ELSEVIER

Global and Planetary Change 19 (1998) 161–179

GLOBAL AND PLANETARY CHANGE

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

Dag Lohmann [a,*], Dennis P. Lettenmaier [b], Xu Liang [a,1], Eric F. Wood [a], Aaron Boone [c,l], Sam Chang [d], Fei Chen [e,2], Yongjiu Dai [f], Carl Desborough [g], Robert E. Dickinson [h], Qingyun Duan [i], Michael Ek [j], Yeugeniy M. Gusev [k], Florence Habets [l], Parviz Irannejad [m], Randy Koster [n], Kenneth E. Mitchell [e], Olga N. Nasonova [k], Joel Noilhan [l], John Schaake [i], Adam Schlosser [o], Yaping Shao [m], Andrey B. Shmakin [p], Diana Verseghy [q], Kirsten Warrach [r], Peter Wetzel [c], Yongkang Xue [s,3], Zong-Liang Yang [h], Qing-cun Zeng [f]

[a] *Department of Civil Engineering and Operations Research, Princeton University, Princeton, NJ, USA*
[b] *Department of Civil Engineering, University of Washington, Seattle, WA, USA*
[c] *Mesoscale Dynamics and Precipitation Branch, NASA / GSFC, Greenbelt, MD, USA*
[d] *Air Force Research Laboratory, Hanscom AFB, Hanscom, MA, USA*
[e] *Environmental Modeling Center (NOAA / NCEP), Camp Springs, MD, USA*
[f] *Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China*
[g] *Climatic Impacts Center, Macquarie University, Sydney, Australia*
[h] *Institute of Atmospheric Physics, University of Arizona, Tucson, AZ, USA*
[i] *Office of Hydrology, NOAA / NWS, Silver Spring, MD, USA*
[j] *Oregon State University, Corvallis, OR, USA*
[k] *Institute of Water Problems, Moscow, Russian Federation*
[l] *Meteo-France / CNRM, Toulouse, France*
[m] *Centre for Advanced Numerical Computation in Engineering and Science, The University of N.S.W., NSW 2052, Australia*
[n] *Hydrological Sciences Branch, NASA / GSFC, Greenbelt, MD, USA*
[o] *NOAA / GFDL, Princeton, NJ, USA*
[p] *Institute of Geography, Moscow, Russian Federation*
[q] *Climate Research Branch, Atmospheric Environment Service, Toronto, Ontario, Canada*
[r] *GKSS Research Center, Geesthacht, Germany*
[s] *Center for Ocean–Land–Atmosphere Studies, Calverton, MD, USA*

Received 8 September 1997; accepted 9 February 1998

* Corresponding author.
[1] Now at JCET UMBC/NASA, Climate and Radiation Branch, Code 913, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA.
[2] Currently at NCAR, Boulder, CO, USA.
[3] Currently at Dept. of Geography, Univ. of Maryland, College Park, MD 20742, USA.

PII: S0921-8181(98)00046-0

Abstract

The water-balance components of 16 Soil–Vegetation Atmospheric Transfer (SVAT) schemes were evaluated by comparing predicted and observed streamflow, predicted evapotranspiration and evapotranspiration inferred from an atmospheric moisture budget analysis, and soil moisture storage changes for a seven-year period (1980–1986) using data from the Red–Arkansas River basins of the Southern Great Plains of the USA. The evaluations support the following suggestions: (a) The mean annual runoff of all models follows, at least generally, the strong climatic East–West gradient of precipitation, although most models predict too much runoff in the dry part of the basin. (b) The mean monthly storage change tends to be underestimated, even though all models capture reasonably well the seasonality of the evapotranspiration. (c) The wide range of conceptualizations used for generation of surface and subsurface runoff strongly affect runoff generation on seasonal, and shorter, time scales. Model responses to summer precipitation ranged from almost no summer runoff (one model) to the (more common) situation of persistent overprediction of summer runoff, especially in the driest part of the basin. (d) All models tended to underpredict evapotranspiration in summer and overpredict in winter. (e) Model-derived mean seasonal cycles of changes in soil moisture storage are qualitatively similar to those inferred from observations, but most models do not predict the decrease in April soil moisture storage and the increase in October that is inferred from observations.

Keywords: PILPS; water balance; routing model; continental river basin modeling; Red–Arkansas River basin

1. Introduction

This paper is Part 3 of a three-part series summarizing the results of Phase 2(c) of the Project for Intercomparison of Land-surface Parameterization Schemes (PILPS). The goal of PILPS is to improve the parameterization of the land-surface schemes used in climate- and weather-prediction models. Details of the project are described by Henderson-Sellers et al. (1993, 1995). PILPS has facilitated a series of experiments: initially, point evaluations using model output as forcings in Phase 1 and, then, progressing to model evaluations using field data in Phase 2.

In PILPS Phase 2(c), the design of which is described in detail in Part 1 of this series (Wood et al., this issue), offline simulations from 16 land-surface schemes were compared to observations of streamflow and basin-scale evapotranspiration in the Red–Arkansas River basin. Part 2 of this series (Liang et al., this issue) evaluates the performance of the models, in terms of their predictions of surface energy fluxes (including latent heat or, equivalently, evapotranspiration). The focus of this paper is the evaluation of the water balances simulated by the 16 models that participated in PILPS Phase 2(c).

The PILPS Phase 2(c) experimental design was, briefly, as follows. Participants were provided with surface atmospheric forcings, at a one-degree scale, for 61 grid cells that constitute the Red–Arkansas River basin. They were also provided with forcings for six small catchments (drainage areas ranging from 100 to 1000 km^2), and, for three of these ('calibration catchments'), they were also provided with the coincident streamflow observations. The participants were allowed (and encouraged) to specify model parameters in such a way as to best 'fit' the observed streamflow series. Model simulations of surface energy and moisture fluxes (including streamflow) were then evaluated for three verification catchments of similar size and location to those used for calibration. Energy and moisture fluxes were also evaluated at the (much larger) scale of the entire Red–Arkansas River basin. Participants were not provided streamflow observations for either the validation catchments or for the major tributaries of the Red–Arkansas river system, at which model simulations were evaluated.

In evaluating the surface water budgets of different models using large-scale streamflow data, the problem immediately arises that most land-surface schemes simulate runoff but not streamflow. For most models, runoff is simply an excess of precipitation over evapotranspiration and local moisture storage change which combines various terms (includ-

ing, e.g., fluxes to or from groundwater, direct surface runoff, and baseflow) that may eventually be evidenced as streamflow. Therefore, a routing model is needed to translate model-simulated runoff into streamflow. Effectively, the routing model can be considered a surrogate for all of the physical processes that control movement of water into and through the stream-channel system to the basin outlet, and/or other locations where streamflow is measured.

