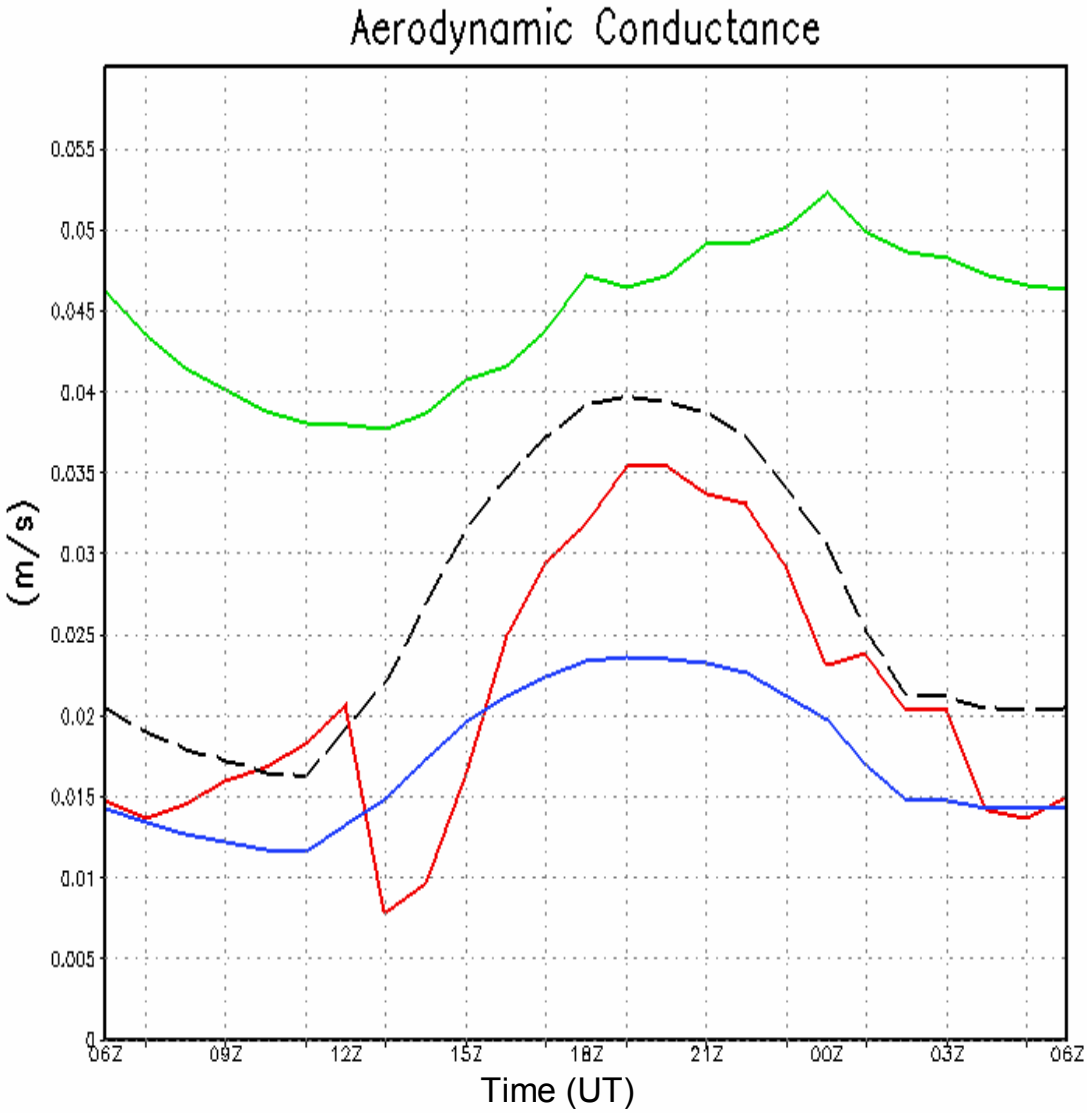

Figure 21. Monthly mean diurnal cycle of aerodynamic conductance (ms^{-1}) averaged over all ARM/CART SIRS sites for July 1998 for the control runs (solid colors) of Noah (blue), VIC (green) and Mosaic (red), and a test run of Noah (dashed black) using a modified treatment for the roughness length for heat.

