Streamflow and Water Balance Intercomparisons of four Land-Surface Models in the North American Land Data Assimilation System Project

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

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

[2]NASA GSFC Hydrological Sciences Branch

[3]SAIC / NASA GSFC Hydrological Sciences Branch & Data Assimilation Office

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

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

[6]Department of Environmental Sciences, Rutgers University

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

[8]Department of Meteorology, University of Maryland

[9]NESDIS Office of Research and Applications

For submissions to JGR Atmospheres, GEWEX 3 special issue, February 2003

Corresponding author address:

Dag Lohmann

NOAA / NCEP/ EMC

Room 204 Phone:301-763-8000 ext. 7278

5200 Auth Road Fax: 301-763-8545

Suitland, MD 20746 Email: Dag.Lohmann@noaa.gov

ABSTRACT

This paper compares and evaluates streamflow and water balance results from four different Land Surface Models (LSMs) which participated in the multi-institutional North American Land Data Assimilation System (NLDAS). These LSMs have been run for the retrospective period 10/01/1996 to 09/30/1999 forced by atmospheric observations from the Eta Data Assimilation System (EDAS) analysis and ETA model output, measured precipitation and downward solar radiation. We have evaluated these simulations using measured daily streamflow data within 9 large major basins within the US and with 1145 smaller basins from 23 km^2 to 10,000 km^2 distributed over the NLDAS domain from the U.S. Geological Survey (USGS). Model runoff was routed with a distributed and a lumped optimized linear routing model. The diagnosis of the model results shows that the LSMs have a wide spread in their partitioning of precipitation into evapotranspiration and runoff. The modeled mean annual runoff shows large regional differences by a factor of up to four between models. The corresponding difference in mean annual evapotranspiration is about a factor of two. Runoff timing for the LSMs is influenced by snow melt timing with differences in the streamflow peaks of up to four months. While the modeled mean annual runoff shows large regional differences among the models and between the models and observations, it is under-estimated in areas with significant snowfall by all models. The monthly water budget shows that in the summer the model differences in runoff are as large as are soil water storage changes.

1. INTRODUCTION

The first phase of the multi-institutional North American Land Data Assimilation System (NLDAS) project [Mitchell et al., 2000, 2003] investigates the ability of four land surface / hydrology models (LSM) to reproduce measured fluxes. On small scales Luo et al. [2003] and Robock et al. [2003] looked at the forcing data and the performance of LSMs over the Southern Great Plains. Sheffield et al. [2003] and Pan et al. [2003] investigated the snow cover extent and snow water equivalent of the models at the point scale and over large spatial areas. This paper focuses on the ability of land surface models to reproduce measured streamflow and also inter-compares the large scale water budget of these models.

In previous off-line tests of land surface or hydrological models it has been shown that models are generally capable of reproducing streamflow time series on a monthly to annual time scale for large river basins up to 10^7 km^2 [Lohmann, 1998b, Oki, 1999; Mauerer, 2002; Bowling, 2003; Nijssen, 2003]. The resulting errors of the models can be attributed to an incorrect amount of runoff or an incorrect timing of the modeled runoff. The reasons for the over- or under-prediction of the total runoff amount on annual or seasonal time scales were addressed in the following major off-line studies. The Global Soil Wetness Project [GSWP, Dirmeyer, 1999] experiment showed that biases in the precipitation forcing led to biases of mean annual runoff [Oki et al., 1999, also Chapelon et al., 2002]. The biases in the resulting modeled streamflow were identified as a function of the precipitation station density. The Project for Intercomparison of Land-surface Parameterization Schemes (PILPS) phase 2(e) showed that differences in the sublimation physics of the models [Bowling, 2002, and Nijssen, 2002] are mainly due to snow surface roughness. Models with high sublimation loose their snow pack too early and

consequently under-predicted observed runoff. The PILPS phase 2(c) [Lohmann, 1998b] demonstrated that differences in runoff production parameterization affect the seasonal cycle of runoff. Models with almost no runoff production during summer precipitation events produced more realistic streamflow time series in the summer. These models produced runoff mainly by subsurface runoff. However, during periods of intense runoff production the resulting timing of runoff in these models was delayed. It was argued that this problem could be solved with a careful calibration of the model.

The question of runoff timing was also addressed in these off-line studies. In most of these studies a simple linear river routing algorithm was used to transform modeled runoff into modeled streamflow [see e.g. Lohmann et al., 1998a, 1998b, or Oki et al., 1999]. All these models are mathematically identical linear models and therefore can be described by the impulse response function of the governing equations. Differences in runoff timing were explained by different factors. Snowmelt timing differences were significant in Bowling [2003] and Boone [2003]. The storage of snowmelt in either the snowpack, surface ponding, or in the soil column influenced the timing of the runoff, but not the absolute magnitude [Bowling et al., 2003]. The resulting differences in peak runoff timing between models were on the order of days to up to 3 months. Boone et al. [2003] confirmed these results and documented that snowmelt timing on large spatial scales where differences in orography are significant can be improved by the introduction of snow bands. Differences in runoff production parameterizations introduced differences in streamflow peaks [Lohmann et al., 1998b]. Models with more sub-surface runoff production showed time delays for the peak streamflow on the order of one day to about one week for major flow events. Delays were mainly the result of different vertical water transport equations within the soil. Differences in the routing parameters lead to different horizontal travel

times of water in the river system [Oki et al., 1999]. Horizontal travel times in river beds for large basins are typically on the order of 0.5 to 5 m/s. Assuming a meandering ratio of 1.4 [Oki et al., 1999] this means that a flood wave will pass through one NLDAS grid cell (1/8 degree) in about 1 to 10 hours. We therefore expect a maximum timing uncertainty for basins for up to 10,000 km^2 to be about one or two days for an un-calibrated routing model. In a previous NLDAS related study for the large US basins, these uncertainties were on the order of weeks [Maurer et al., 2002]. It should be noted that a full implementation of the physically based hydraulic St-Venant equations [Chow, 1959] could improve this runoff timing, but would be computationally more expensive and more difficult to set up since more parameters are required.

During the last 10 years many publications showed that land surface models can successfully reproduce streamflow on daily to annual time scales for many river basins around the globe. However, it has never been reported for models other than the Bucket model that there are structural problems within a model that prevent it from modeling streamflow correctly. Entekhabi et al. [2000] point correctly to the limitations of current land surface and hydrology models which are used at scales from 1 km to 300 km. Most models are lumped single column models which operate outside of the spatial range for which the governing equations were derived. The underlying assumption is that the equations still capture the basic behavior of the system for which we can find effective parameters. This paper tries to uncover the strengths and weaknesses of the participating models and recommends areas in which future developments are needed.

The main areas of investigation are:

1) On the annual mean, LSMs show a wide range of partitioning precipitation into evapotranspiration and runoff. Monthly inter-model differences in runoff production have the same order of magnitude as inter-model differences in soil moisture storage change.

2) Comparing daily modeled and observed streamflow in many small to medium sized basins (23 km^2 to 10.000 km^2) within a region gives us a good idea about the spatial distribution of runoff modeling skills of LSMs. In areas with significant snowfall, the main factor in runoff timing is snowmelt timing. Daily streamflow data are needed to evaluate the snowmelt timing of the models.

3) For basins smaller than 10,000 km^2 lumped routing was used for daily streamflow intercomparisons and validation.

4) Results from the larger basins are consistent with the small basin results, but show significant signs of water resource regulation in the western basins.

To keep the impact of model spin-up to a minimum we decided to analyze only the model output from October 1997 to September 1999, the last 2 out of 3 years of model results The results are not necessarily encouraging; they show that we cannot model streamflow in most basins within the US without more work in parameter estimation techniques, model structure and processes, and input data.

2. NLDAS SET-UP AND NLDAS MODELS

The NLDAS configuration and models are described in detail in Mitchell et al. [2003], here only a short summary is given. NLDAS is an off-line data assimilation system in which four land surface models are driven by hourly atmospheric forcing data from the Eta Data

Assimilation System (EDAS, Rogers et al., 1997) and unified gage based precipitation analysis [Higgins et al., 2000] and satellite retrieval [Pinker et al., 1999] as described in Cosgrove et al. [2003, this issue] on a 1/8° latitude-longitude resolution over a domain that covers the continental US, part of Canada, and part of Mexico (125°W – 67°W, 25°N – 53°N). The hourly input data include precipitation, air temperature, air specific humidity, air pressure at the surface, wind speed, incoming solar radiation, and incoming longwave radiation. Hourly output fields from the LSMs include surface state variables such as soil moisture, soil temperature, snow water equivalent and surface fluxes such as latent, sensible, and ground heat flux, and runoff [Mitchell, et al., 2003]. The following models were used in the NLDAS system.

The Noah model is the model of the National Centers for Environmental Prediction (NCEP/EMC) also used as the lower boundary condition in many atmospheric models [Chen et al., 1996, Koren et al., 1999]. It participated in all major off-line land surface experiments conducted under the Project for Intercomparison of Land-surface Parameterization Schemes (PILPS) [Henderson-Sellers et al., 1993], the Global Soil Wetness Project (GSWP) [Dirmeyer et al., 1999], the Distributed Model Intercomparison Project (DMIP) [Smith, 2002], and the Rhone Aggregation Project [Boone et al., 2003]. It has also been used as a model in parameter estimation techniques.

The Mosaic land surface model developed by Koster and Suarez [1996] is a surface-vegetation-atmosphere transfer scheme (SVATS) that accounts for the sub-grid heterogeneity of vegetation and soil moisture with a "mosaic" approach. It also participated in most of the off-line intercomparison studies.

The variable infiltration capacity (VIC) model [Liang et al., 1994; 2002; Cherkauer et al., 1999, 2002] has been widely applied to large continental river basins, for example the Columbia

[Nijssen et al., 1997]; the Arkansas-Red [Lettenmaier et al., 1996], the Weser Lohmann et al., 1998] River, the Elbe River [Lobmeyr, 1999] and the Upper Mississippi [Cherkauer and Lettenmaier, 1999], as well as at continental scales in Maurer et al. [2002] and global scales [Nijssen et al., 2001]. It has also participated in most other off-line projects within PILPS and GSWP.

The Sacramento model (SAC) is run together with the SNOW-17 model, both part of the National Weather Service River Forecast System [Burnash et al., 1973; Anderson, 1973]. SAC is a conceptual rainfall-runoff model. It has a two-layer structure, and each layer consists of tension and free water storages. The input data requirement of the SAC model is different from all the other models. The basic inputs needed to drive SAC are rain plus snowmelt from SNOW-17 and potential evapotranspiration. The outputs include estimated evapotranspiration and runoff. For the NLDAS runs, the potential evaporation was obtained from the Noah model output.

3. STREAMFLOW DATA AND FLOW DIRECTION MASK

Daily streamflow data for the time period of the retrospective forcing for the entire NLDAS domain were obtained from the USGS website (http://waterdata.usgs.gov/nwis.sw) and the Army Corps of Engineers. We selected 1145 small basins for which data were available from 10/01/1996 to 09/30/1999. Criteria for the selection were basin size (smaller than 10,000 km^2), no visible signs for reservoir operation (following OHD/NWS/NOAA) and no missing data. The 1145 basins represent 15041 grid points of the NLDAS domain, about 25% of the total land area of the reduced (cutoff at 50°N) NLDAS grid. Figure 1 shows the spatial distribution of mean annual measured runoff from these basins. The USGS and Army Corps streamflow data is stored

in cfs [cubic feet per second], for this paper we re-mapped these values to [mm/year] to get an idea about the spatial distribution of annual average runoff. The distribution follows closely the distribution of the mean annual precipitation as shown in Cosgrove et al. [2003, this issue] with maximum values in the southeast and the northwest sections of the USA.

The river flow direction mask was provided for 12 River Forecast Centers (RFC) by the Office of Hydrologic Development of the National Weather Service [Reed, 2002, personal communication]. They used a modified method of Wang et al. [2000] to assign to each NLDAS grid point an integer value between 1 and 8 to characterize the eight main flow directions within each grid cell. This is sometimes referred to as a D8 model [Fairfield and Leymarie, 1991]. This map was merged into one NLDAS map and error corrected for loops and incorrect flow directions. Similar data sets have been used on various scales by above cited studies and by other authors [Vörösmarty et al., 1989, Oki 1999]. Figure 2 shows the simulated river network for the Arkansas River. To show the reasonable agreement with the natural river network, we also plotted the river reach file RF1 data set from the Environmental Protection Agency (EPA). The complete simulated NLDAS river network is shown in Figure 3. We plotted the log10 of the upstream area in km^2 for each grid cell within the 12 RFC's.

4. ROUTING MODEL

The routing model used for this study is identical with the one used in previous PILPS experiments [Lohmann et al, 1998b; and Bowling et al, 2003]. It calculates the timing of the runoff reaching the outlet of a grid box, as well as the transport of the water through the river network. It can be coupled directly into a land surface scheme, thus adding a state variable

"surface water" to that LSM, or it can be used off-line (like in this study) from the LSM with no further feedback. It is assumed that water can leave a grid cell only in one of its eight neighboring grid cells, given by the river flow direction mask. Each grid cell can also function as the sink of runoff from its upstream area, like in the Great Basin (Utah, Nevada). Both with-in grid cell and river routing time delays are represented using linear, time-invariant and causal models [Lohmann et al., 1998a] which are represented by non-negative impulse-response functions.

The equation used for the transport within the river is the linearized St.Venant Equation.

$$\frac{\partial Q}{\partial t} = -D \cdot \frac{\partial^2 Q}{\partial x^2} + C \frac{\partial Q}{\partial x} \quad (1)$$

where Q is the discharge, D a dispersion of diffusion coefficient, and C the velocity. The coefficients which were set initially in Equation 1 to C = 2m/s and D = 50 m^2/s.

For this study we used the distributed approach of Lohmann et al. [1998b] for the major basins and small basins; as well as a simplified lumped approach for the small basins, in which we optimized the routing parameters for each model and each basin separately. The lumped approach convolutes the sum of each models runoff in each basin with one impulse response function UH(t). This function is solved by de-convoluting

$$streamflow_{meas}(t) = \frac{\Delta\tau}{86.4} \sum_{\tau=0}^{\tau\max} \left(\sum_i area^i \cdot R_{t-\tau}^i \right) \cdot UH_\tau \quad (2)$$

where streamflow is the measured streamflow in [m^3/s], $\Delta\tau$ is the time interval of the measurements [1 day], $area^i$ is the area of a grid cell in a basin in [km^2], R^i is the modeled runoff of a grid cell in [mm/day], 86.4 is the factor to account for the different units. τmax is the length of the impulse response function in units of $\Delta\tau$, which did not exceed 7 days for all basins. It reflects the maximum concentration time of runoff within the basins. Equation 2 was typically applied for a time period of 1 year to calculate the resulting impulse response function.

Figure 4 shows the travel time distribution of surface water for the US. With the current parameters of the distributed routing model all runoff produced by the LSMs reaches the outlet of the river basins with maximal 50 days.

Model streamflow is compared to the measured streamflow with the relative runoff bias

$$Bias = \frac{\overline{\mathrm{mod}} - \overline{meas}}{\overline{meas}} \tag{3}$$

and with the Nash-Sutcliffe Efficiency criterion

$$Efficiency = 1 - \frac{\sum_i (meas_i - \mathrm{mod}_i)^2}{\sum_i (meas_i - \overline{meas})^2}, \tag{4}$$

where mod_i is the modeled streamflow with mean $\overline{\mathrm{mod}}$ and $meas_i$ is the measured streamflow with mean $\overline{meas}$ for any given time period. The Nash-Sutcliffe coefficient is a

measure of the prediction skill of the modeled streamflow compared to mean daily observed streamflow. Efficiency below zero indicates that average daily measured streamflow would have been as good a predictor as the modeled streamflow. A perfect model prediction has the score equal to one.