The general form of the routing model used is based on assumptions of linearity and time invariance (see Lohmann et al., 1996 for details). The routing model requires the specification of a set of parameters, which were estimated independently for each of the land surface schemes (however, the form of the model was the same for all of the land-surface schemes). The routing model uses baseflow separation techniques, which are well established in the hydrologic literature (see, e.g., Linsley et al., 1975) to determine the parameters for the routing within each of the 1°-by-1° grid cells. Similar routing models have been used for other large scale applications by Abdulla (1995) and Wetzel (1994) or flood forecasts (Todini, 1996) and references therein.

Zhao (1997) estimated the atmospheric budget for a region which roughly corresponds to the Red–Arkansas River basin (see Fig. 1), using radiosonde data. The estimated basin-scale evapotranspiration was compared with the results of the land-surface schemes. The residuals from the measurements and estimations of the terrestrial and atmospheric water balance is the storage change of water at the land surface. It will be shown that the utilization of this storage has a profound impact on the evapotranspiration results in PILPS Phase 2(c) simulations.

2. Water-balance climatology

The Red and Arkansas River basins are located in the southern part of the Great Plains of the USA. The drainage areas are 409,273 km^2 for the Arkansas River and 156,978 km^2 for the Red River. For the years 1980–1986 the mean annual precipitation for the 61 grid cells was 767 mm/yr, ranging from about 1400 mm/yr in the southeastern part of the basin to around 200 mm/yr in the western arid part. The mean annual runoff climatology (GGHYDRO, Cogley, 1991) tends to follow this gradient (see Fig. 2). Runoff is highest in the eastern part of the basin (> 400 mm/yr) and lowest in the western part (< 5 mm/yr).

Fig. 1 shows schematically the simplified stream network that connects the 61 1°-by-1° grid cells, and is used to route the grid-generated model runoff into streamflow. Each grid-cell unit-area runoff was multiplied by the cell area to compute the grid-cell discharge. Since the basin boundary of the Red–Arkansas River basin deviates slightly from the boundary implied from the 61 grid cells, weights, ranging from 0.3 to 1.25, were assigned to the discharge generated in the boundary cells. This assures that the simulated streamflow is not affected by area discrepancy. The adjusted boundary changes the mean annual precipitation slightly, from 767 to 755 mm/yr.

The model intercomparisons described in this paper were based on comparisons of model-simulated streamflow with naturalized streamflow data provided by the Tulsa office of the U.S. Army Corps of Engineers. The naturalized streamflow data have all the effects of reservoirs and diversions removed. For the period 1980–1986, the sum of the mean observed (at USGS gages at Little Rock and Shreveport on the Arkansas and Red rivers, respectively) and naturalized flows (112 mm/yr) were within about 2%. For this period, the annual average runoff ratio

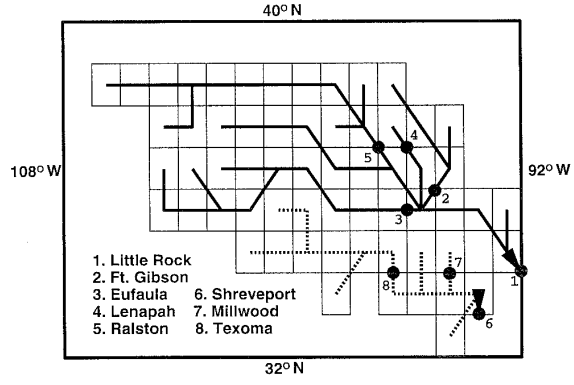

Fig. 1. Red–Arkansas River basin schematized routing directions and streamflow evaluation sites. Atmospheric budgets were computed for the region bounded by 92°W to 108°W longitude and 32°N to 40°N latitude.

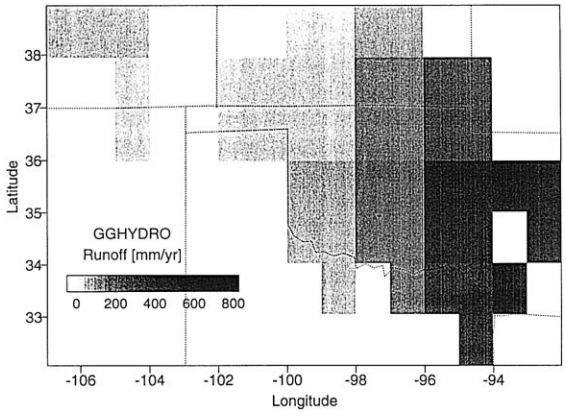

Fig. 2. Runoff climatology for the 61 1°-by-1° grid cells which form the Red–Arkansas River basin (GGHYDRO, Cogley, 1991). The value for 106.5°W, 38.5°N was adjusted to its neighboring grid cell, as streamflow data indicate that the climatology is not representative for that part of the basin.

(defined as runoff divided by precipitation) was 14.8%.

An atmospheric moisture budget calculated by Zhao (1997) for the Red–Arkansas River basin was used by Liang et al. (this issue) to evaluate the model-simulated evapotranspiration. As discussed in Section 3, the estimated mean water vapor convergence for the years 1980–1986 was 108 mm/yr leading to a mean annual evapotranspiration, based on the atmospheric budget calculation, of 658 mm/yr. This resulted in the seven-year annual average naturalized streamflow exceeding the estimated water vapor convergence by 4 mm/yr, or about 0.5% of the annual precipitation—a relatively small value, given that these estimates are independently derived. As noted in Liang et al. (this issue), the mean annual atmospheric budget estimate of evapotranspiration is thought to be accurate to within 5%, and monthly evapotranspiration (during the warm-season months) is thought to be accurate to within 10%.

3. Water-balance intercomparison method

This section briefly describes the governing water-balance equations, and the methodology by which measured data were compared to model output.

3.1. Water-balance equations

The governing equations of the atmospheric water balance can be written as (see, e.g., Starr and Peixoto, 1958 or Peixoto, 1973):

$$\frac{\mathrm{d}W}{\mathrm{d}t} + \nabla_{\mathrm{H}} \cdot \vec{\boldsymbol{Q}} = E - P \tag{1}$$

where W is the precipitable water, $\nabla_{\mathrm{H}} \cdot \vec{\boldsymbol{Q}}$ is the atmosphere water vapor flux divergence, $\vec{\boldsymbol{Q}}$ is the vertically integrated horizontal vapor flux vector, E is the evapotranspiration and P is the precipitation. The solid- and liquid-phase water content in the atmosphere is generally small and was neglected.