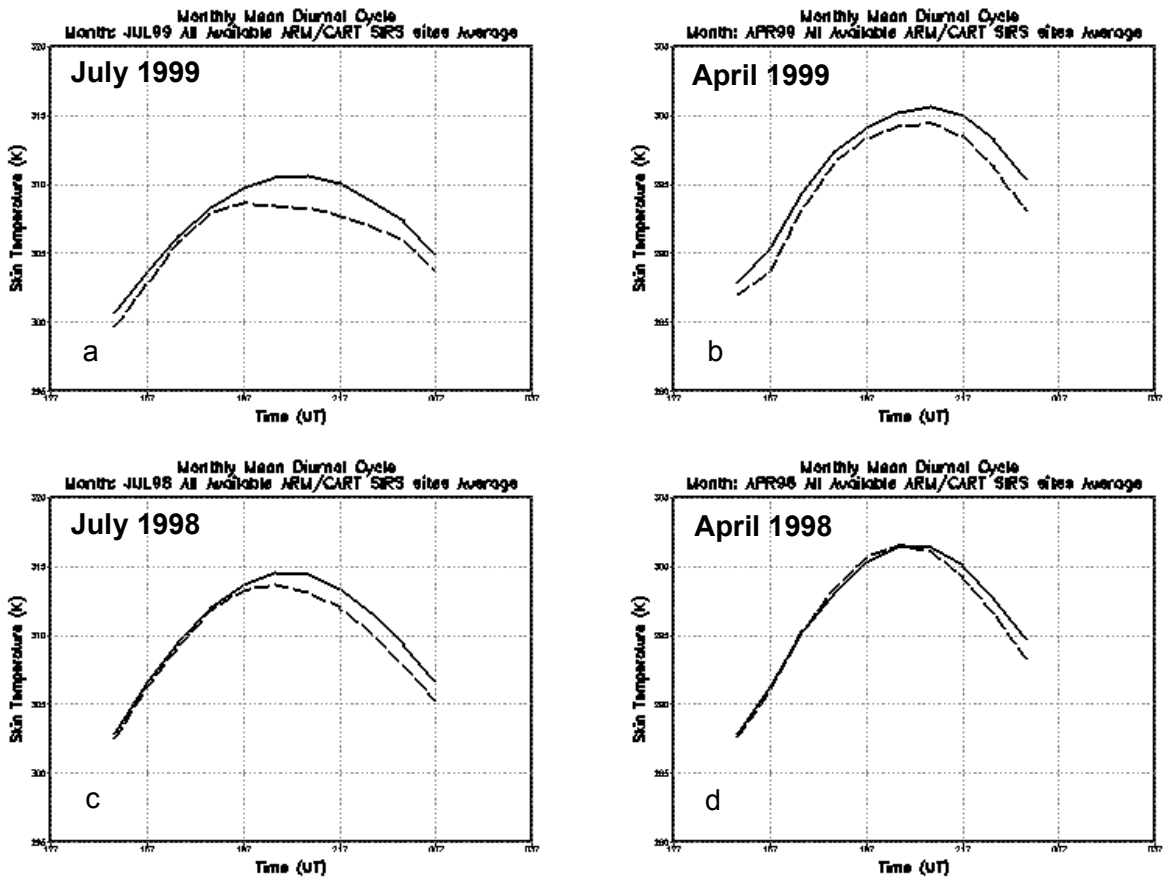

Figure 22. Monthly mean diurnal cycle of *LST* (K) averaged over all ARM/CART SIRS sites for July (left) and April (right) during 1999 (top) and 1998 (bottom) from SIRS observations (solid) and GOES-East retrieval (dashed).