In most cases we found that using a simple lumped unit-hydrograph model for the small basin improved the resulting modeled streamflow as compared to distributed routing with default parameters. Of course, we could also optimize the distributed routing to archive the same result, but for the basin size we chose we cannot expect too much improvement as compared with lumped routing. The reason why the lumped routing is relatively successful is because precipitation, vegetation and soil types are spatially highly correlated in the small basins. Therefore runoff production is spatially highly correlated for the models within the NLDAS domain.

5. RESULTS AND DISCUSSION

We analyzed the water budget and streamflow on two different spatial scales, large and small river basins and three different time scales (daily, monthly, annual). While the small basins give us insight into spatially distributed processes for runoff production, the large major basins give us information about the large scale water balance.

5.1 Large Scale Water Balance Intercomparison

Figure 5 shows the spatial distribution of the mean annual evapotranspiration of the time period 10/1/1997 to 09/30/1999. Figure 6 shows the corresponding spatial distribution of mean annual runoff. The model results vary significantly in the eastern US, but show similarities in the

water limited drier western part. The similarity between the Mosaic model and Sacramento model is quite remarkable, given that the Sacramento model uses the potential evapotranspiration computed by the Noah model as a surrogate for the atmospheric forcing. To highlight these differences we divided the US into 4 quadrants (NW, NE, SW, SE). Figure 7 shows the mean annual sum of evapotranspiration and runoff for these four areas, with the diagonal line as the mean annual precipitation. Model symbols below the diagonal line indicate a positive storage change for the analysis time period. We can see that in the western part the total and relative differences of evapotranspiration and runoff are not as pronounced as in the eastern part, where they are quite dramatic in the eastern US. Mean annual runoff in the NE quadrant varies by a factor of 4 between the VIC model and the Sacramento model, and by a factor of 3 in the SE between the VIC model and the Mosaic and the Sacramento model, with the Noah model falling in between. Figure 8 analyzes this water balance for the 1145 small basins from Figure 1 and excludes all other grid cells. The vertical lines indicate the mean annual measured runoff values. For the NE, SE, and SW the Noah is closest to the mean annual observed streamflow, with the Mosaic and the Sacramento model producing less runoff than observed and the VIC model producing more. However, all models do not produce enough runoff in the NW quadrant. We think a significant low bias in NLDAS is mainly due to the precipitation forcing in this mountainous area (see companion papers by Sheffield et al. [2003] and Pan et al. [2003], this issue).

Two parallel VIC modeling efforts were carried out as part of the NLDAS project. Maurer et al [2002] performed a 50-year retrospective LSM run over the NLDAS domain, at a $1/8^{th}$ degree spatial and 3-hourly temporal resolution. Wood and colleagues at Princeton University performed the VIC real-time NLDAS runs, which are analyzed in this and companion

NLDAS papers. The real-time VIC NLDAS runs use essentially the same parameters as Maurer et al. [2002]. One significant difference in the Maurer et al and real-time VIC runs is that the 50-year runs were performed at 3-hour time step, and used an equal partitioning of the daily gridded station data into 3-hour time intervals within the day. Maurer et al. [2002, Figure 2] analyzed the impact of the equal distribution of daily precipitation within 3-hour time steps as opposed to a more realistic disaggregation scheme based on observed hourly precipitation. The Maurer et al results show that the differences, for the subregion (Lower Mississippi basin) analyzed, were modest. Nonetheless, comparisons between the retrospective and real-time VIC runs show that the impact of the temporal disaggregation, and the impact of another difference in the runs, namely spatial disaggregation of precipitation, which was implemented in the real-time runs, but not by Maurer et al, can have a much larger effect than the lower Mississippi results suggest. This apparently has to do with a) the difference between 3-hourly time steps, used by Maurer et al, and hourly time steps in the real-time runs; and b) interactive effects of temporal and spatial disaggregation of precipitation. The differences (shown for a transect across the eastern and central U.S. at 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. A more detailed examination of the differences, and development of parameter transformations to account for the temporal disaggregation issues, will be the subject of a future paper. Robock et al. [2003] looked into model differences in the SGP ARM region. They showed that the Mosaic model had too much evapotranspiration over vegetated areas, while the VIC and the Noah model had too little.

Figure 9 shows the monthly water budget for the first year of the same analysis period for each of the models in the four quadrants. The black line is the precipitation and the deviation of

the red triangles from the solid black line indicates snow processes as follows: A red triangle is as much above (below) the solid black line as snow melts (accumulates). The main differences between the model results are: For the SE area the Mosaic model has the largest soil water storage changes within its annual cycle, associated with a higher evapotranspiration rate in July and August than any other model, about 20% of its total evapotranspiration. All models vary also by the way they produce runoff as surface and sub-surface runoff, with VIC producing the most sub-surface runoff and the Sacramento model the least. In the NE the models show very similar as in the SE, the major difference is that snow processes become more important. The Sacramento model shows the largest snow accumulation and melt, the Noah model has almost no snow accumulation and melt. This is also the case in the NW region. In both western areas summer evapotranspiration is largely influenced by the storage change. The Mosaic and the VIC model have the largest storage changes, while the Sacramento model has the smallest ones. Another noticeable difference is the different runoff production mechanism of the models. The VIC model produces most of its runoff as sub-surface runoff, Noah and Mosaic produce slightly more surface runoff and the Sacramento model produce most of its runoff as surface runoff.

5.2 Small Scale Runoff Validation

Figure 10 shows the observed (black curve) and modeled streamflow for the Nehalem River near Foss in Oregon (USGS code 14301000) for all four NLDAS models. In this basin all models have relatively low biases (less than 5%) and high correlation (R) values. Noticeable differences between modeled and observed data are the high baseflow of VIC during the summer. All modeled streamflow curves are similar. There are about 20 basins within the NLDAS area for which all four models perform similarly well. Figure 11 shows the derived

lumped unit-hydrograph (from equation 2) for the 4 models which minimizes the least square difference between modeled and measured data for each simulation. It was calculated using the iterative de-convolution technique explained in detail in Lohmann et al. [1998a,b] with modeled runoff instead of effective precipitation (see Equation 2). This unit-hydrograph represents the best linear lumped routing procedure for this catchment for each model and takes into consideration the different runoff production mechanisms of the four models. It is a measure of the distribution of the residence time of surface water in the catchment after it has been produced as runoff from the LSM. The different hydrographs for each model can be explained as follows: the Sacramento model and the VIC model produce more fast runoff then other two models. To match the measured flow they therefore need to keep this runoff longer in a horizontal routing model. This interplay between runoff production and horizontal transport is often neglected. We used this optimization procedure for all 1145 basins. While for most basins the resulting model predictions improve as compared to a distributed model with default parameters, the iterative de-convolution scheme failed whenever there was no parsimonious unit-hydrograph to transform modeled runoff into streamflow. This mainly occurred in areas with significant snowfall or low runoff ratios. In these cases the distributed runoff routing model was used to calculate the error statistics.

Figure 12 shows the observed (black curve) and modeled streamflow for the Wind River near Crowheart, Wyoming (USGS code 06225500). The streamflow measurement station is 1718 m above sea level and the runoff timing is highly influenced by snow melt. The modeled hydrographs contain interesting information about the snowmelt timing of the four models. The Noah model starts melting snow in early March and has no more snow to melt in the basin at the

end of June. The other models start melting the snow significantly later. SAC and Mosaic start melting their snow late April and early May for both years, but Mosaic melting period lasts until the end of August, while SAC has a shorter melt period until the end of June. The VIC model has different snow melt timing than the other models. It does not start melting until mid-May and has its peak in snowmelt induced runoff production in July, about three weeks later than the observed streamflow. Its snowmelt period lasts about as long as Mosaic's. The data emphasize the point that for basins of this size daily streamflow data are important to understand inter-model differences as well as check the LSMs parameterizations of snowmelt with measured streamflow data. It is also important to note that Pinker et al. [2003] found a substantial high bias in NLDAS solar insolation over areas with winter snow cover. The low NLDAS precipitation bias over the Northwest combined with the high solar insolation bias suggests that the Noah model would improve its snow melt timing with revised forcing data and that also the Mosaic and the VIC model would melt the snow even later in the season.

For annual time scales, one can assume the storage change in the water balance equation to be close to zero, and therefore precipitation equals the sum of runoff and evapotranspiration for long time scales, although Cosgrove [2003] showed that one year of spin-up might not be long enough for some parts of the NLDAS domain. From Figures 7 and 8 we can see that storage change (deviation from the diagonal line) is an order of magnitude smaller than precipitation itself. Figure 13 shows the relative bias of the mean annual runoff of all four NLDAS models for the time period from 10/01/1997 to 09/30/1999. All models under-predict the mean annual runoff in the northern Rocky Mountains in most basins by 20% to around 80%. The main reason for this can be found in the companion papers by Sheffield et al. [2003] and Pan et al. [2003]. They show

that for 110 screened SNOTEL stations within the NLDAS area the NLDAS precipitation forcing is more than 50% too low compared to station measurements. They showed that by increasing the amount of precipitation by a constant factor of 2.1693 (from regression analysis) most of the errors in the snow water equivalent can be reduced significantly, However, other potential forcing errors (e.g. solar radiation or air temperature) also would need to be addressed.

The VIC model overestimates runoff by more than 60% in the southeast and midwest. The corresponding evapotranspiration Figure 5 shows that the VIC model produces less evapotranspiration than all the other models. The largest relative runoff biases are in the southeast and the midwest region. The Noah model has a similar spatial structure of relative biases; however the biases tend to be smaller and of either sign. For large areas in the east, for example, the Noah model also under-predicts streamflow. The Mosaic and the SAC model have very similar patterns of relative runoff biases. Both consistently under-predict runoff throughout the NLDAS area. The only exceptions are basins in North-Texas, New-Mexico, and Oklahoma where all models show too much runoff. The reasons for this have not been investigated for this paper, but it should be noted that about 50% of the US is cultivated farmland, with almost 5% of the total area receiving irrigation of more than 500 mm/year water on average. These effects are not included in the current NLDAS set-up.

To gain more insight into seasonal differences between the models, we computed the cold (Figure 14) and the warm season (Figure 15) runoff biases of the models. The Noah model slightly over-predicts runoff in most basins in the east to the mid-west in the cold season, but under-predicts runoff in the east and over-predicts runoff in the mid-west in the warm season.

The VIC model shows a similar spatial pattern, just with more runoff, and therefore a consistently larger positive bias. The Mosaic and the Sacramento model have very similar bias distributions in the cold season, but different patterns in the warm season. While the Mosaic model under-predicts runoff more in the Atlantic region, the Sacramento model under-predicts runoff more in the northern and upper mid-west.

Figure 16 shows the Nash-Sutcliffe efficiency for daily modeled streamflow for all basins for the same time period. All models share the same spatial structure of efficiency in the eastern US, with the SAC and the Noah model being slightly better than the other two models for most basins. The Mosaic model has the highest efficiency scores in the northern midwest, while the VIC model has the highest values of all the models in the Rocky Mountains and the northern part of the east. The relationship between mean annual snowfall and the correlation between modeled and observed streamflow is shown in Figure 17. All models show their lowest correlation in basins with high snowfall, but noticeably the Noah needs to improve its snowmelt timing. This indicates that more efforts are needed to produce reliable forcing data for these areas, as well as the need for intensified research for large scale snow models.

Figure 18 shows the temporal location of the maximum of the cross-correlation function between measured and observed streamflow data. Negative numbers indicate the number of days that the modeled streamflow peaked before the observed streamflow. Positive numbers show the number of days that the model lagged behind the observation. Basically over most of the country the models reproduce the streamflow peaks within plus or minus 3 days. This is also due to the fact that we optimized the routing model for large parts of the country. However, for the Rocky

mountains and the northeast, the cross-correlation function clearly indicates model trends for snow covered areas. The Noah model has many of its peak streamflows more than 2 months prior to the event. Mosaic and the Sacramento model have errors of about one month for many basins, while the VIC model seems to model runoff timing (and therefore snowmelt timing) very well, though sometimes it predicts snowmelt too late, as was also seen in Figure 12. These results are consistent with the companion NLDAS papers by Sheffield et al. [2003] and Pan et al [2003, this issue] and also with results from the PILPS 2(e) and Rhone Experiment (only VIC and Noah model participated from the four models).

Figure 19 was inspired by the work of Oki et al. [1999]. They showed that for the Global Soil Wetness Project (GSWP) there was a high correlation between runoff biases and the precipitation station density. Note that the station density in the NLDAS project is about an order of magnitude larger than in the GSWP project. We used the average station density for July 1997 to compute the station density. Each grid cell which had more than 30% of its area within one basin was counted as a station within that basin. The resulting pattern in the NLDAS project is not as prominent here compared to the GSWP pattern. It is possible that the biases are better explained by model physics (specifically snow, evapotranspiration and runoff parameterization) and precipitation amounts in mountain and snow covered areas then by the density of the precipitation network itself. The red dots in the figure are the basins with more than 100 mm/year snowfall. For all models the majority of the red dots show a clear negative bias in modeled runoff. For the basins with little annual snowfall (black dots), the Noah model has the least bias.

5.3 Large Scale Runoff Validation

The results for the major US rivers are consistent with the analysis for the 1145 small basins. Figure 20 shows the location of the 9 large basins and their gauging stations. The corresponding mean monthly streamflow is shown in Figure 21. These large basins can be seen as integrators of all the headwater systems upstream. However, it is unclear whether our modeling efforts need to include groundwater systems which are larger than the current grid box in the NLDAS setup. None of the models include such a horizontal water transport.

The largest differences between the modeled and the measured runoff are visible in the strongly regulated basins in the west. The Columbia and the Colorado River show a smaller seasonal signal than the modeled runoff. Previous modeling studies therefore used naturalized streamflow data [e.g. Maurer et al., 2002, and Lohmann et al. 1998], which are reconstructed time series were the influence of damns and reservoirs are removed. The main feature of the modeled streamflow in the Columbia and Colorado River is the difference in runoff timing between the models. Earlier results from the small scale basins are confirmed that show that the Noah model has a much earlier runoff production in the snow-melt season. For all other basins all models capture the seasonality of runoff reasonably well, with large differences in the low flows in the summer. Again, the same pattern as with the small basins can be observed. The VIC model produces too much runoff in the almost all basins, however, the Ohio River basin is modeled fairly well. In these large basins the integrative effect of the large areas seems to benefit the Noah model results, which showed a spatially varying slight under- and over-prediction in the east, and therefore reflects the water balance of most major basins the best of all models. The Mosaic and the Sacramento model consistently under-predict streamflow in the eastern region.

6. CONCLUSIONS

We presented results from a streamflow validation and water balance intercomparison for four NLDAS models used over the continental US. The results point to areas where more research, data collection and parameter estimation needs to be done. Also, some of the results are open for interpretation. Can we accept that the current default parameters in our models seem to introduce relative runoff biases on the order of 20 to 50%? Is it more important to get better precipitation and solar radiation estimates before we try to improve large scale snow models? Are there better ad-hoc parameter estimation routines?

The major results of this study are: all models failed to produce correct amounts of runoff in all basins, but some models did significantly better than other models in various parts of the country. Although VIC produced too much runoff in most parts of the NLDAS domain, it has the best snowmelt timing. This is consistent with results from the Rhone experiment [Boone et al., 2003]. It therefore seems to be advisable that other models adopt VIC's approach and introduce VIC's probably most distinguishing feature, elevation bands into their models. Of all models the Noah model is closest to the observed streamflow amount for most basins, but could improve its snow modeling, which seems to be the major shortcoming at this time. Given the experience and the success in recent modeling of large scale basins [Maurer et al., 2002] of the VIC model in modeling large scale basins, some of its results seem surprising. This will be addressed in a future publication. The Sacramento model is the most advanced and flexible model when it comes to the parameterization of runoff production processes, but it lacks the energy part of land surface modeling. It is unclear at this time whether a different parameter choice would have improved the model results. It is our suspicion that the different climatology introduced by

different forcing data could largely responsible to the under-prediction of streamflow. There are not too many calibration or parameter estimation studies with the Mosaic model and therefore it is not clear where the consistent under-prediction of streamflow comes from.