Similarly, the water-balance equation for the land surface can be written as:

$$\frac{\mathrm{d}S}{\mathrm{d}t} + R = -(E - P) \tag{2}$$

where S is the water stored at the land surface, P is the precipitation, E is the evapotranspiration and R is runoff. S describes all the different water storages found on the land surface, like soil moisture, groundwater, canopy storage, and storage in rivers and lakes. Analogous to Eq. (1), runoff in Eq. (2) can be written as the divergence of the vertically integrated liquid water flux vector $\vec{\boldsymbol{r}}$:

$$R = \nabla_{\mathrm{H}} \cdot \vec{\boldsymbol{r}} \tag{3}$$

In order to apply Eqs. (1) and (2) to large spatial scales, as in the PILPS Phase 2(c) experiment, the equations have to be spatially integrated. Introducing the space average operator for the area A:

$$\langle (\cdot) \rangle = \frac{1}{A} \int\int (\cdot)\, \mathrm{d}A \tag{4}$$

the spatially averaged Eq. (2) becomes:

$$\left\langle \frac{\mathrm{d}S}{\mathrm{d}t} \right\rangle + \frac{1}{A} \oint_l (\vec{\boldsymbol{r}} \cdot \vec{\boldsymbol{n}})\, \mathrm{d}l = -\langle E - P \rangle \tag{5}$$

where $\vec{\boldsymbol{n}}$ is the outward normal vector at any point on the boundary l of the area A. $< R > = \{1\}/\{A\} \oint_l (\vec{\boldsymbol{r}} \cdot \vec{\boldsymbol{n}})\, \mathrm{d}l$ is the mean net water outflow per unit area from the area A, which in hydrological studies normally represents a river basin. It can further be subdivided into river outflow and groundwater or subsurface outflow from the basin. The subsurface water exchange between different basins,

especially when they are large, can normally be neglected (Peixoto, 1973). If groundwater outflow from the basin is neglected, the basin area averaged runoff $< R >$ times the basin area A equals the basin streamflow, as $\vec{r}$ is zero everywhere at the basin boundary, except at the basin outlet.

Spatial averaging of Eq. (1) leads to:

$$\left\langle \frac{\mathrm{d}W}{\mathrm{d}t} \right\rangle + \left\langle \nabla_{\mathrm{H}} \cdot \vec{Q} \right\rangle = \langle E - P \rangle \tag{6}$$

The difference of evapotranspiration minus precipitation $< E - P >$ is common to Eqs. (5) and (6) and, therefore, connects the terrestrial and the atmospheric water balance to:

$$\left\langle \frac{\mathrm{d}S}{\mathrm{d}t} \right\rangle + \langle R \rangle = - \left\langle \frac{\mathrm{d}W}{\mathrm{d}t} \right\rangle - \left\langle \nabla_{\mathrm{H}} \cdot \vec{Q} \right\rangle \tag{7}$$

The spatial averaging has direct consequences on the interpretation of storage change $\mathrm{d}S/\mathrm{d}t$ and runoff R (and their spatial means) in Eqs. (2), (5) and (7). All land-surface schemes participating in the PILPS Phase 2(c) are formulated as point models and follow, for their water balances, Eq. (2). This means that land surface schemes do not have any prior information about how fast locally produced runoff is transported from an area A. In order to compare modeled and observed runoff, the modeled runoff has to be routed out of the control volume.

3.2. Routing model

The routing model calculates the timing of the runoff reaching the outlet of a grid box, as well as the transport of water through the river network. It is assumed that water can leave a grid cell only in the direction of one of its eight neighboring grid cells. The runoff is then combined with the river discharge and routed downstream.

Both the within-grid cell and river routing contributions were represented using the linear and time-invariant models of Lohmann et al. (1996). The

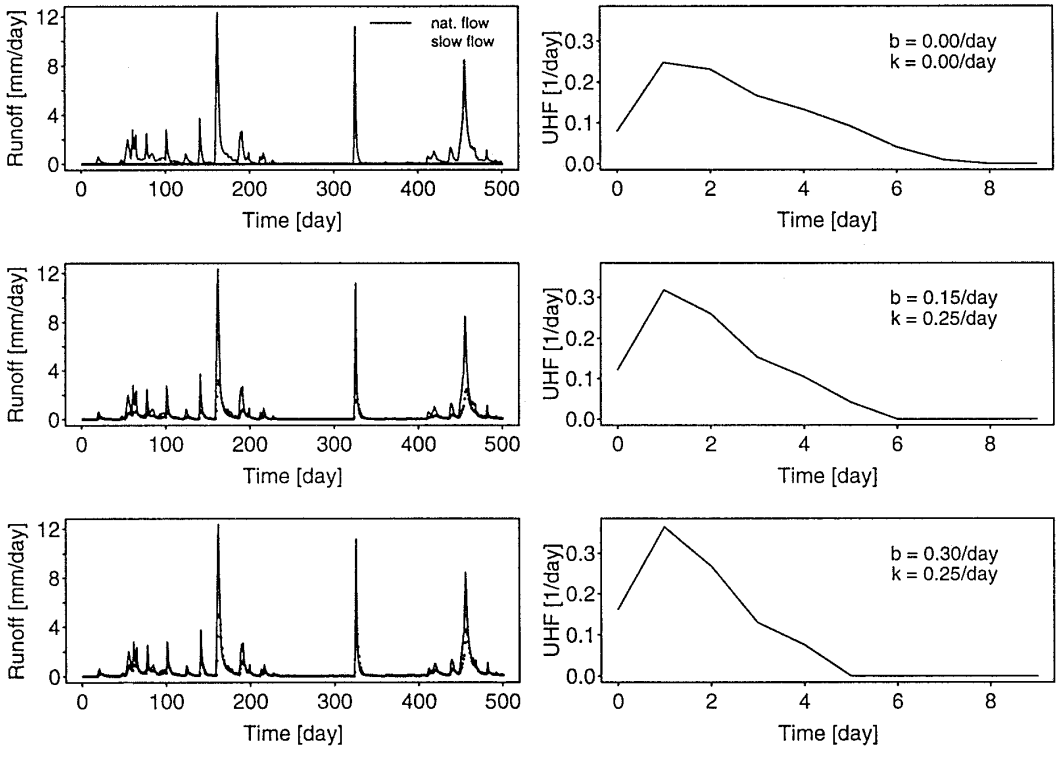

Fig. 3. Baseflow separation and hydrograph estimation for three different sets of parameters for b and k, for the Arkansas River above Lenapah. Higher fractions of separated slow components correspond to shorter impulse response functions for the fast component.

models were derived using naturalized streamflow and the same precipitation data that were provided to the participants. In addition to linearity and time invariance, the routing model assumes causality and nonnegativity of the impulse-response functions: two concepts that are well established in the hydrologic literature (see, for example, Dooge, 1979; Singh et al., 1982; Duband et al., 1993).