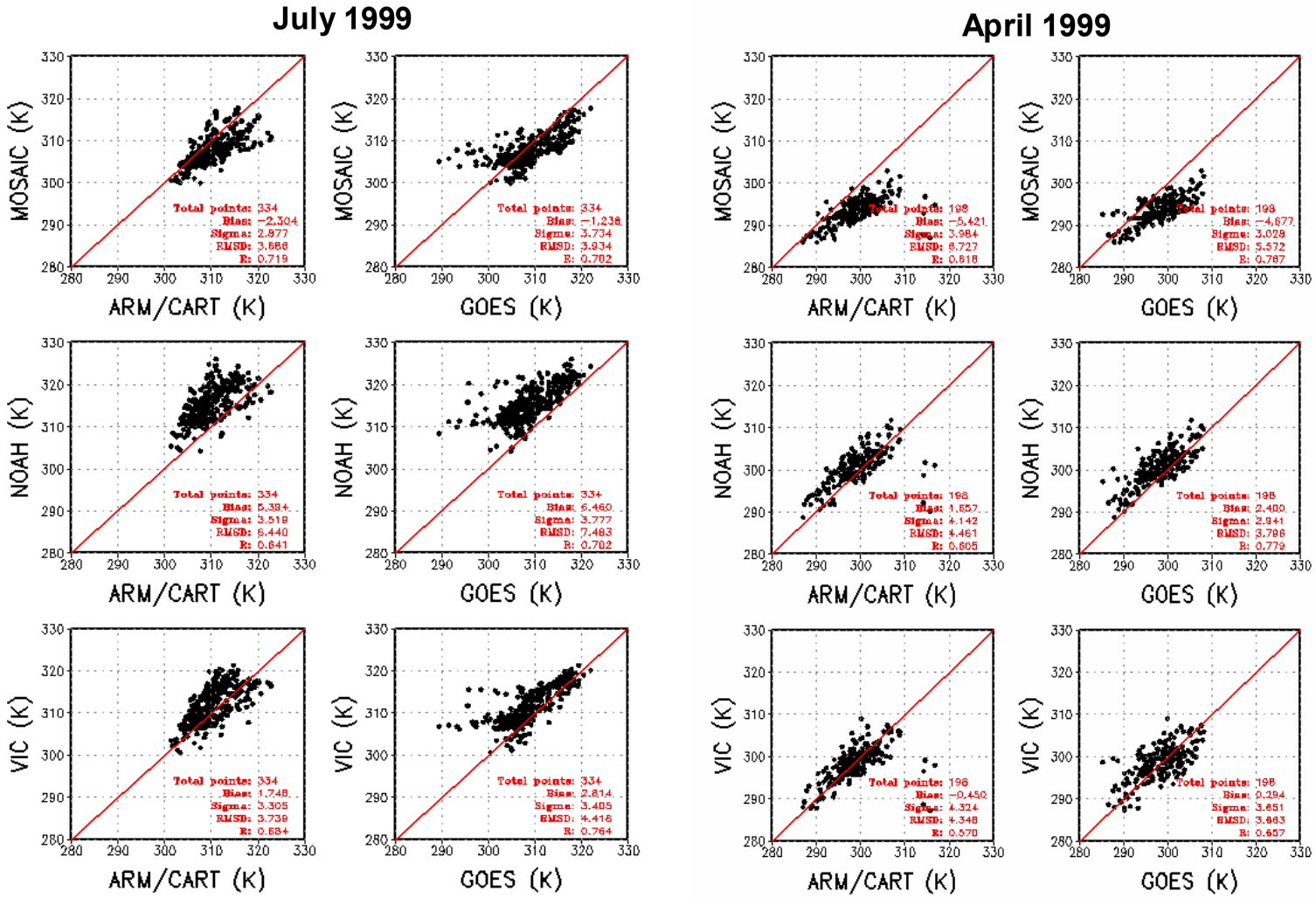

Figure 23. Comparison of model (y-axis) versus observed (x-axis) *LST* (K) at 18 UTC over all ARM/CART SIRS sites for July (columns 1-2) and April (columns 3-4) 1999 for Mosaic (top), Noah (middle) and VIC (bottom) versus SIRS observations (columns 1,3) and GOES-East observationsl (columns 2,4). Match-up point included only if GOES *LST* available (cloud free), yielding sample size of 334 in columns 1-2 and 198 in columns 3-4.