The model performance raises question of how we continue to develop models. It should be helpful if every model continues to test its simulation of runoff production and snowmelt in a variety of climate regimes. Running it through the PILPS 2(c) experiment [Woods et al., 1998], the PILPS 2(d) experiment at Valdai, Russia [Schlosser et al., 1997, 2000; Slater et al., 2001; Luo et al., 2003] and the PILPS 2(e) experiment [Bowling et al., 2003, Nijssen et al., 2003] could be a valuable first step.

Overall, the spread of the model results indicates that more work needs to be done. We need better forcing data in the mountains and over snow cover, mainly precipitation and solar radiation. In the current realtime NLDAS setting [Mitchell, 2000], unlike the retrospective runs here, we use the PRISM climatology to interpolate precipitation spatially, which might help with the low bias in the west. Parsimonious parameter estimation routines need to be implemented over large spatial areas. We should be able to reduce the spread amongst the models significantly and move closer to the observations. We should note that the VIC and the Sacramento model were developed and tested mainly with off-line data sets within medium to large river basins, and in the case of VIC also in the PILPS project. The Noah and Mosaic model have an early legacy of coupling to atmospheric models and only in the last couple of years have included more realistic descriptions of runoff production processes. It therefore remains to be seen whether large parts of the uncertainty in the modeling of streamflow comes from a lack in parameter estimation techniques or from incorrect parameterizations.

7. ACKNOWLEDGEMENTS

The work on this project by NCEP/EMC, NWS/OHD, and NESDIS/ORA was supported by the NOAA OGP grant for the NOAA Core Project for GCIP/GAPP (co-PIs K. Mitchell, J. Schaake, D. Tarpley). The work by NASA/GSFC/HSB 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, and additionally. The work by Princeton was supported by NOAA OGP GAPP grant NA86GP0258 (E. Wood, PI). The work by NCEP/CPC was supported by NOAA/NASA GAPP Project 8R1DA114 (R. W. Higgins, PI). The work by University of Maryland was supported by grants NA56GPO233, NA86GPO202 and NA06GPO404 from NOAA/OGP and by NOAA grant NA57WC0340 to University of Maryland's Cooperative Institute for Climate Studies (R. Pinker, PI).

We thank USGS for the streamflow data that were provided to the project at no cost.

8. REFERENCES

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

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 processes in the Torne-Kalix basin: PILPS Phase 2(e) 1: Experiment description and summary comparisons, *Journal of Global and Planetary Change*, accepted 2003.

Burnash, R.J.C., Ferral, R. L., and McGuire, R. A., A Generalized Streamflow Simulation System -Conceptual Modeling for Digital Computers, *Tech. Rep*., Joint Federal and State River Forecast Center, U.S. National Weather Service and California State Department of Water Resources, Sacramento, California, p 204., 1973.

Chapelon, N., H. Douville, P. Kosuth, T. Oki, Off-line simulation of the Amazon water balance: a sensitivity study with implications for GSWP *Climate Dynamics 19: 141-154, 2002.*

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

Cherkauer, K. A. and D. P. Lettenmaier, Hydrologic effects of frozen soils in the upper Mississippi River basin, *J. Geophys. Res.*, 104(D16), 19,599-19,610, 1999.

Cherkauer, K. A., L. C. Bowling and D. P. Lettenmaier, Variable Infiltration Capacity (VIC) Cold Land Process Model Updates, accepted *Global and Planetary Change*, 2003.

Chow, VT, Open channel hydraulics, Mc-Graw-Hill, New-York, USA, 1959.

Dirmeyer, P. A., A. J. Dolman, and N. Sato, The Global Soil Wetness Project: A pilot project for global land surface modeling and validation. *Bull. Amer. Meteor. Soc.*, 80, 851-878, 1999.

Entekhabi, D., Ghassem R. Asrar, Alan K. Betts, Keith J. Beven, Rafael L. Bras, Christopher J. Duffy, Thomas Dunne, Randal D. Koster, Dennis P. Lettenmaier, Dennis B. McLaughlin, William J. Shuttleworth, Martinus T. van Genuchten, Ming-Ying Wei, Eric F. Wood, An Agenda for Land-Surface Hydrology Research and a Call for the Second International Hydrological Decade*, Bulletin of the American Meteorological Society,* 80(10), 2043-2058, 1999.

Koren, V., J. Schaake, K. Mitchell, Q.-Y. Duan, F. Chen, J. Baker. A parameterization of snowpack and frozen ground intended for NCEP weather and climate models, *J. Geophys. Res.,* 104 (D16), 19.569-19585, 1999.

Koster R. D., Suarez M. J., Energy and Water Balance Calculations in the Mosaic LSM, *NASA Tech Memorandum* 104606, Vol.9, 1996.

Lettenmaier, D.P., F.A. Abdulla, 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.

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

Liang, X., and Z. Xie, A new surface runoff parameterization with subgrid-scale soil heterogeneity for land surface models, *Advances in Water Resources* (special issue on "Nonlinear Propagation of Multi-Scale Dynamics through Hydrologic Subsystems"), in press, 2002.

Lohmann, D., E. Raschke, B. Nijssen, D.P. Lettenmaier, Regional scale hydrology, Part II: Application of the VIC-2L model to the Weser River, Germany, Hydrological Sciences Journal, Vol.43(1) p. 143-158, 1998a

Lohmann, D., Lettenmaier, D.P., Liang, X., Wood, E.F., Boone, A., Chang, S., Dai, Y., Desborough, C., Dickinson, R.E., Duan, Q., Ek, M., Gusev, Y., Habets, F., Irannejad, P., Koster, R., Michell, K., Nasonova, O.N., Noilhan, J., Schaake, J., Schlosser, A., Shao, Y., Shmakin, A.,Verseghy, D., Warrach, K., Wetzel, P., Xue, Y., Yang, Z.-L., and Zeng, Q., 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. *Journal of Global and Planetary Change, Special Issue: Coupling Land and Atmosphere*, 19, 161-179, 1998b.

Luo, L., A. Robock, K. Vinnikov,, C, A. Schlosser, A. Slater, A. Boone, H. Braden, P. Cox, P. de Rosnay, R. Dickinson, Y.-J. Dai, Q. Duan, P. Etchevers, A. Henderson-Sellers, N. Gedney, Y. Gusev, F. Habets, J. Kim, E. Kowalczyk, K. Mitchell, O. Nasonova, J. Noilhan, A. Pitman, J. Schaake, A. Shmakin, T. Smirnova, P. Wetzel, Y. Xue, Z.-L. Yang, Q.-C. Zeng, Effects of frozen soil on soil temperature, spring infiltration, and runoff: Results from the PILPS 2(d) experiment at Valdai, Russia. *J. Hydrometeorology*, accepted 2003.

Luo, L., Robock, A., Mitchell, K., Houser, P., Wood, E., Schaake, J., Lohmann, D., Cosgrove, B., Sheffield, J., Duan, Q., Lettenmaier, D., Higgins, W., Pinker, R., and Tarpley, D., 2002, Validation of the North American Land Data Assimilation System Retrospective Forcing Over the Southern Great Plains, Submitted *to J. Geophys. Res., Atmospheres*, this issue, 2003.

Maurer, E.P., A.W. Wood, J.C. Adam, D.P. Lettenmaier, and B. Nijssen,, A Long-Term Hydrologically-Based Data Set of Land Surface Fluxes and States for the Conterminous United States, *J. Climate* 15(22), 3237-3251, 2002.

Mitchell, K., Lin, Y., Rogers, E., Marshall, C., Ek, M., Lohmann, D., Schaake, J., Tarpley, D., Grunmann, P., Manikin, G., Duan, Q., Koren, V., Recent GCIP-Sponsored Advancements In Coupled Land-Surface Modeling And Data Assimilation In The NCEP ETA Mesoscale Model, Preprints, 15th AMS Conference on Hydrology, 180-184, 2000.

Mitchell, K., Lohmann, D., Houser, P., Wood, E., Robock, A., Schaake, J., Cosgrove, B., Sheffield, J., Luo, L., Duan, Q., Lettenmaier, D., Pinker, R., Tarpley, D., Higgins, W., Meng, J., Marshall, C., Bailey, A., Wen, F., and Entin, J., The Multi-institution North American Land Data Assimilation System (NLDAS): Leveraging multiple GCIP products in a realtime and retrospective distributed hydrological modeling system at continental scale, Submitted to *J. Geophys. Res., Atmospheres*, this issue, 2003.

Nijssen, B., D. P. Lettenmaier, X. Liang, S. W. Wetzel, and E. F. Wood, Streamflow simulation for continental-scale river basins. *Water Resour. Res*., 33, 711-724, 1997.

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. Nishimura, and P. Dirmeyer, Assessment of annual runoff from land surface models using Total Runoff Integrating Pathways (TRIP), *J. Meteor. Soc. Japan*, **77**, 235-255, 1999.

Pan, M., J. Sheffield, E. Wood, K. Mitchell, P. Houser, J. Schaake, A. Robock, D. Lohmann, B. Cosgrove, Q. Duan, L. Luo, W. Higgins, R. Pinker, D. Tarpley, Snow Process Modeling in the North American Land Data Assimilation System (NLDAS). Part II: Evaluation of Model Simulated Snow Water Equivalent, Submitted to *J. Geophys. Res., Atmospheres*, this issue, 2003.

Pinker, R.T., J. D. Tarpley, I. Laszlo, and K. Mitchell, Houser, P., Wood, E., Robock, A., Schaake, J., ., Lohmann, D., Cosgrove, B., Sheffield, J., Luo, L., Duan, Q., Lettenmaier, D., Pinker, R., Tarpley, D., Higgins, W., Surface Radiation Budgets in Support of the GEWEX Continental Scale International Project (GCIP) and the Gewex Americas Prediction Project (GAPP) , Submitted to *J. Geophys. Res., Atmospheres*, this issue, 2003.

Robock, A., L. Luo., Wood, E.F., F. Wen, Mitchell, K., Houser, P., Schaake, J., Lohmann, D., Cosgrove, B., Sheffield, J., Duan, Q., Higgins, W., Pinker, R., Tarpley, D., Jeffery B. Basara,, Kenneth C. Crawford, Evaluation of the North American Land Data Assimilation (NLDAS) over the Southern Great Plains during the Warm Season, Submitted to *J. Geophys. Res., Atmospheres*, this issue, 2003.

Rogers, E., D. Parris and G. DiMego, 1999: Changes to the NCEP operational Eta Analysis. Technical Procedures Bulletin, (not numbered), NOAA/NWS. [National Weather Service, Office of Meteorology, Silver Spring, MD].

Sheffield, J., M. Pan, E. Wood, K. Mitchell, P. Houser, J. Schaake, A. Robock, D. Lohmann, B. Cosgrove, Q. Duan, L. Luo, W. Higgins, R. Pinker, D. Tarpley, B. Ramsay, Snow Process Modeling in the North American Land Data Assimilation System (NLDAS). Part I: Evaluation of Model Simulated Snow Cover Extent, Submitted to J. *Geophys. Res., Atmospheres*, this issue, 2003.

Schlosser, C.A., A. Robock, K. Vinnikov, N. Speranskaya,, Y. Xue, 18-Year land-surface hydrology model simulations for a midlatitude grassland catchment in Valdai, Russia. *Mon. Weather Rev.,* 125, 3279-3296, 1997.

Schlosser, C. A., A. Slater, A. Robock, A. Pitman, K. Vinnikov, A. Henderson-Sellers, N. Speranskaya, K. Mitchell, A. Boone, H. Braden, P. Cox, P. de Rosnay, R. Dickinson, Y.-J. Dai, Q. Duan, P. Etchevers, A. Henderson-Sellers, N. Gedney, Y. Gusev, F. Habets, J. Kim, E. Kowalczyk, O. Nasonova, J. Noilhan, J. Schaake, A. Shmakin, T. Smirnova, P. Wetzel, Y. Xue, Z.-L. Yang, Q.-C. Zeng, Simulations of a boreal grassland hydrology at Valdai, Russia: PILPS Phase 2(d). *Mon. Weather Rev*., 128, 301-321, 2000.

Slater, A. G., C. A. Schlosser, C. E. Desborough, A. J. Pitman, A. Henderson-Sellers, A. Robock, K. Ya. Vinnikov, N. A. Speranskaya, K. Mitchell, A. Boone, H. Braden, F. Chen, P. M.

Cox, P. de Rosnay, R. E. Dickinson, Y-J. Dai, Q. Duan, J. Entin, P. Etchevers, N. Gedney, Ye. M. Gusev, F. Habets, J. Kim, V. Koren, E. A. Kowalczyk, O. N. Nasonova, J. Noilhan, J. Schaake, A. B. Shmakin, T. G. Smirnova, D. Verseghy, P. Wetzel, Y. Xue, Z-L. Yang, and Q. Zeng, The representation of snow in land-surface schemes: Results from PILPS 2(d*). J. Hydrometeorology*, 2, 7-25, 2001.

Smith, M., Distributed model intercomparison project (DMIP), Project webpage http://www.nws.noaa.gov/oh/hrl/dmip, 2002

Vörösmarty, C.J., B. Moore, M.P. Gildea, B. Peterson, J. Melillo, D. Kicklighter, J. Raich, E. Rastetter, and P. Steudler, A continental-scale model of water balance and fluvial transport: Application to South America. *Global Biogeochemical Cycles 3*: 241-65, 1989.

Wang, M., A.T. Hjelmfelt, and J. Garbrecht, DEM Aggregation for Watershed Modeling, *Journal of the American Water Resources Association*, 36, 3, 2000.

Wood, E.F., D. Lettenmaier, X. Liang, D., Lohmann, A. Boone, S. Chang, F. Chen, Y. Dai, C. Desborough, R. Dickinson, Q. Duan, M. Ek, Y. Gusev, F. Habets, P. Irannejad, R. Koster, O. Nasanova, J. Noilhan, J. Schaake, A. Schlosser, Y. Shao, A. Shmakin, D. Verseghy, J. Wang, K. Warrach, P. Wetzel, Y. Xue, Z.-L. Yang, 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, *Global and Planetary Change*, 19(1-4), 115-135, 1998.

8. Figure Captions

Figure 1. Annual mean observed runoff in [mm/year] for 1145 small basins in the NLDAS domain for the time period 10/01/1997 to 09/30/1999. Data were provided by the USGS through their web-site http://www.usgs.gov.

Figure 2. Example of the NLDAS 1/8 degree simulated river flow direction. The red lines indicate the location of the real rivers from the EPA river reach file RF1. The blue triangles are the basin outlets (Arkansas, Missouri, upper Mississippi River).

Figure 3. Logarithm to the base 10 (log10) of the upstream area in [km^2] for all grid cells.

Figure 4. Distribution of travel times within the NLDAS domain for all grid cells to the outlet of each basin for the default parameters of the routing model.

Figure 5: Mean annual evapotranspiration [mm/year] for the NLDAS domain for the time period October 1997 to September 1999.

Figure 6: Mean annual runoff [mm/year] for the NLDAS domain for the time period October 1997 to September 1999.

Figure 7: Partitioning of precipitation of the four NLDAS models (N-Noah, V-VIC, M-Mosaic, S-Sacramento) in four different quadrants for October 1997 to September 1999 into runoff (x-axis) and evapotranspiration (y-axis). The diagonal line is the mean of the precipitation in each area. Models whose symbol falls below the line have a positive storage change for the time period.