The first part of the routing model describes the time delay before runoff produced within a grid cell, and is measured as streamflow at the outlet of a 1°-by-1° grid cell. To estimate this delay, data from the USGS Lenapah gauging station (see Fig. 1) were used, because its catchment area (9420 km^2) is approximately the same size as that of a 1°-by-1° grid. The response time from this catchment was assumed to be representative of all grid cells in the Arkansas River basin. For the Red River, the same approach was used, based on data from the Red River above Millwood (10,700 km^2).

The routing model uses a simple baseflow separation technique to account for different timing responses of the surface and subsurface runoff processes within the land-surface schemes. The fast and slow components of the streamflow are assumed to be linearly related in the manner proposed by Rodriguez (1989) (see also Box et al., 1994; Duband et al., 1993):

$$\frac{dQ^S(t)}{dt} = b \cdot Q^F(t) - k \cdot Q^S(t) \qquad (8)$$

where, $Q^F(t)$ is the fast flow (caused by surface

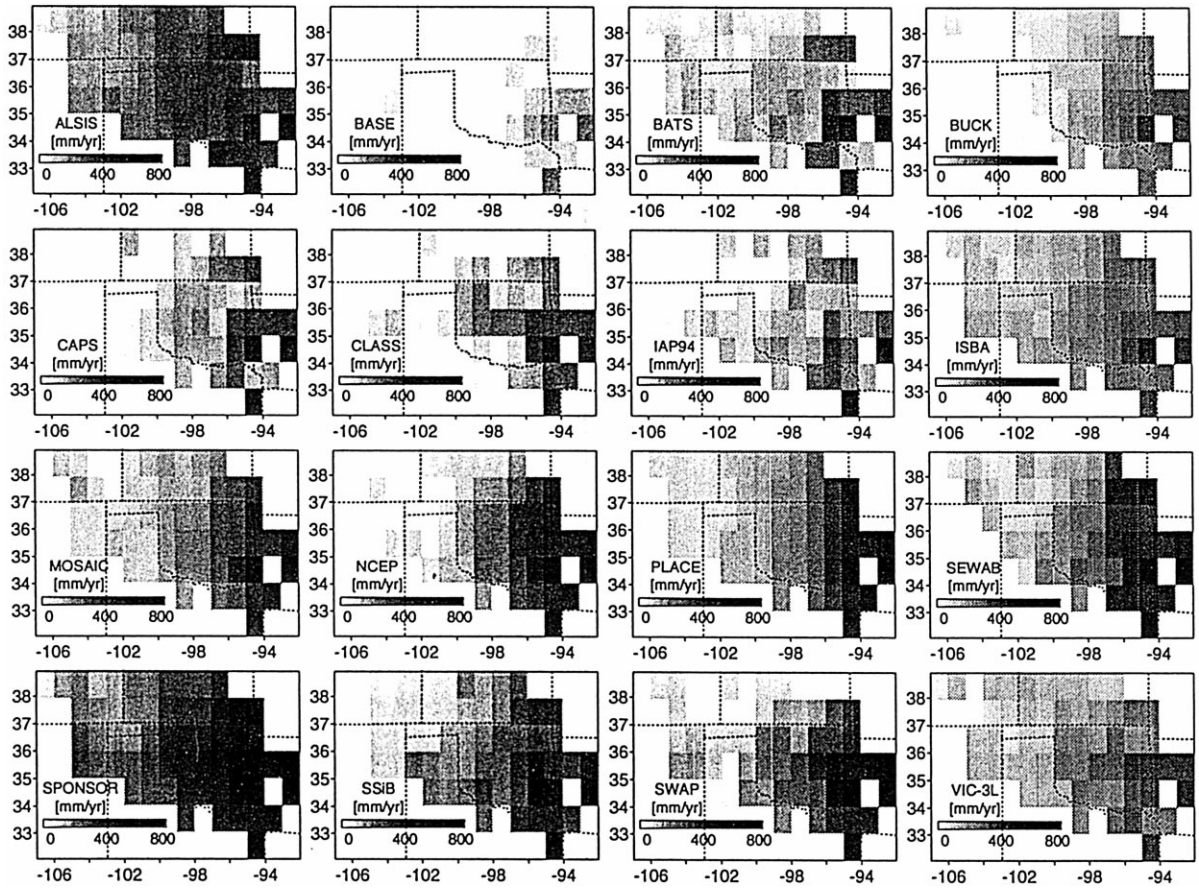

Fig. 4. Spatial runoff distribution of the annual mean (1980–1986) total runoff. All models capture the East–West gradient from the climatology, although some seem to overestimate runoff in dry areas.

runoff) and $Q^S(t)$ is the slow flow (caused by subsurface runoff or baseflow), and:

$$Q(t) = Q^S(t) + Q^F(t) \quad (9)$$

is the total streamflow. Fig. 3 illustrates this partitioning into the fast and slow components of the streamflow for three different sets of parameters for b and k, applied to the Arkansas River at Lenapah. The fast and slow components are related by:

$$Q^S(t) = b\int_0^t \exp(-k(t-\tau))\,Q^F(\tau)\,d\tau + Q^S(0) \times \exp(-kt) \quad (10)$$

Assuming that there is a linear relationship between the naturalized streamflow and that portion of precipitation that becomes streamflow (often referred to as effective precipitation P^{eff}), an impulse-response function can be used to relate the fast component Q^F to P^{eff}, and Eq. (10) can be used to relate fast and slow components of the streamflow. Computationally, the impulse-response function UH^F and P^{eff} is found by iteratively solving:

$$Q^F(t) = \int_{-\infty}^{t} UH^F(t-\tau)\,P^{eff}(\tau)\,d\tau \quad (11)$$

(see Lohmann et al., 1996; Duband et al., 1993).