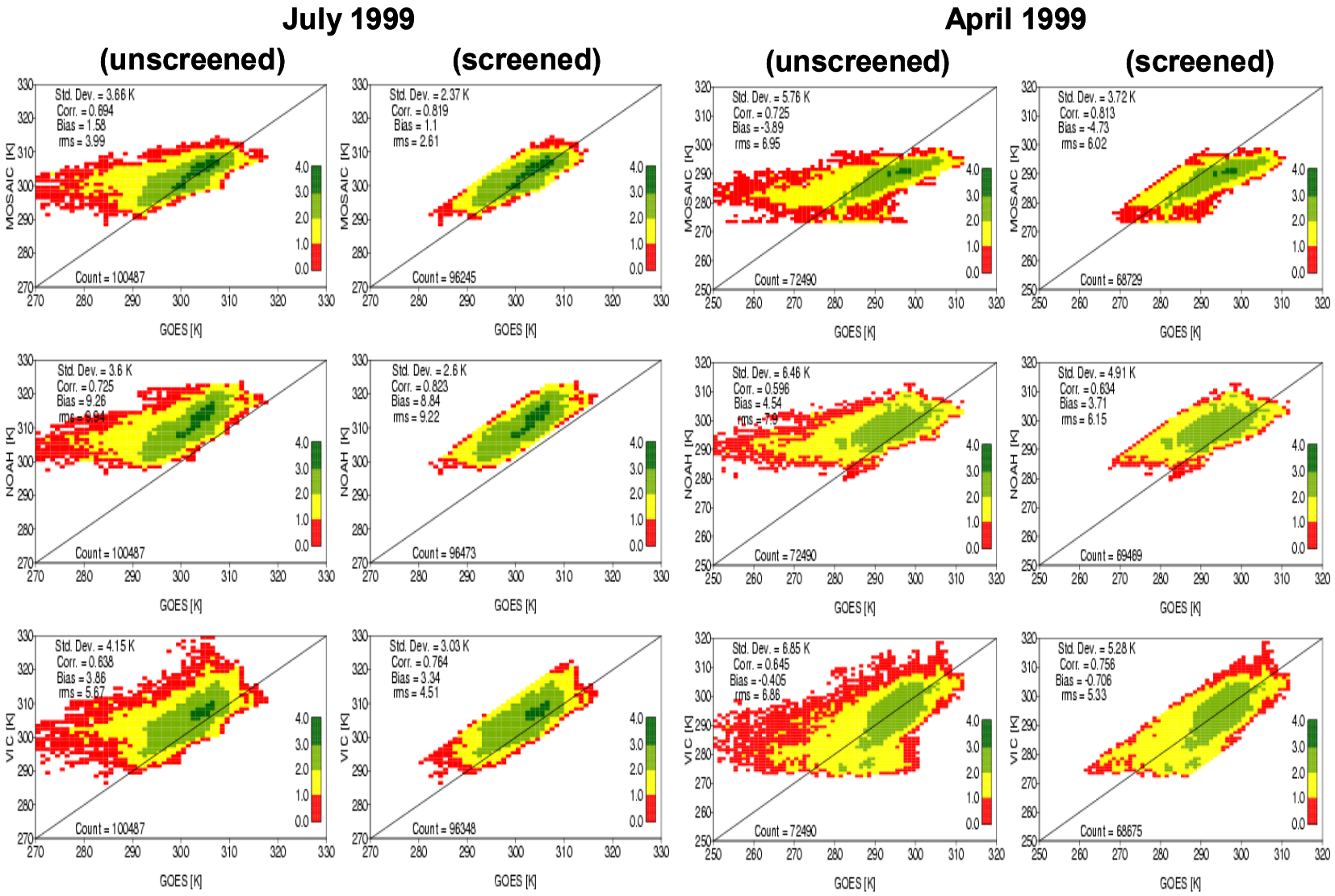

Figure 24. Comparison of model (y-axis) versus GOES-East (x-axis) *LST* (K) at 18 UTC over the northern Midwest during July (columns 1-2) and April (columns 3-4) 1999 for Mosaic (top), Noah (middle) and VIC (bottom) versus unscreened (columns 1, 3) and screened (columns 2,4) GOES *LST*. See text for color scale definition and latitude/longitude range of region.