Figure 8: Partitioning of precipitation of the four NLDAS models (N-Noah, V-Vic, M-Mosaic, S-Sacramento) in all basins which fall into the four different quadrants for October 1997 to September 1999. The partitioning of precipitation is quite similar to Figure 7. Each vertical line is the averaged measured runoff from the 1145 small basins within the NLDAS area. The runoff in some areas varies by a factor of eight (Vic and Sacramento in the northeast). All models under-predict runoff in the northwest.

Figure 9: Monthly water balance of the NLDAS models in the four quadrants of Figs. 3 and 4 for the time period October 1997 to September 1998. Orange = upper soil storage change, red = lower soil storage change, light blue = surface runoff, dark blue = sub-surface runoff, green evapotranspiration, black solid line = precipitation, black doted line = liquid precipitation, red triangles = storage change + evapotranspiration + runoff. The deviation of the red triangles from the solid black line indicates snow processes. A red triangle is as much above (below) the solid black line as snow melts (accumulates).

Figure 10. Observed (black curve) and modeled streamflow for the Nehalem River near Foss in Oregon.

Figure 11. Derived Unit-Hydrograph for the 4 NLDAS models for the Nehalem River near Foss in Oregon. .

Figure 12. Observed (black curve) and modeled streamflow for the Wind River near Crowheart, Wyoming. The station is 1718 m above sea level, and shows the impact of snow melt on runoff.

Figure 13. Relative runoff bias for the 4 NLDAS models for the time period 10/01/1997 to 09/30/1999. Notice that all models show less runoff than observed in the western snow covered areas.

Figure 14: Cold season relative bias (October – March) in runoff for the time period 10/01/1997 to 09/30/1999.

Figure 15: Warm season (April – September) relative runoff bias for the time period 10/01/1997 to 09/30/1999.

Figure 16. Nash-Sutcliffe Efficiency for the NLDAS domain for the time period 10/01/1997 to 09/30/1999 for daily mean modeled and measured data.

Figure 17. Relationship of mean annual snowfall in [mm/year] and the correlation of simulated and observed runoff. Each of the dots represents one of the 1145 basins.

Figure 18. Time to peak delay [days] of streamflow for 1145 basins calculated as the maximum of the cross-correlation function of modeled and observed streamflow. Negative numbers indicate streamflow peaks earlier than observed peaks. In snow covered areas Noah, Mosaic, and the Sacramento model show consistently too early snowmelt and therefore produce streamflow too early (See Ming et al., 2002).

Figure 19. Density of the observing rain-gauges as a function of the annual runoff bias for the time period 10/01/1997 to 09/30/1999. Basins with more than 100 mm/year snow are shown as red dots. This analysis has been done the first time for the GSWP experiment (Oki et al, 1999) and showed that the absolute runoff bias was a function of the gauging station density. That relationship is also observed here, but to a much lesser degree.

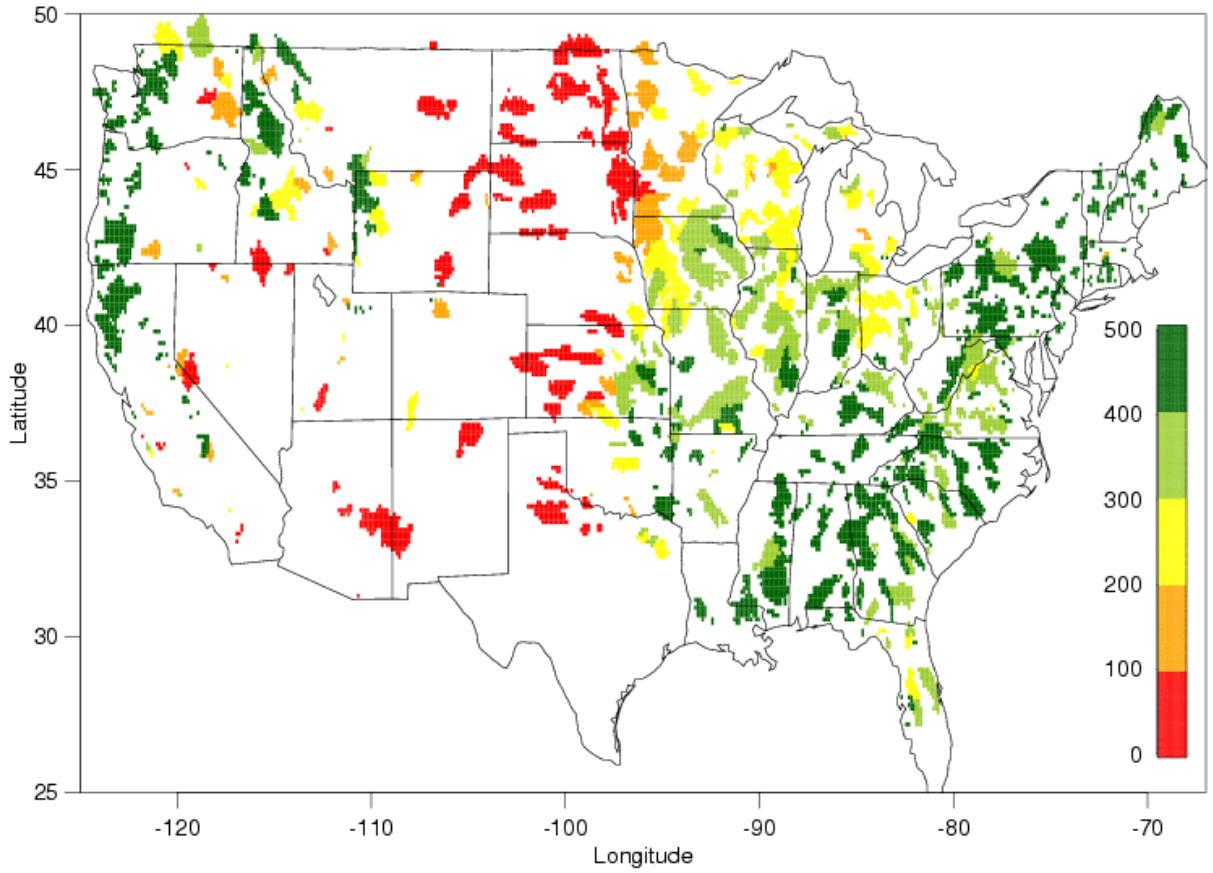

Figure 1. Annual mean observed runoff in [mm/year] for 1145 small basins in the NLDAS domain for the time period 10/01/1997 to 09/30/1999. Data were provided by the USGS through their web-site http://www.usgs.gov.

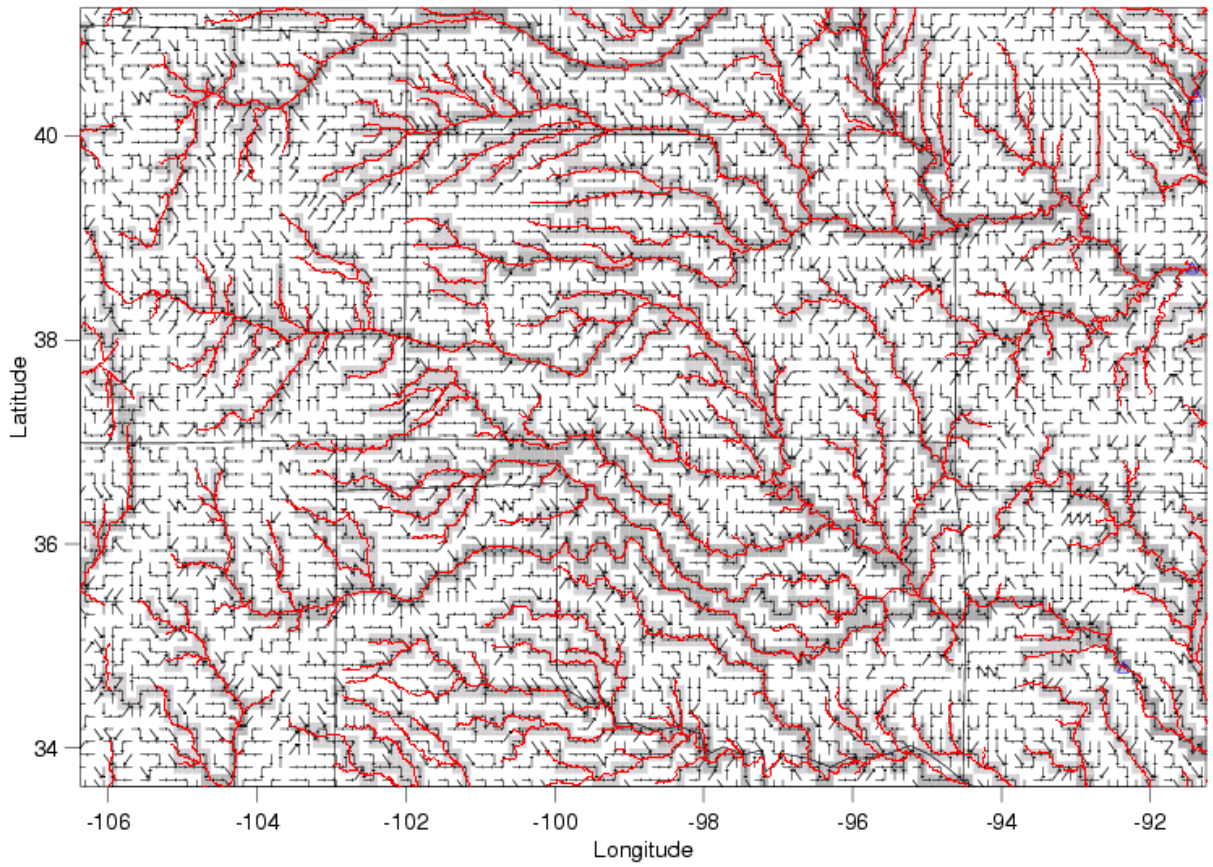

Figure 2. Example of the NLDAS 1/8 degree simulated river flow direction. The red lines indicate the location of the real rivers from the EPA river reach file RF1. The blue triangles are the basin outlets (Arkansas, Missouri, upper Mississippi River).

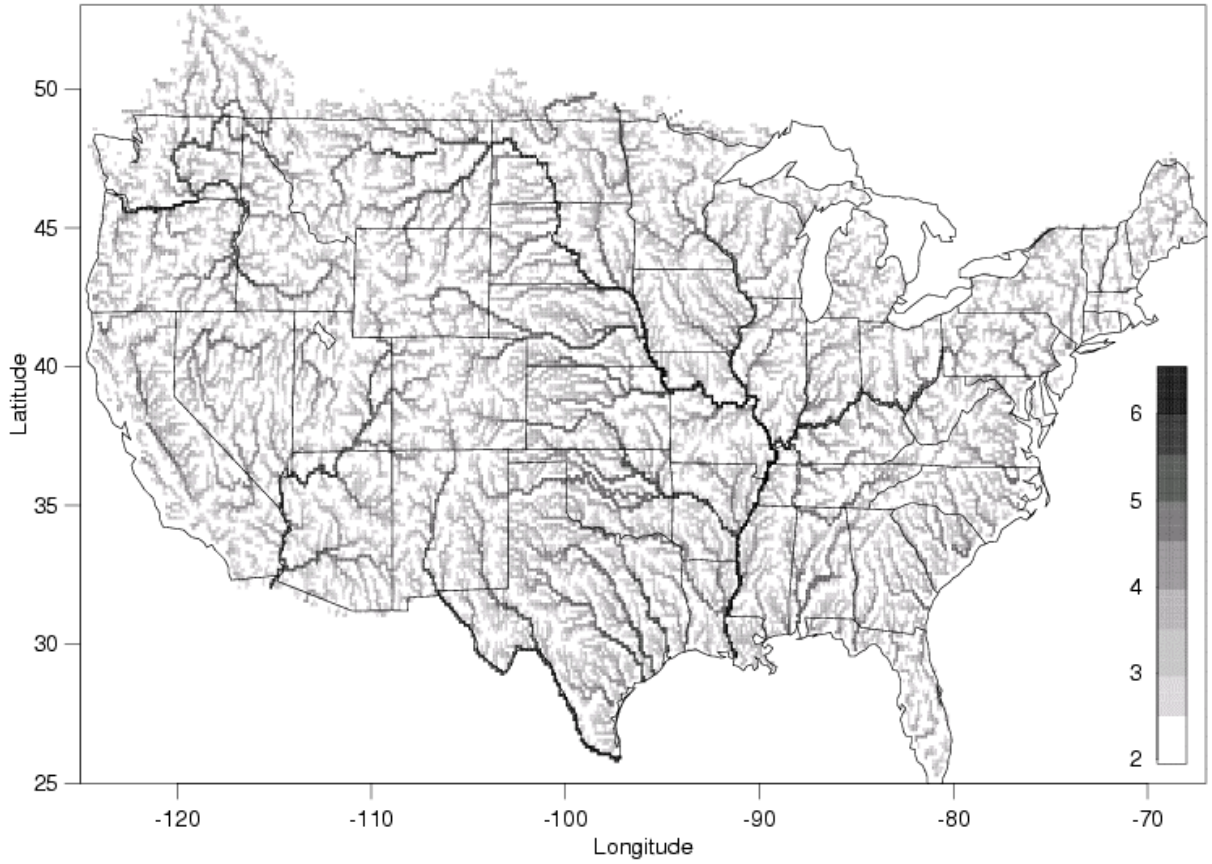

Figure 3. Logarithm to the base 10 (log10) of the upstream area in [km^2] for all grid cells.

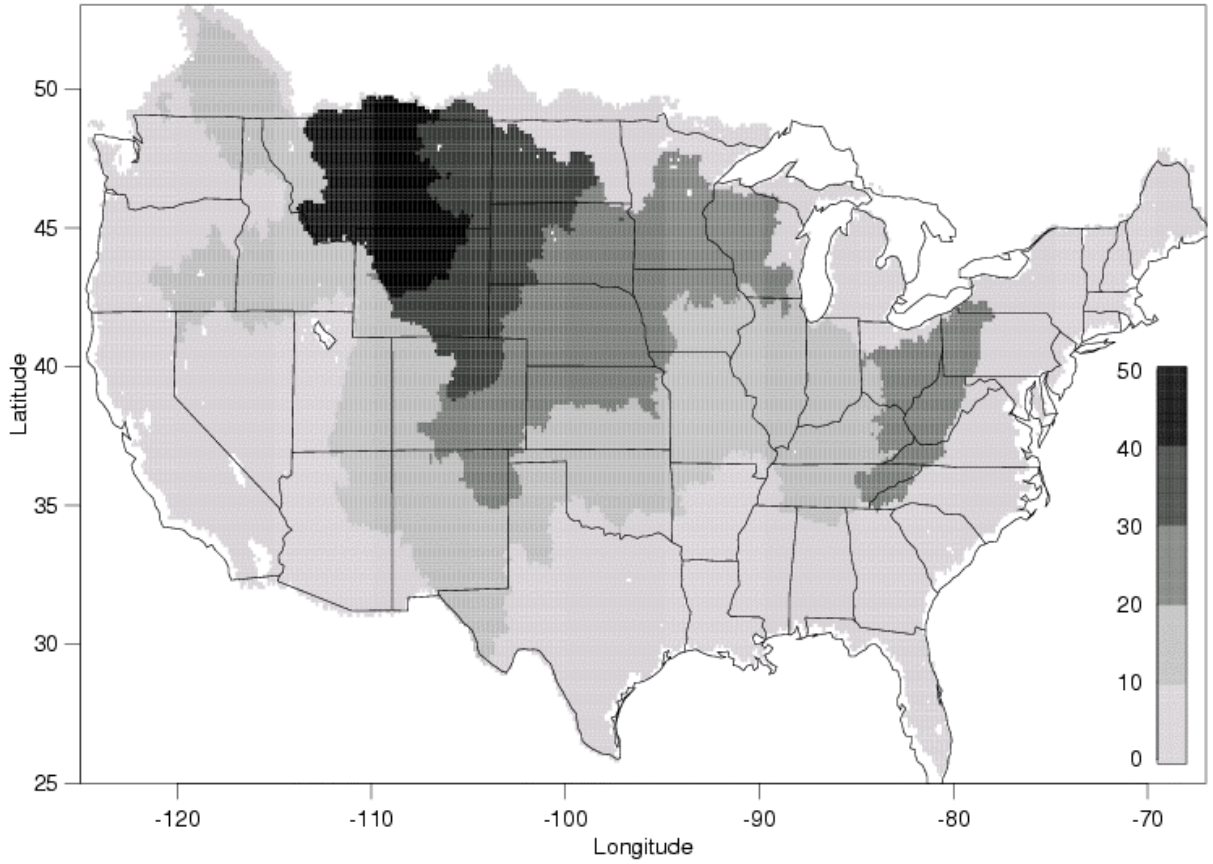

Figure 4. Distribution of travel times within the NLDAS domain for all grid cells to the outlet of each basin for the default parameters of the routing model.