The normalized impulse-response functions UH^F shown in Fig. 3 reflect the shape of the fast runoff hydrograph. Generally, higher fractions of baseflow result in a shorter impulse-response function for the fast flow, because the longer time scales are described by the slow baseflow component.

There is no general rule for estimating b and k, and while one might estimate different impulse-response functions, dependent on the inferred production of surface runoff and baseflow, for each scheme, we instead used the same impulse-response functions for all land surface schemes. The parameters used were $b = 0.15/\text{day}$ and $k = 0.25/\text{day}$, which were inferred from the observed precipitation and runoff series. The justification for using a common routing scheme was to avoid confounding differences in model dynamics with differences in routing schemes.

Once the water is transported out of the grid cell, it is further routed through the stream network (see Fig. 1). River routing is calculated with the linearized Saint-Venant equation (see, e.g., Lettenmaier and Wood, 1993):

$$\frac{\partial Q}{\partial t} = D\frac{\partial^2 Q}{\partial x^2} - C\frac{\partial Q}{\partial x} \quad (12)$$

where C and D are optimized for each grid cell within the basin using a least-quares method. Typical velocities, C, range between 0.8 m/s and 1.5 m/s and diffusivities, D, range between 600 m^2/s and 2000 m^2/s. Eq. (12) is solved with convolution

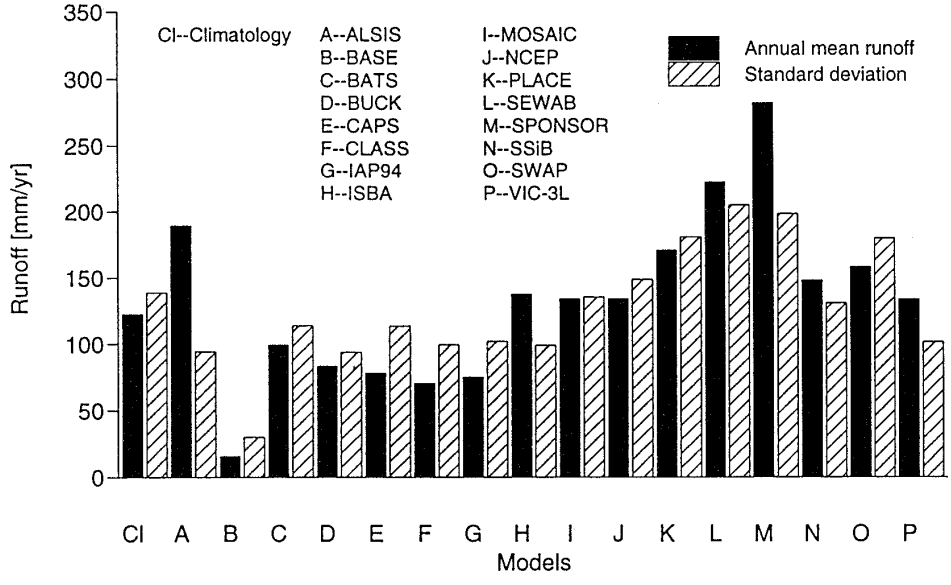

Fig. 5. Mean and spatial standard deviation (over all 61 grid cells) of model-predicted runoff for the years 1980–1986.

integrals of its impulse-response (or Green's) function (see Todini, 1991).

4. Basin-scale water-balance results

This section summarizes spatial and temporal analyses of the various terms in the surface water balance for the PILPS Phase 2(c) land-surface schemes. Some summary intercomparisons for the water balance are given by Wood et al. (this issue), while Liang et al. (this issue) present more detailed results for the evapotranspiration (latent heat) term.

4.1. Spatial runoff distribution

Fig. 4 shows the spatial pattern of annual average runoff (averaged for the period 1980–1986) for all schemes over the Red–Arkansas River basin area. The model results can be compared to the GGHYDRO runoff climatology shown in Fig. 2 for comparison.

The runoff climatology has a general East-to-West gradient, with a higher runoff in the eastern portion of the basin. This spatial pattern is consistent with the precipitation climatology. The results from the PILPS Phase 2(c) schemes are in qualitative agreement with climatological the gradient. However, the schemes also showed a wide range in the magnitude of the total basin runoff and in the strength of the East–West runoff gradient. For some of the schemes, the agreement of the simulated runoff with observations was markedly better in the humid areas than in the semihumid and semiarid regions.

Fig. 5 shows the mean annual runoff (averaged over the basin) for the years 1980–1986, and the