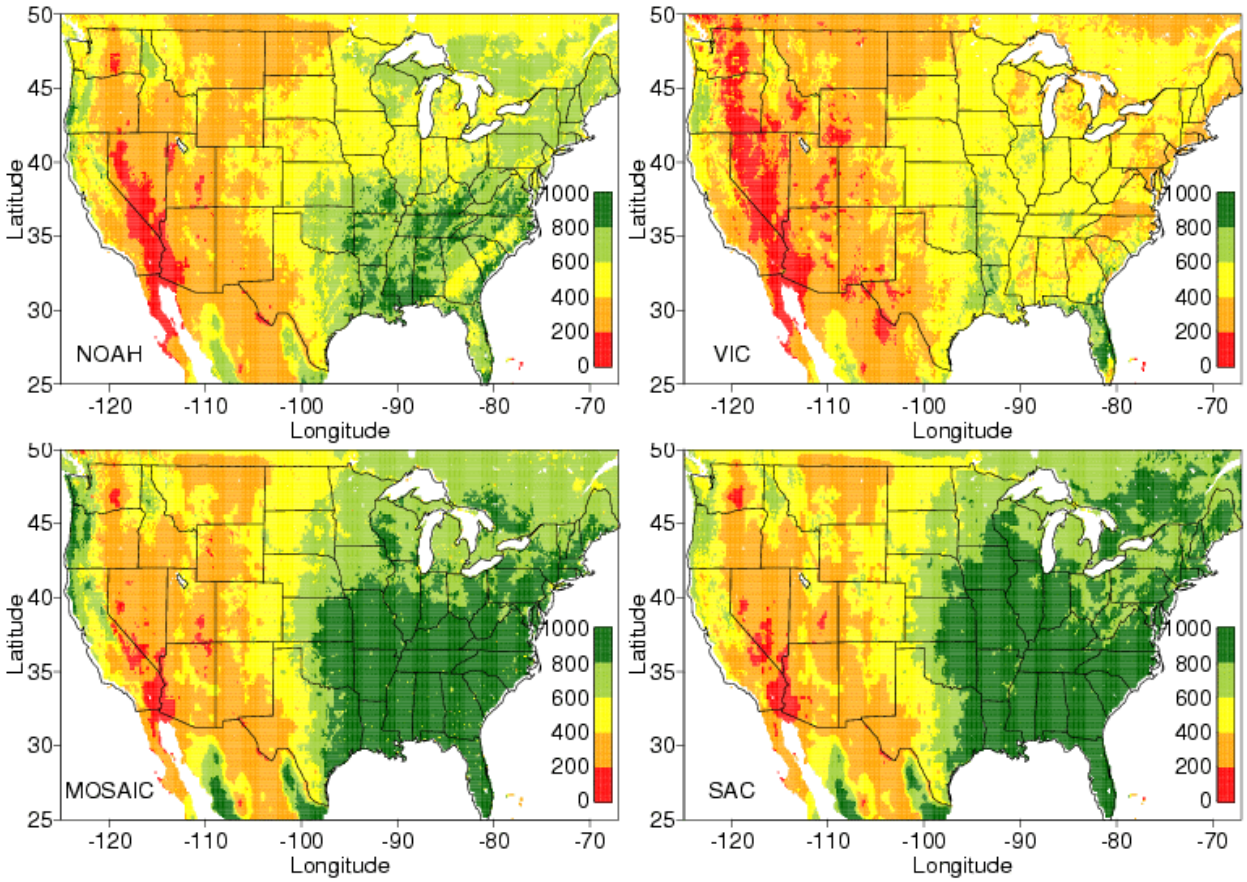

Figure 5: Mean annual evapotranspiration [mm/year] for the NLDAS domain for the time period October 1997 to September 1999.

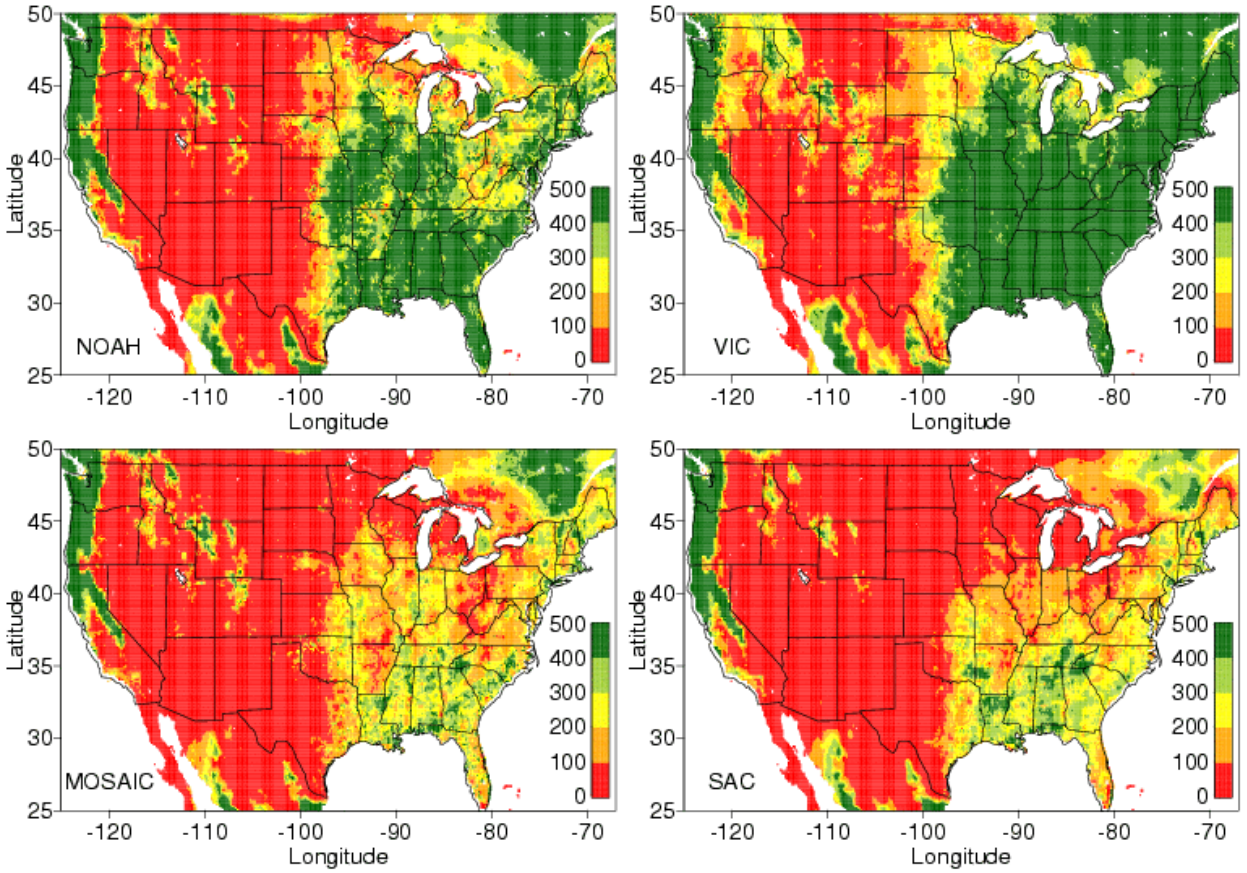

Figure 6: Mean annual runoff [mm/year] for the NLDAS domain for the time period October 1997 to September 1999.

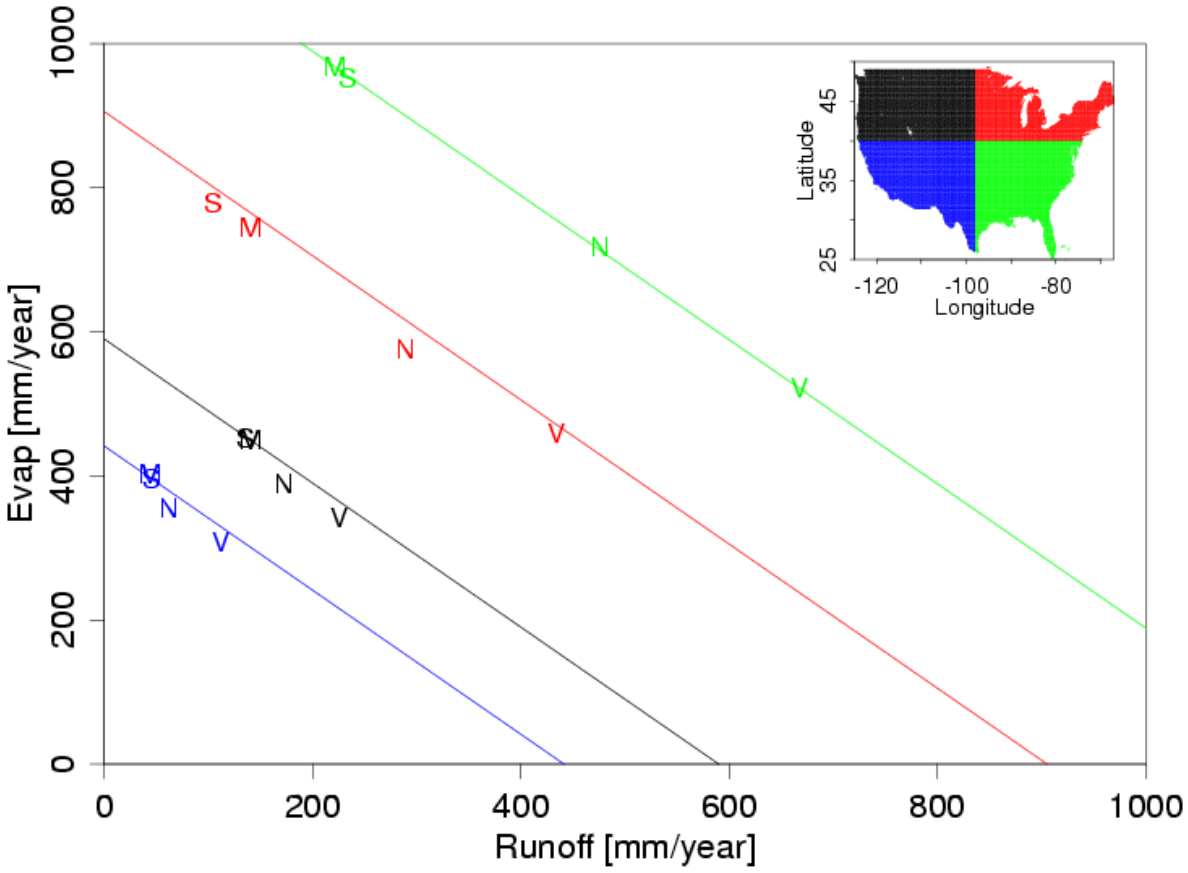

Figure 7: Partitioning of precipitation of the four NLDAS models (N-Noah, V-VIC, M-Mosaic, S-Sacramento) in four different quadrants for October 1997 to September 1999 into runoff (x-axis) and evapotranspiration (y-axis). The diagonal line is the mean of the precipitation in each area. Models whose symbol falls below the line have a positive storage change for the time period.

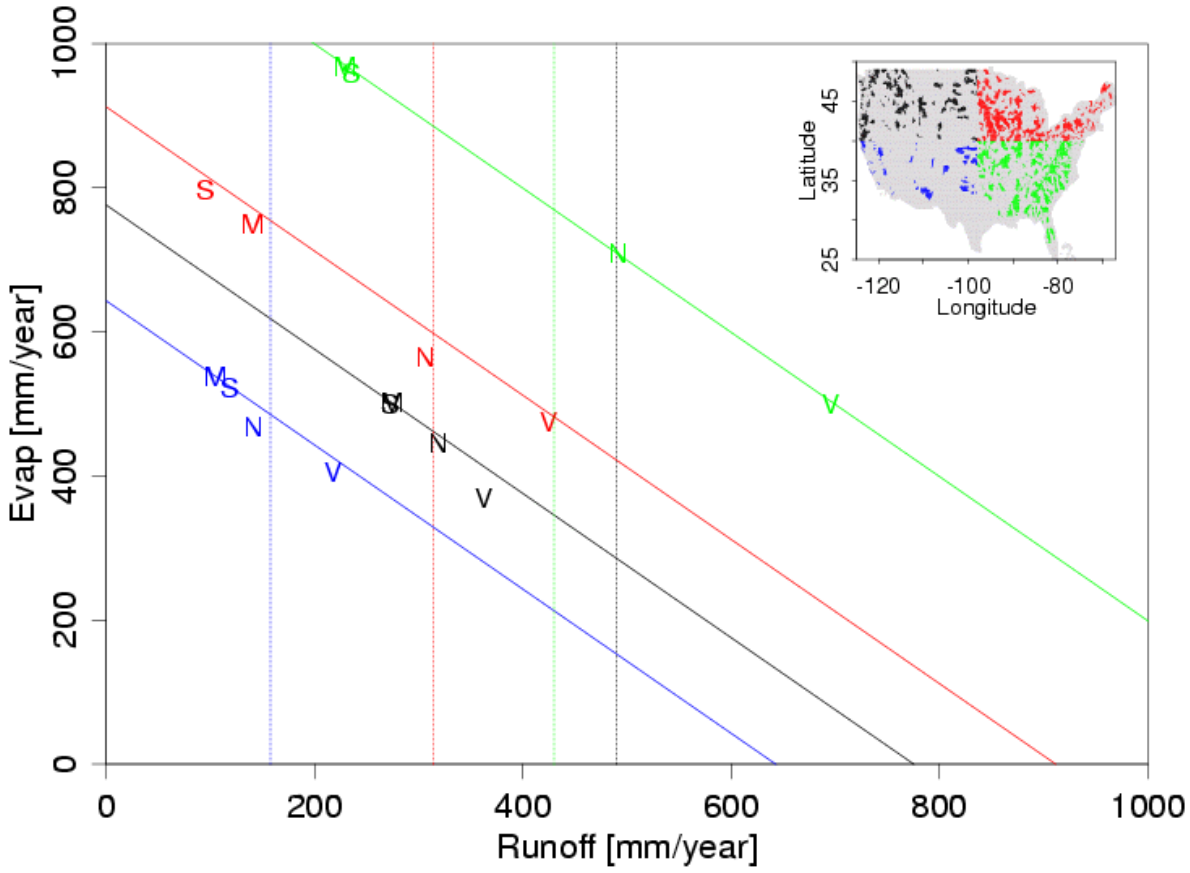

Figure 8: Partitioning of precipitation of the four NLDAS models (N-Noah, V-Vic, M-Mosaic, S-Sacramento) in all basins which fall into the four different quadrants for October 1997 to September 1999. The partitioning of precipitation is quite similar to Figure 7. Each vertical line is the averaged measured runoff from the 1145 small basins within the NLDAS area. The runoff in some areas varies by a factor of eight (Vic and Sacramento in the northeast). All models under-predict runoff in the northwest.