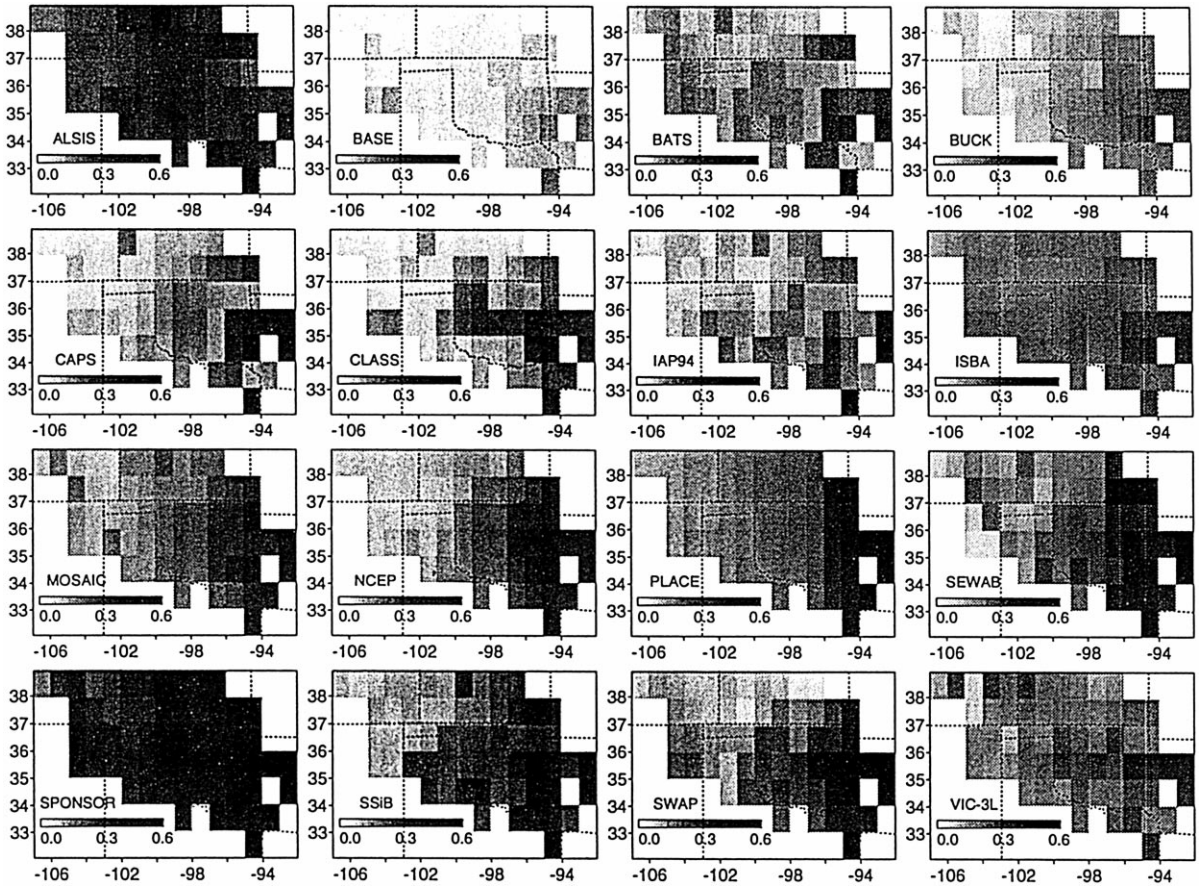

Fig. 6. Spatial distribution of the mean annual (1980–1986) runoff ratio.

spatial standard deviation of the mean annual runoff across the 61 grid boxes. For the observations, the spatial standard deviation was roughly equal to the mean annual runoff. However, models that tended to overpredict runoff in the western (dry) part of the catchment also tended to have a smaller standard deviation than models that predicted low runoff production. BASE produced significantly less runoff (runoff ratio of 0.02) than any other scheme while SPONSOR (runoff ratio of 0.41) produced runoff that was much larger than the climatology. Compared to the runoff climatology, ALSIS, ISBA, SEWAB, SPONSOR and VIC-3L produced too much runoff in the arid and semiarid western part of the catchment. PLACE, SEWAB, SPONSOR and SWAP produced the highest runoff of all models in the eastern part of the catchment.

Fig. 6 shows the spatial pattern of the mean annual runoff ratio for all schemes. Modeled runoff ratios varied between 0.0 in the western arid part of the catchment for some models to 0.6 (for SPONSOR) in the eastern part. From Fig. 6, it is apparent that the spatial patterns of the various schemes were very different, a fact that tended to be obscured by summary statistics, such as those shown in Fig. 5. Some schemes showed a very smooth West–East distribution, without large spatial gradients, and an almost monotonically increasing runoff ratio (BUCK, ISBA, NCEP, PLACE), while others had much more spatial texture (BATS, CLASS, IAP94, SEWAB, SPONSOR). ALSIS had the least spatial variation in its runoff ratio, which was almost constant for large areas. It is surprising that some models (BASE, CLASS, SPONSOR, SWAP) had a local maximum in the southwestern part of the basin, while others (IAP94, SEWAB, SSiB) had a local minimum there. The same held for single grid boxes in the southeast, where runoff ratios varied considerably among the schemes.

4.2. Streamflow comparison

Fig. 7 shows the partitioning of total runoff into surface and subsurface runoff for each scheme, aver-

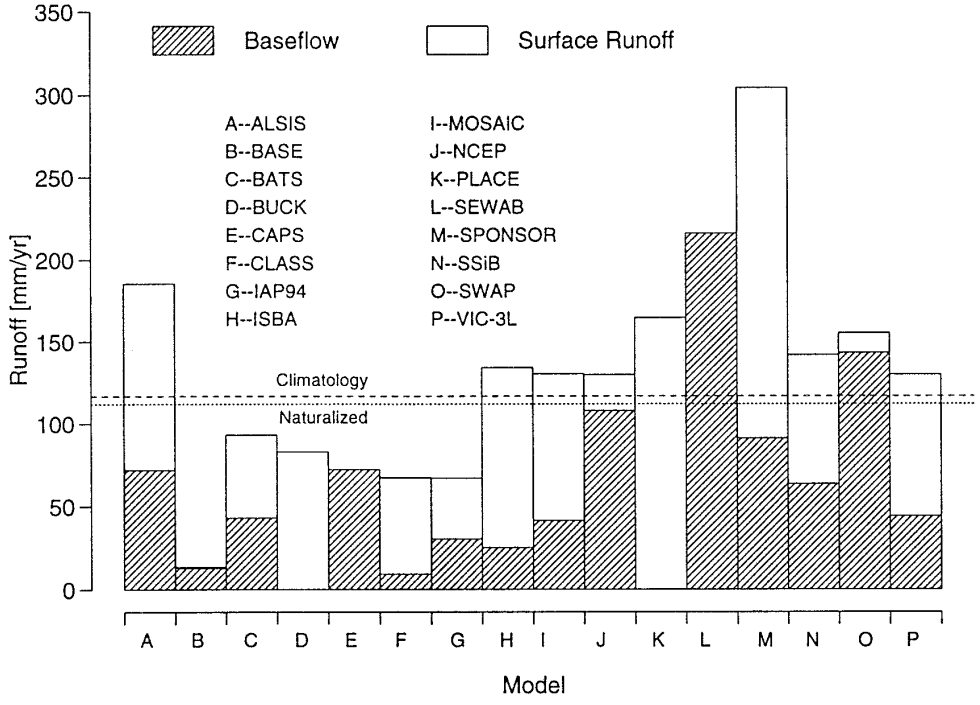

Fig. 7. Model partitioning of runoff into surface and subsurface runoff, averaged over the years 1980–1986. Also shown are the runoff from the GGHYDRO climatology and from the naturalized streamflow data.

aged over the years 1980–1986. The dotted line shows the naturalized streamflow and the dashed line the climatological value. The close agreement of these values suggests that the 1980–1986 period was climatologically representative. As noted in Section 2, the total runoff volumes from the models should deviate slightly from those shown in Fig. 5, because of the adjustment made for the basin boundary (see Section 2).