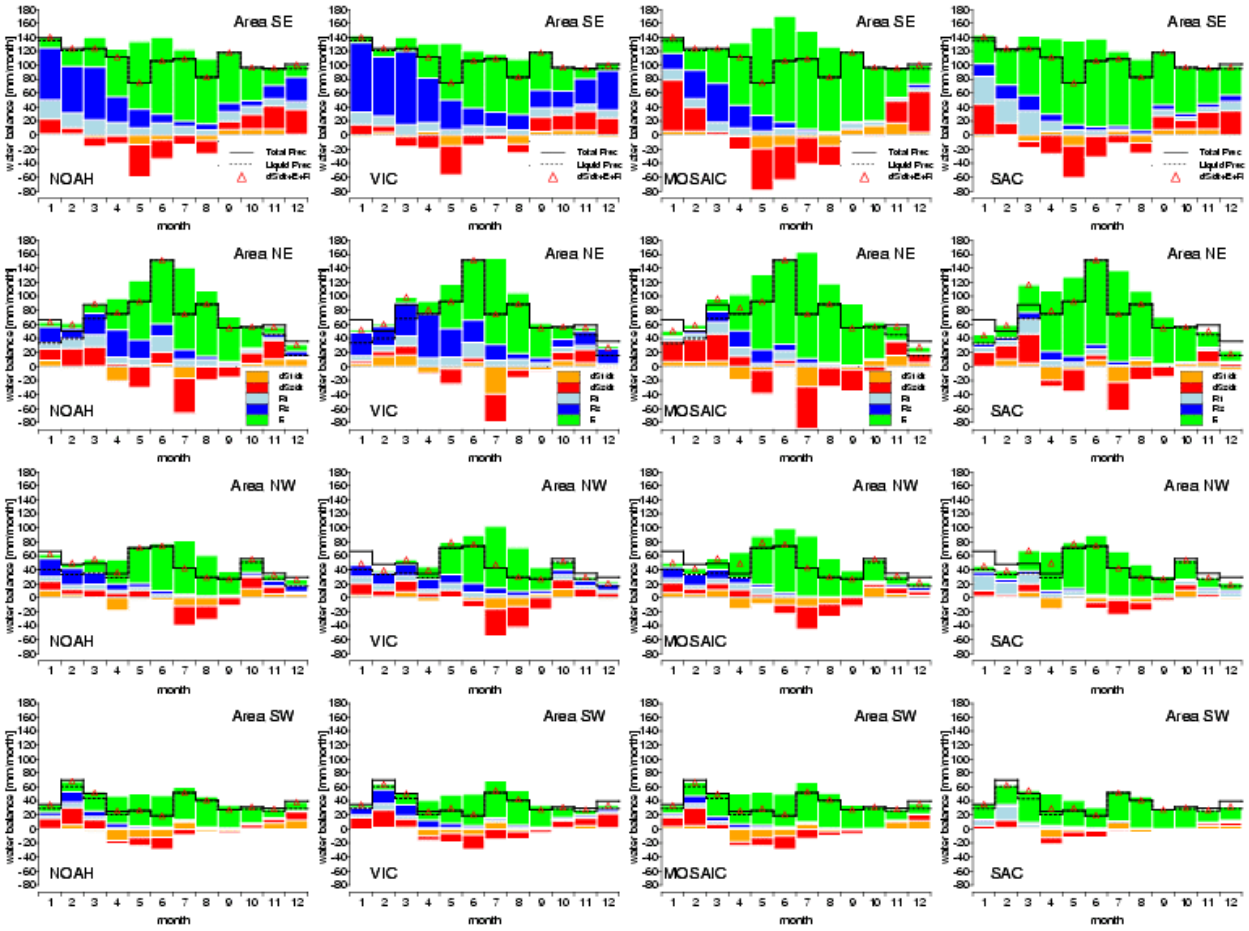

Figure 9: Monthly water balance of the NLDAS models in the four quadrants of Figs. 3 and 4 for the time period October 1997 to September 1998. Orange = upper soil storage change, red = lower soil storage change, light blue = surface runoff, dark blue = sub-surface runoff, green evapotranspiration, black solid line = precipitation, black doted line = liquid precipitation, red triangles = storage change + evapotranspiration + runoff. The deviation of the red triangles from the solid black line indicates snow processes. A red triangle is as much above (below) the solid black line as snow melts (accumulates).

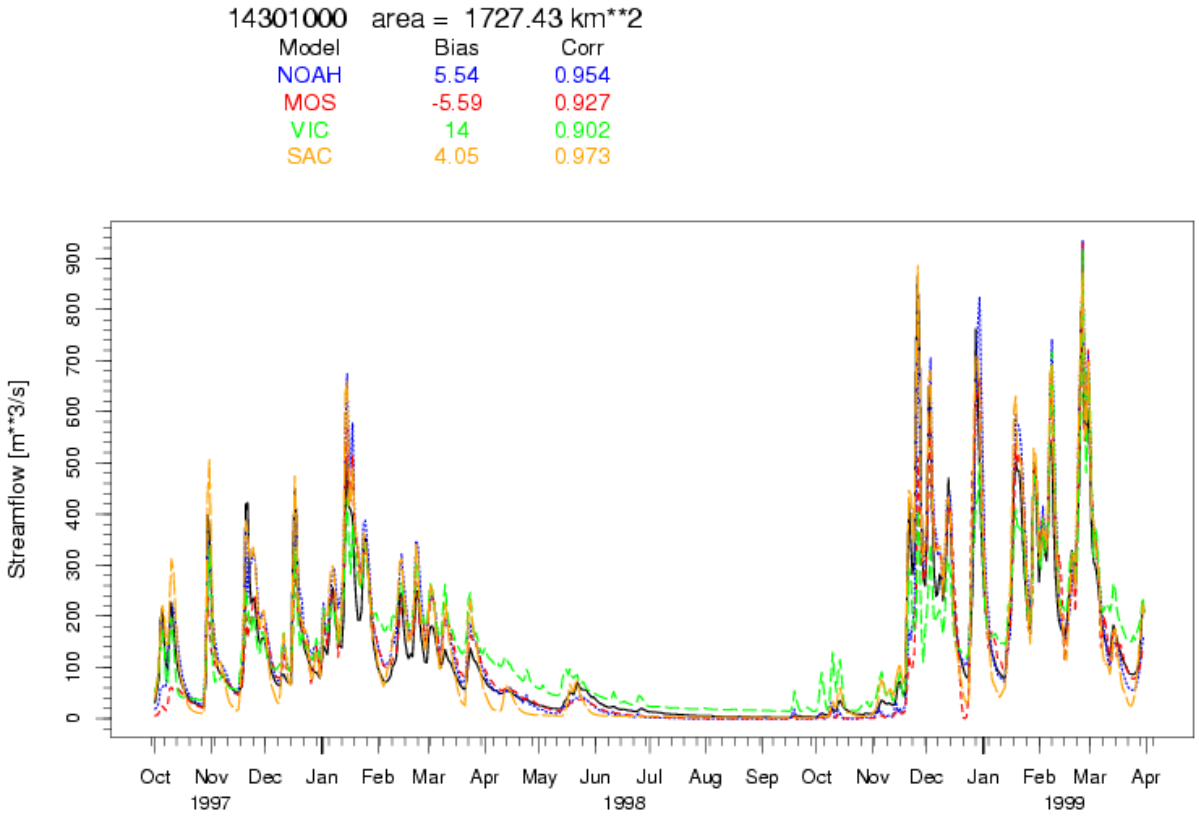

Figure 10. Observed (black curve) and modeled streamflow for the Nehalem River near Foss in Oregon.

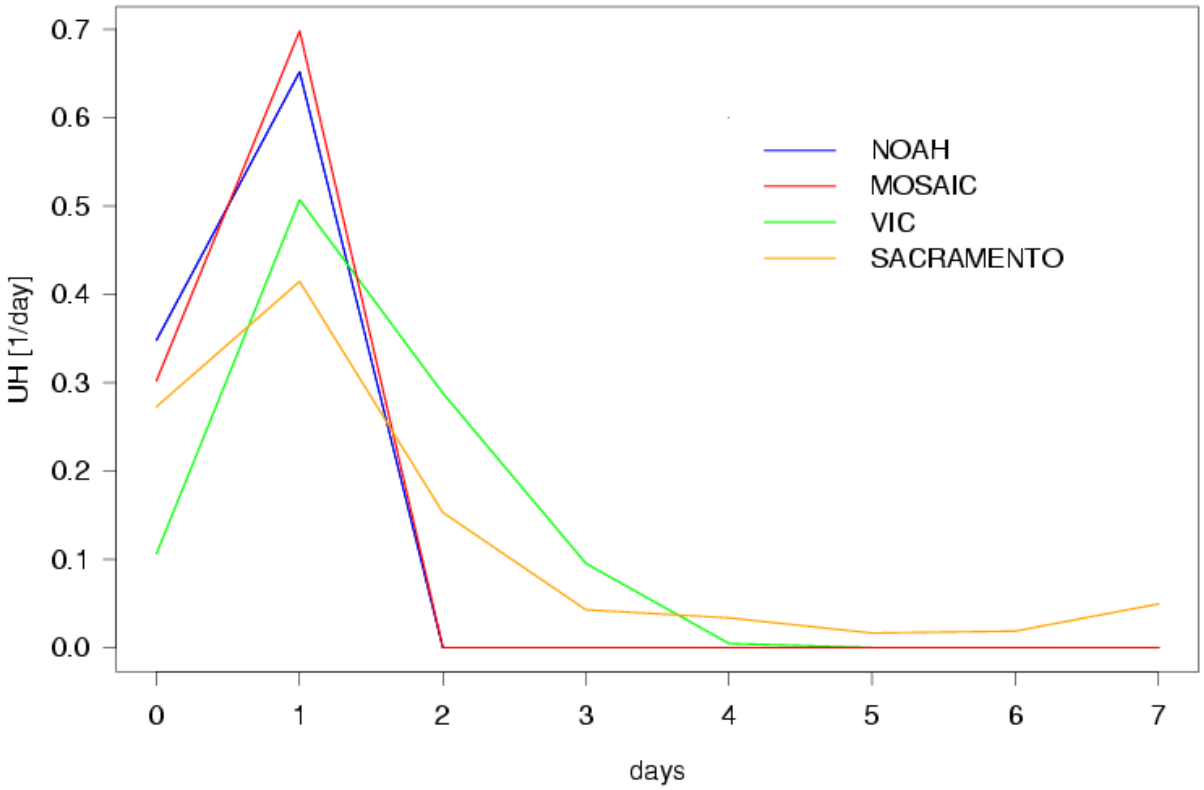

Figure 11. Derived Unit-Hydrograph for the 4 NLDAS models for the Nehalem River near Foss in Oregon. .

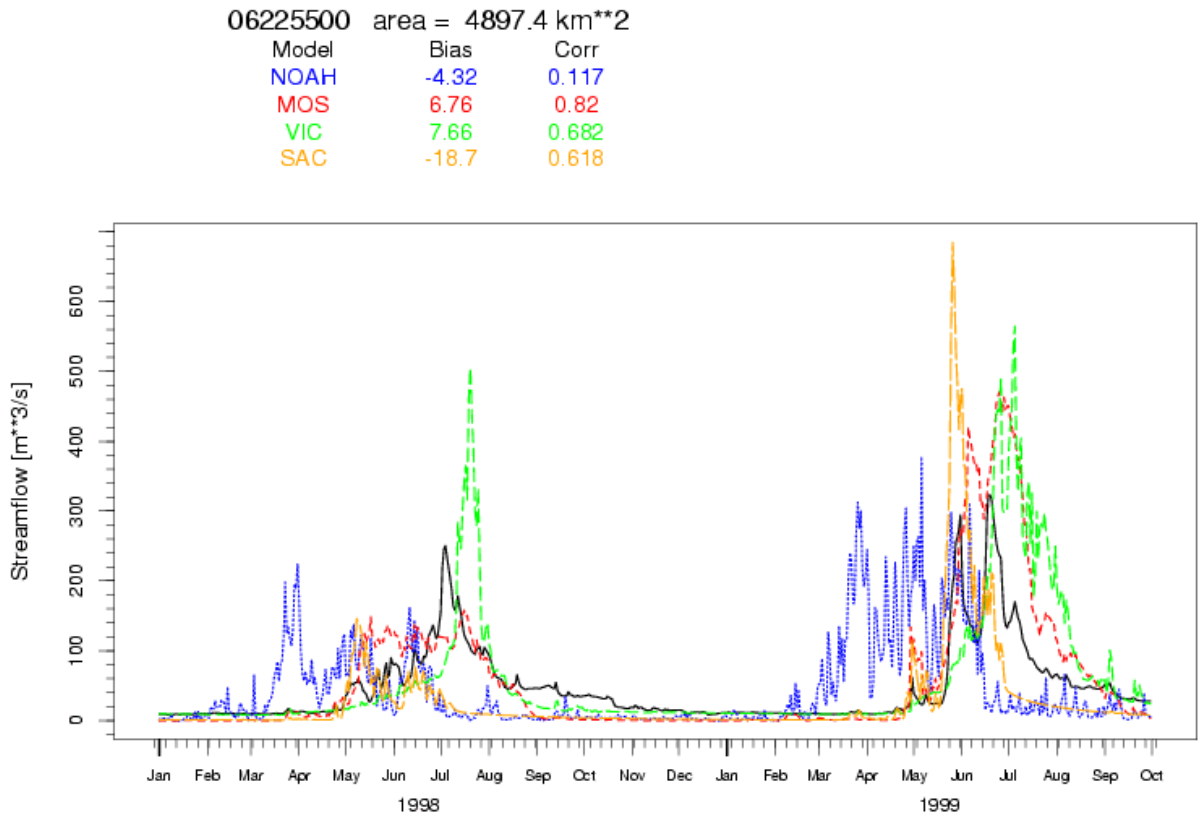

Figure 12. Observed (black curve) and modeled streamflow for the Wind River near Crowheart, Wyoming. The station is 1718 m above sea level, and shows the impact of snow melt on runoff.

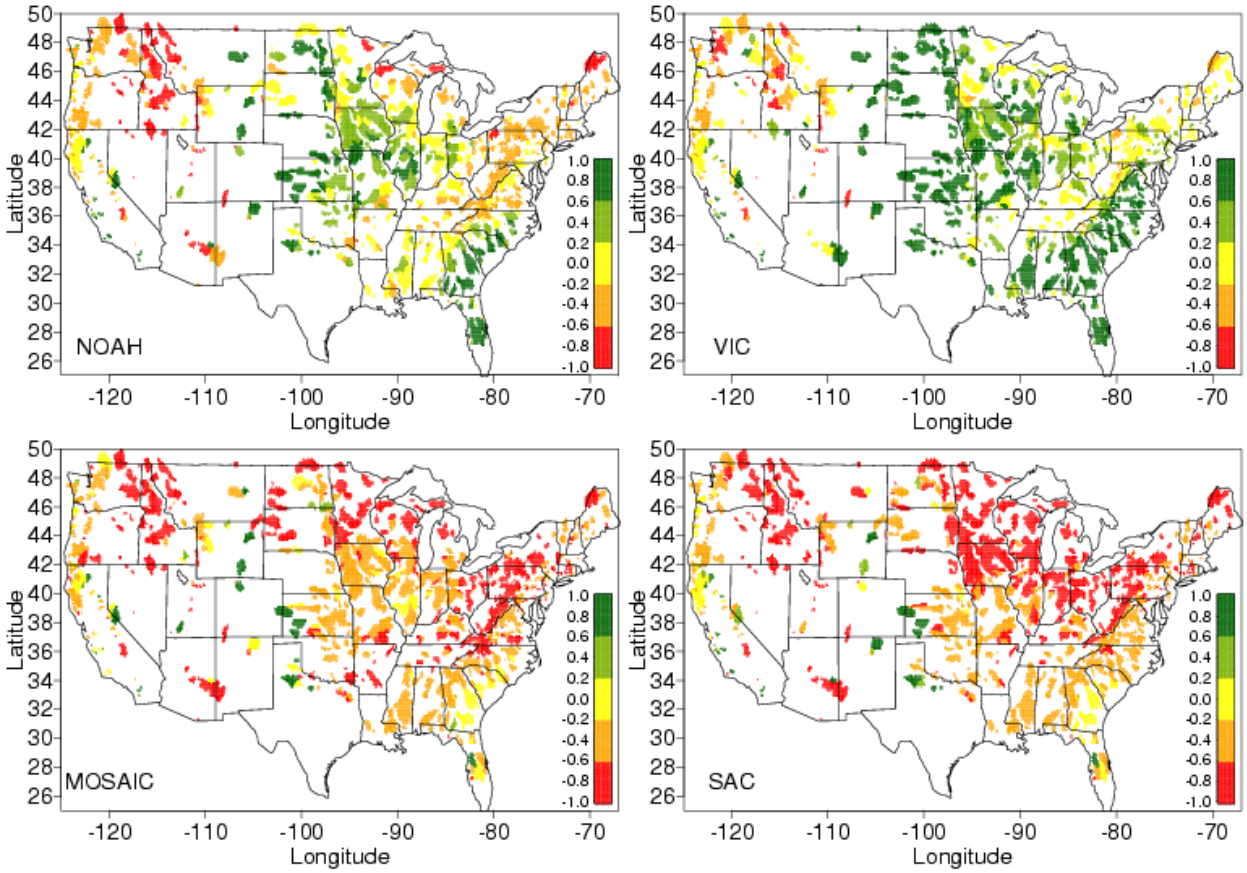

Figure 13. Relative runoff bias for the 4 NLDAS models for the time period 10/01/1997 to 09/30/1999. Notice that all models show less runoff than observed in the western snow covered areas.

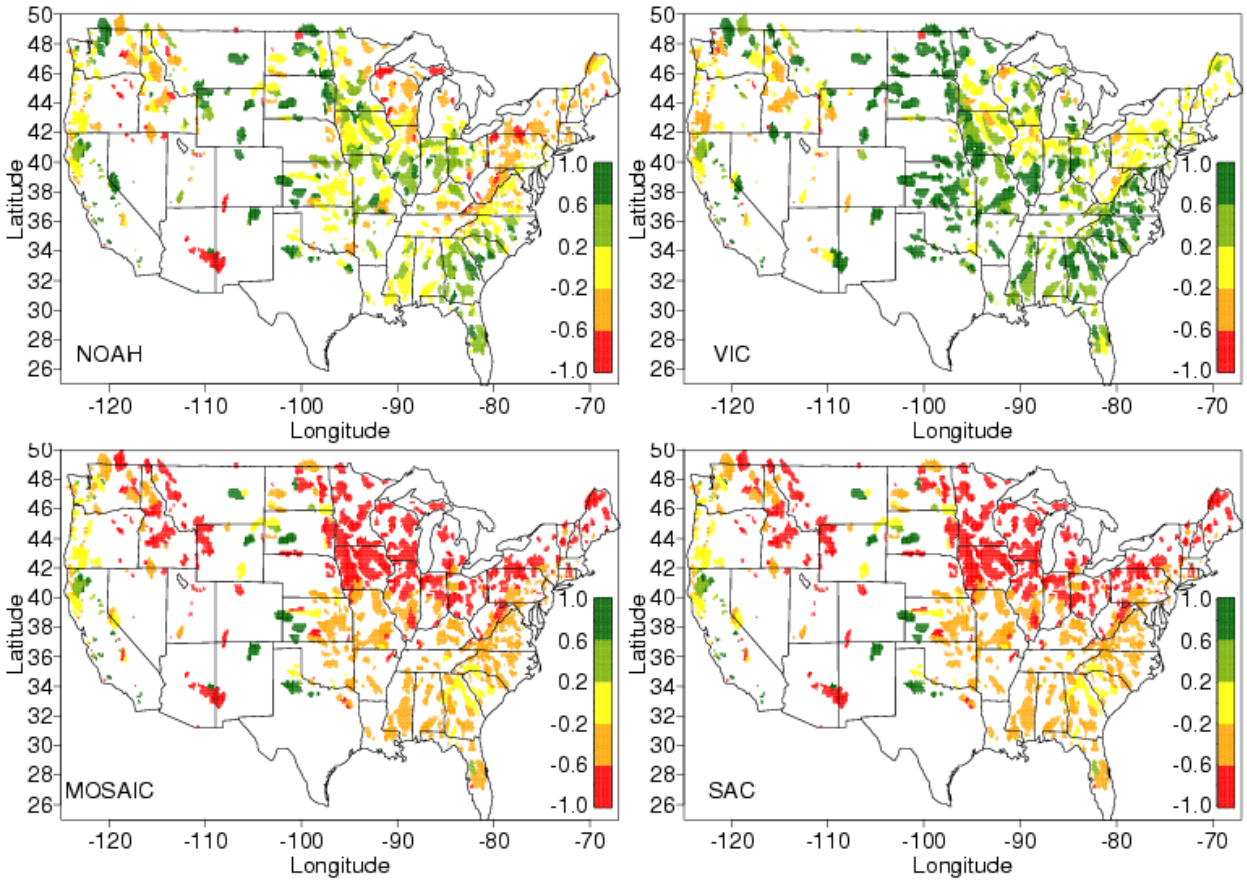

Figure 14: Cold season relative bias (October – March) in runoff for the time period 10/01/1997 to 09/30/1999.

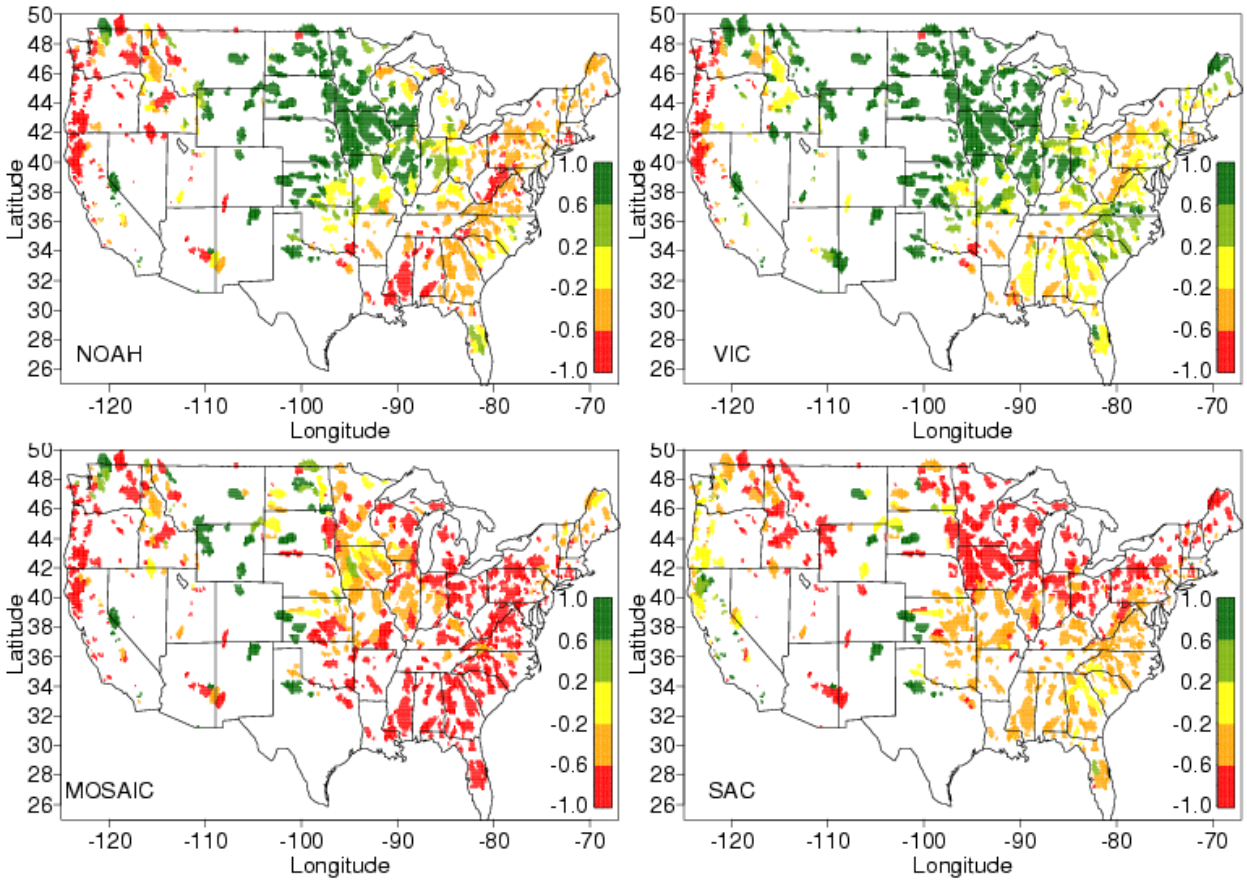

Figure 15: Warm season (April – September) relative runoff bias for the time period 10/01/1997 to 09/30/1999.

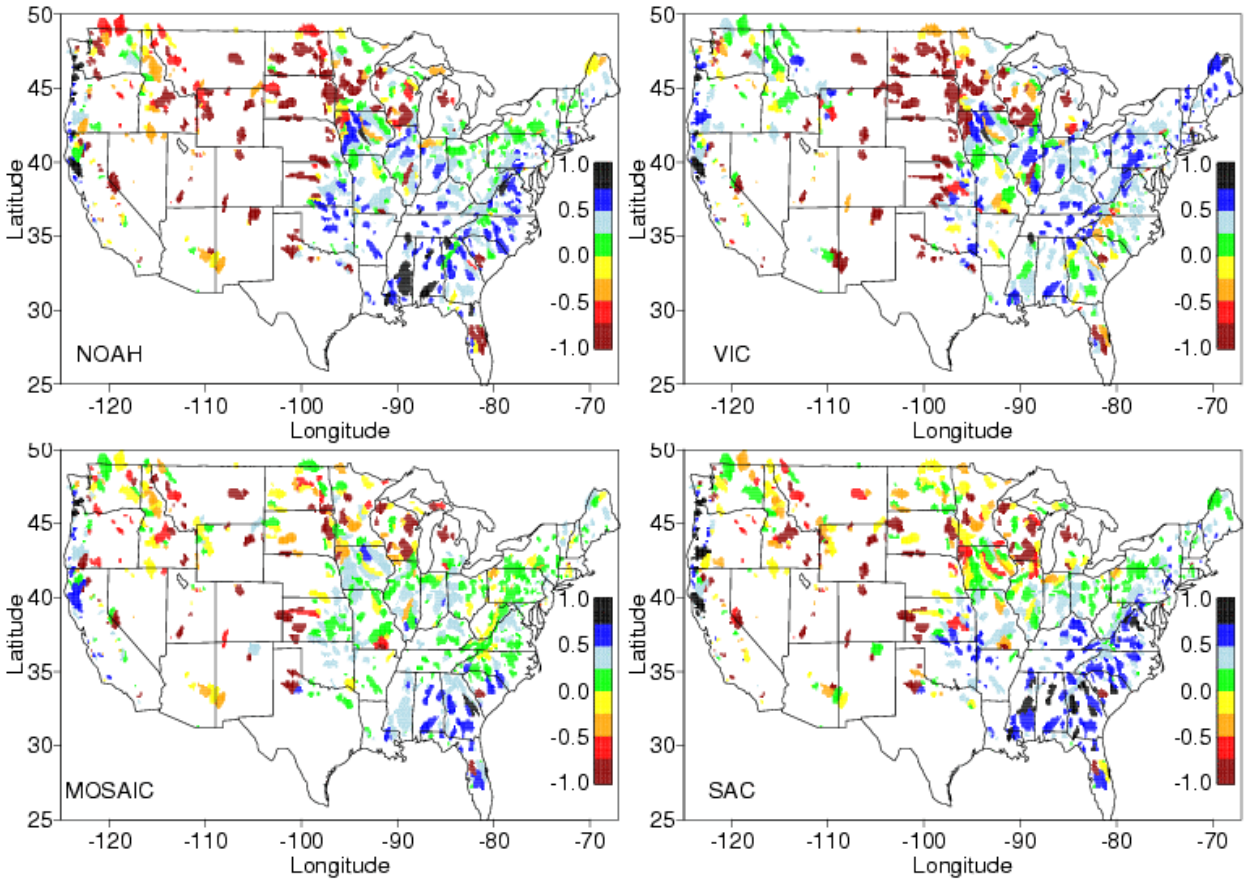

Figure 16. Nash-Sutcliffe Efficiency for the NLDAS domain for the time period 10/01/1997 to 09/30/1999 for daily mean modeled and measured data.

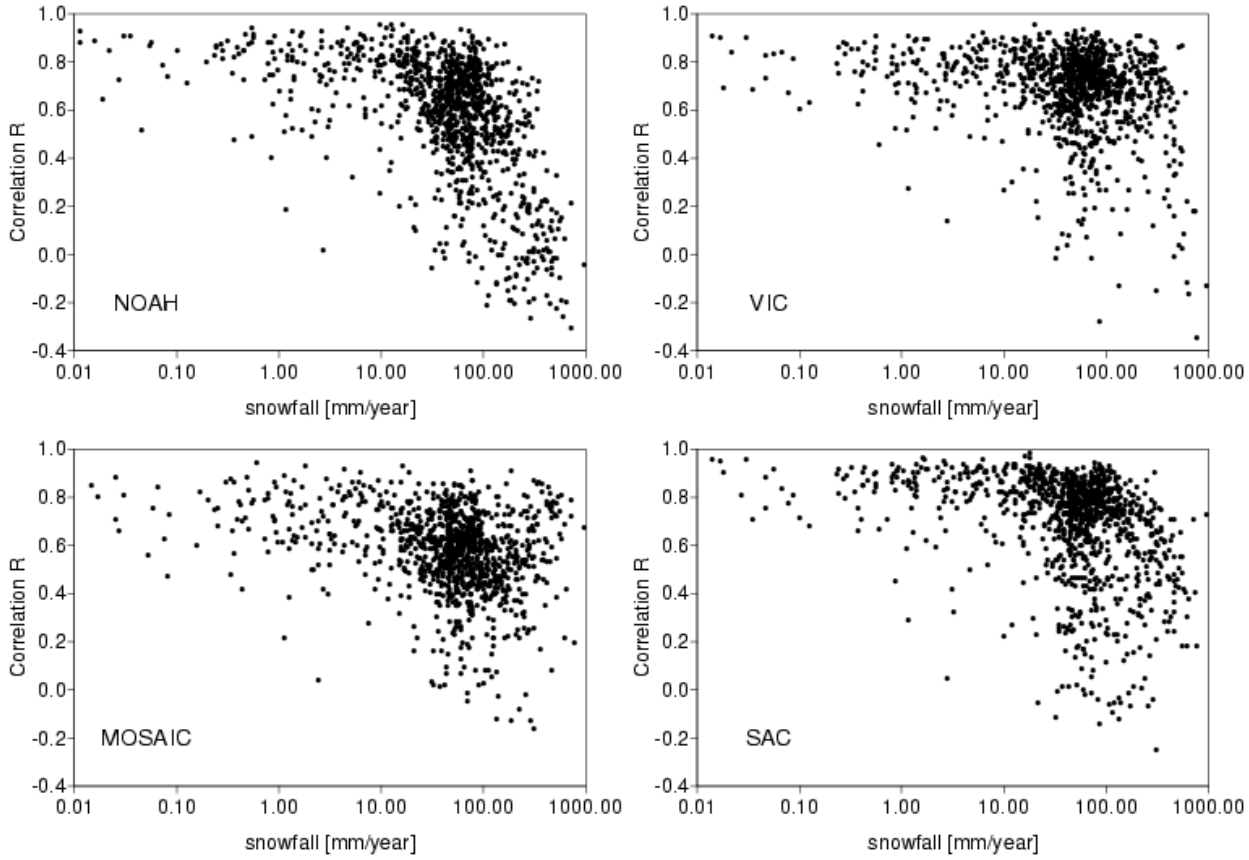

Figure 17. Relationship of mean annual snowfall in [mm/year] and the correlation of simulated and observed runoff. Each of the dots represents one of the 1145 basins.