For some schemes, total runoff was dominated by subsurface drainage (BASE, CAPS, NCEP, SEWAB and SWAP), while others were dominated by surface runoff (BUCK and PLACE). The runoff of the BUCK model can also be interpreted as subsurface runoff. However, the runoff timing (as shown later) was more similar to that of surface runoff. The large surface runoff in PLACE is a result of the simple calibration scheme developed for the PILPS Phase 2(c), in which the surface runoff was highly correlated with incident precipitation. The remaining schemes appeared to have a more balanced division of surface and drainage runoff (ALSIS, BATS, CLASS, IAP94, ISBA, MOSAIC, SPONSOR, SSiB and VIC-3L). The importance of the different runoff components becomes clearer when the routed model output is compared to measured (naturalized) streamflow data. Some of the comparisons have to be done on a daily basis, since weekly or monthly averaged values do not reveal important differences in runoff response among models.

Figs. 8 and 9 show daily predicted streamflow from all schemes, compared to naturalized stream-

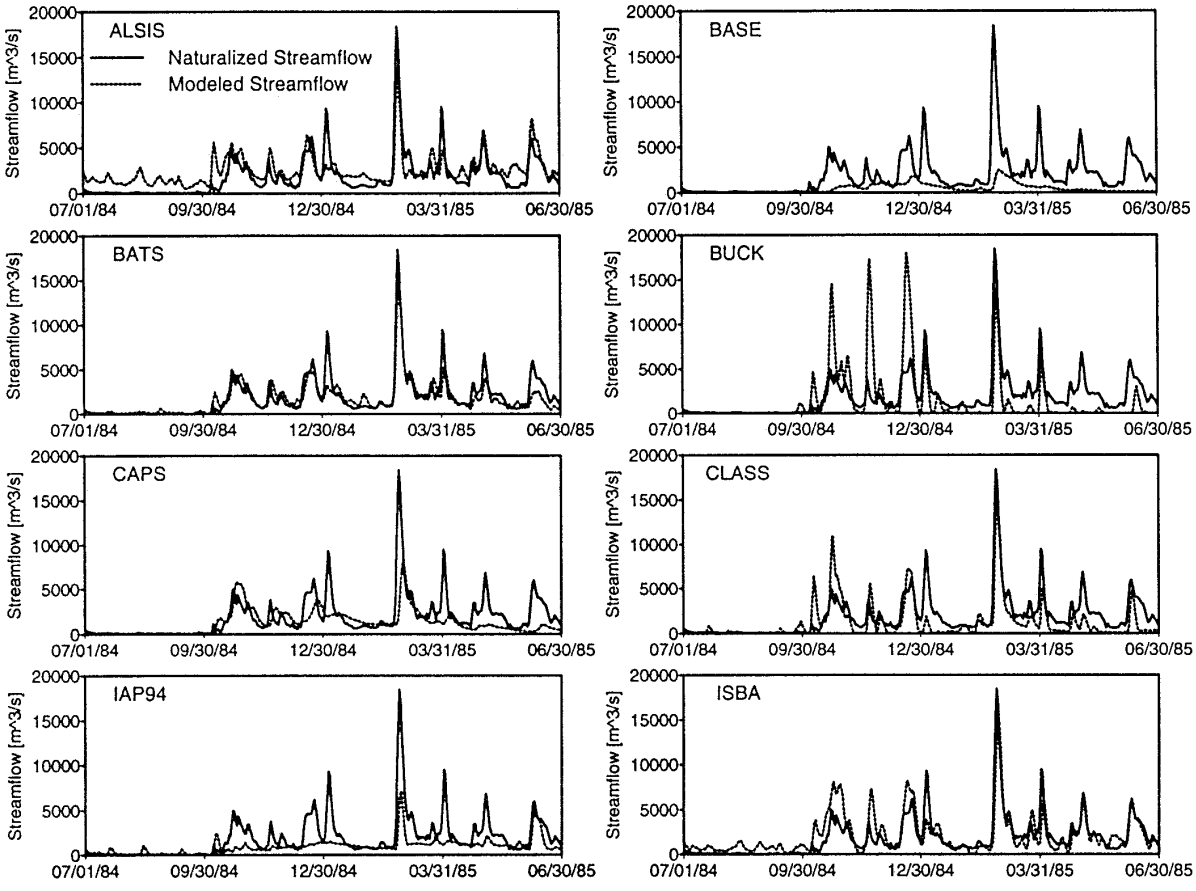

Fig. 8. Comparison of modeled and naturalized streamflow for ALSIS, BASE, BATS, BUCK, CAPS, CLASS, IAP94 and ISBA for a one-year period (July 1, 1984 to June 30, 1985).