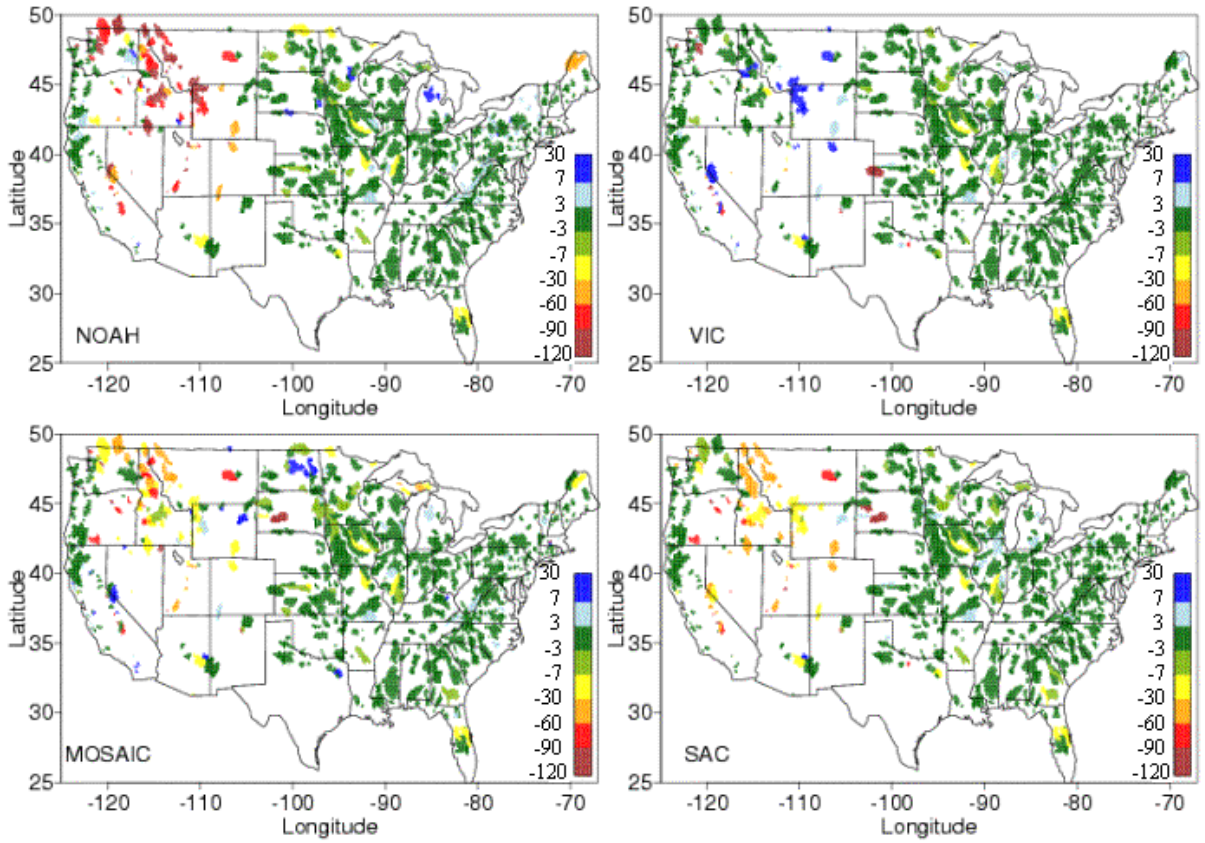

Figure 18. Time to peak delay [days] of streamflow for 1145 basins calculated as the maximum of the cross-correlation function of modeled and observed streamflow. Negative numbers indicate streamflow peaks earlier than observed peaks. In snow covered areas Noah, Mosaic, and the Sacramento model show consistently too early snowmelt and therefore produce streamflow too early (See Ming et al., 2002).

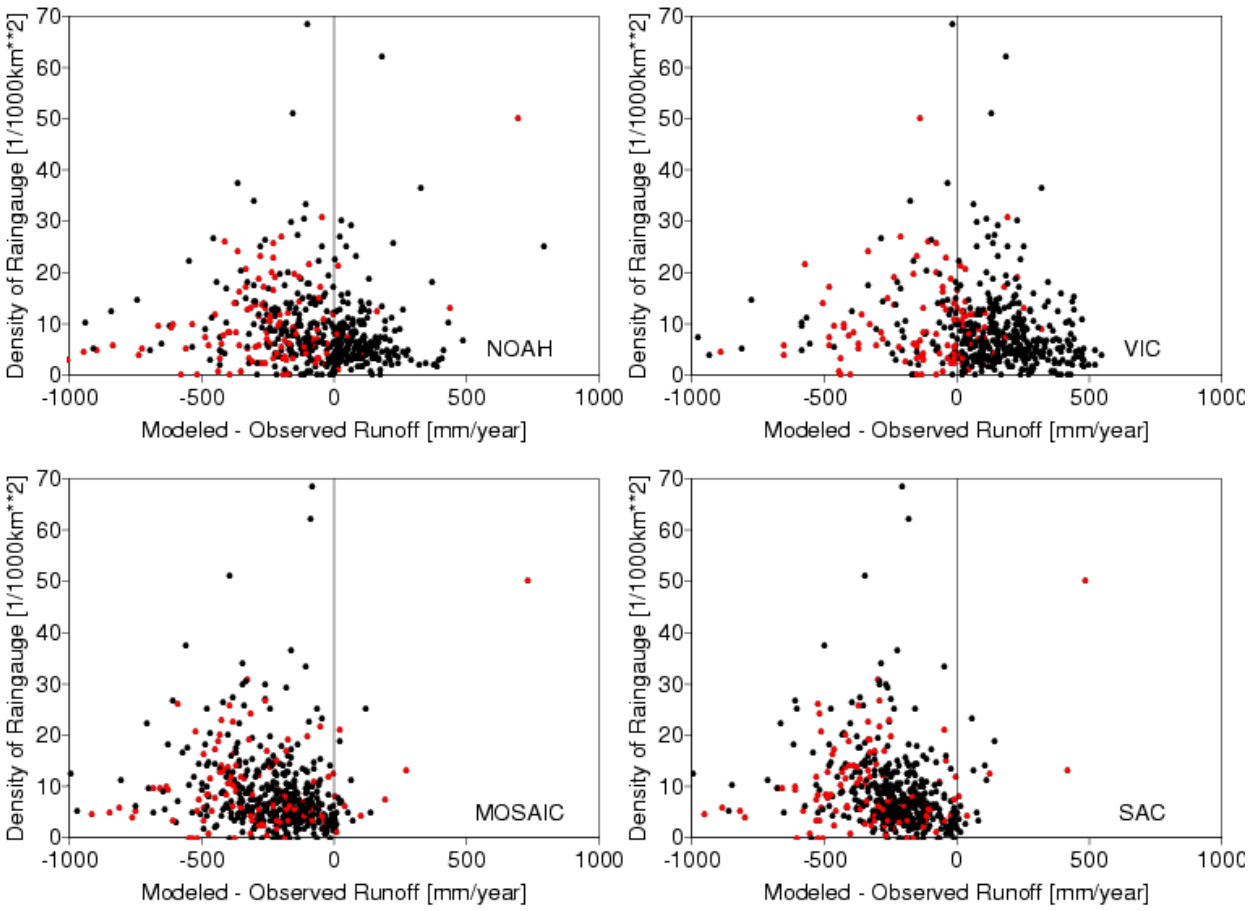

Figure 19. Density of the observing rain-gauges as a function of the annual runoff bias for the time period 10/01/1997 to 09/30/1999. Basins with more than 100 mm/year snow are shown as red dots. This analysis has been done the first time for the GSWP experiment (Oki et al, 1999) and showed that the absolute runoff bias was a function of the gauging station density. That relationship is also observed here, but to a much lesser degree.

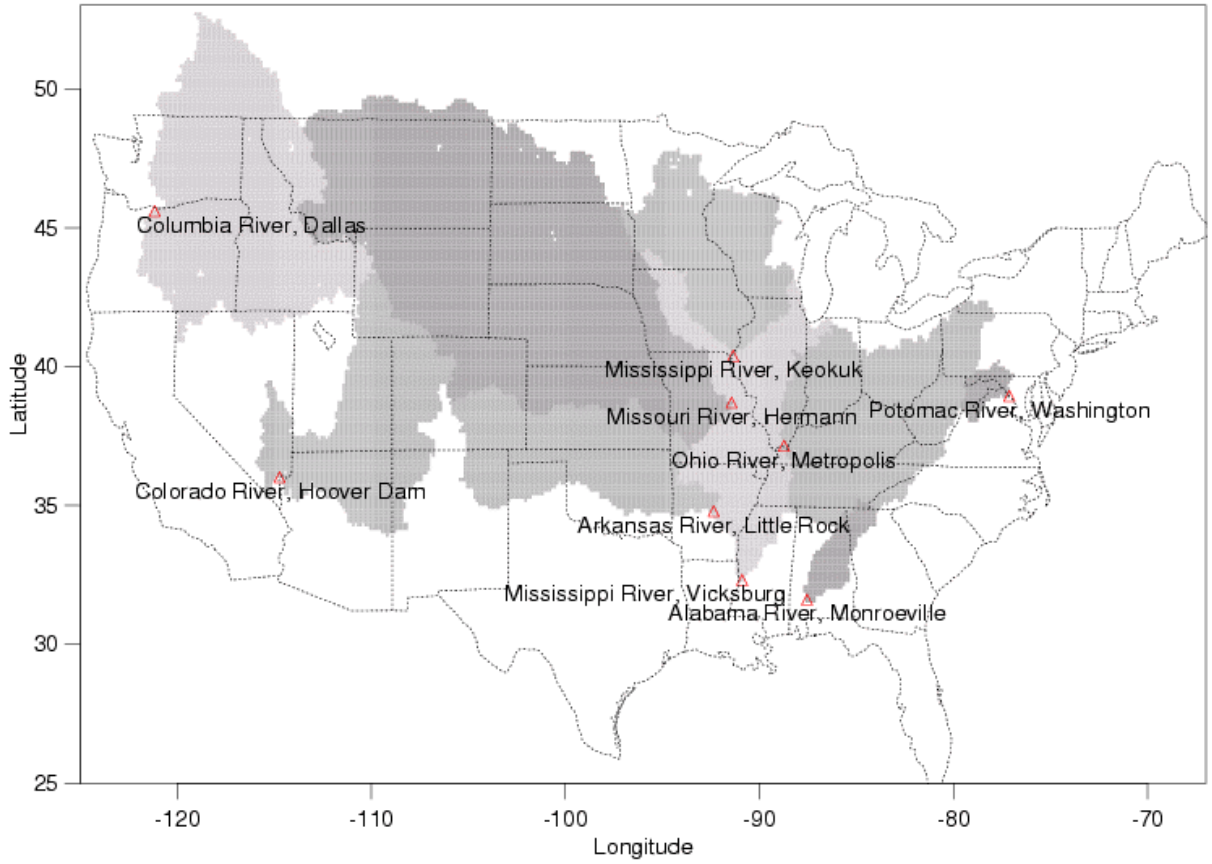

Figure 20. Location of the nine major basins and USGS gauging stations used for this intercomparison and validation study within the US.

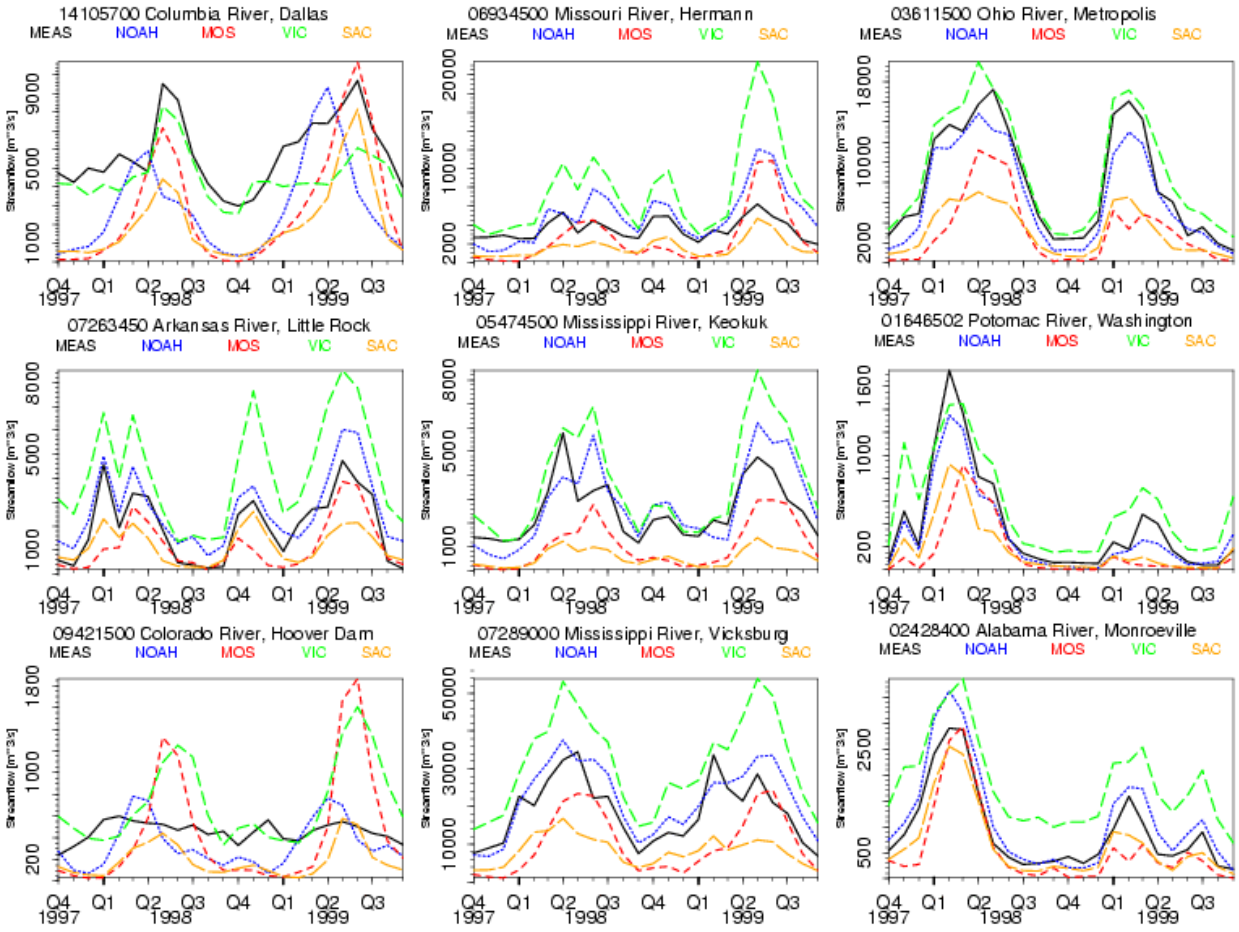

Figure 21. Monthly mean streamflow for the nine major basins from Figure 20 for the time period 10/1997 to 09/1999.