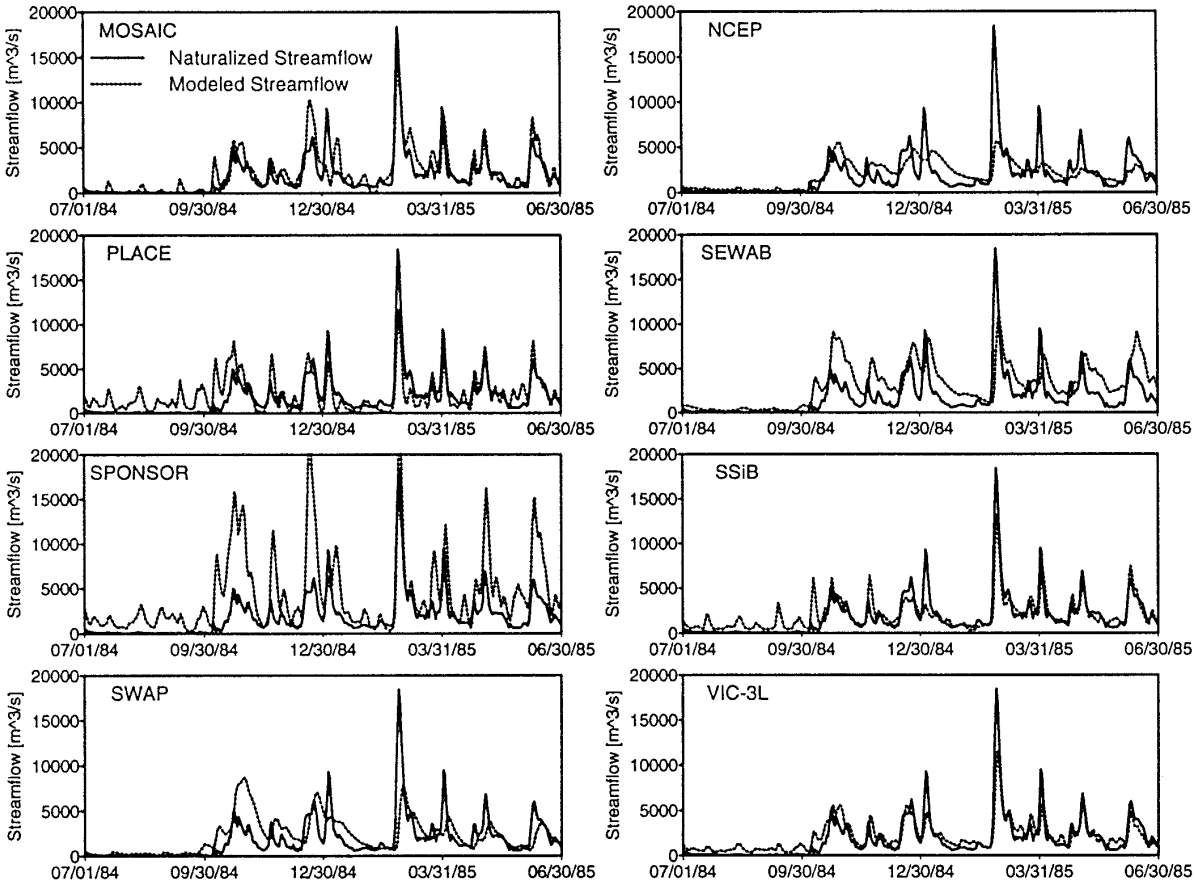

Fig. 9. Comparison of modeled and naturalized streamflow for MOSAIC, NCEP, PLACE, SEWAB, SPONSOR, SSiB, SWAP and VIC-3L for a one-year period (July 1, 1984 to June 30, 1985).

flow for the Arkansas River at Little Rock from July 1, 1984, to June 30, 1985. This period was chosen because it contains typical flow conditions during the year, with a low flow period from July to September, 1984, and medium to high flows during the rest of the year. Important criteria for evaluating daily streamflow are runoff timing, how often precipitation causes a runoff (or streamflow) event, and whether the land surface schemes are able to capture the summer low-flow period. These criteria show the general characteristics of runoff-production functions and their functional dependence on soil moisture and precipitation. The model results can be classified into two basic, distinct groups that share common characteristics. The subsurface-dominated results from BASE, CAPS, NCEP, SEWAB and SWAP are in one group; all other schemes are in another.

The subsurface-dominated schemes (BASE, CAPS, NCEP, SEWAB and SWAP) generally had their peak streamflow 2 to 7 days later than the naturalized streamflow (see Figs. 8 and 9). Also, their modeled streamflow generally had a slower recession than the naturalized streamflow. In addition, some schemes (e.g., NCEP) used different (deeper) soil depths than did most of the other schemes. The subsurface-dominated schemes did not produce runoff with each precipitation event, since precipitation was almost always able to infiltrate. The timing of runoff in all of these schemes is given by the solution of approximations to the Richards equation (Hillel, 1982; Richards, 1931), since only downward percolating water produces runoff. Figs. 8 and 9 show that models with a significant surface runoff production had more realistic hydrograph timing, as

evaluated using the naturalized streamflow. However, the amount of modeled streamflow varied widely. The interpretation of these results, in terms of runoff production processes (infiltration, drainage, surface and subsurface runoff), is more complicated than for the case of the subsurface-flow-dominated models.

Table 1 summarizes the runoff processes, as given by the participating Phase 2(c) schemes. All the schemes parameterize infiltration as a function of the precipitation or throughfall intensity and soil moisture. However, the models apply a wide range of concepts regarding the generation of either surface or subsurface runoff. Some of the schemes (BATS, ISBA, MOSAIC, SPONSOR, VIC-3L) include the concept of a contributing area (see Beven and Kirby, 1979), but its calculation varies widely among the schemes. MOSAIC and SPONSOR use the upper soil layer only in their surface runoff parameterization, while BATS and VIC-3L use the upper two layers and ISBA uses the entire soil column. Others use point- or local-infiltration equations for calculating the surface runoff (ALSIS, BASE, CAPS, CLASS, NCEP, PLACE, SEWAB, SSiB, SWAP), which are only applied to the top surface layer.

The low-flow period from July to September in Figs. 8 and 9 shows some of the characteristics of the surface-runoff processes in each model. BUCK, IAP94 and SEWAB only allow surface runoff when their upper layer (or the whole soil column in BUCK) is saturated. These approaches to parameterizing runoff can be seen in the daily streamflow pattern of BUCK and IAP94 as a threshold behavior, as different portions of the soil column wet–up or dry–down. Some models (ALSIS, PLACE and SPONSOR) produced significantly more runoff than the other schemes during this time period, mainly in the eastern part of the catchment. ALSIS predicted too much subsurface and surface runoff in the summer, while PLACE responded to all precipitation events with surface runoff, due to the added runoff scheme. ISBA, MOSAIC, SEWAB, SSiB and VIC-3L also produced too much runoff. BASE, BATS, BUCK, CAPS, CLASS, IAP94, NCEP and SWAP produced hardly any runoff in summer. It also seems that most subsurface-dominated models produced reasonable

Table 1
Runoff production mechanisms of the models

Model	Surface runoff	Subsurface runoff
ALSIS	excess throughfall above infiltration capacity	free drainage
BASE	excess of point-based infiltration capacity for top soil layer reduced by simple soil wetness factor	free drainage
BATS	power law similar to VIC curve, dependent on soil moisture for upper two soil layers	free drainage
BUCK	bucket overflow	bucket drainage
CAPS	excess above maximal infiltration rate	free drainage
CLASS	overflow of surface retention capacity	free drainage
IAP94	no surface runoff before upper soil layer is saturated	
ISBA	VIC curve for total soil moisture	free drainage
MOSAIC	throughfall on saturated fraction becomes runoff	free drainage, slope dependent
NCEP	variable fraction of throughfall from canopy, depending on soil moisture	free drainage
PLACE	residual excess water above interception, storage capacity and maximum infiltration	topography-driven lateral discharge and free drainage
SEWAB	no surface runoff before upper soil layer is saturated	free drainage
SPONSOR	saturated contributing areas, Horton mechanism, dependent on precipitation intensity and upper layer soil moisture	free drainage
SSiB	saturation excess, point infiltration equation	free drainage
SWAP	residual excess water above interception and infiltration, determined by Green–Ampt model	free drainage, slope dependent
VIC-3L	variable infiltration capacity (VIC) for upper two soil layers	nonlinear Arno baseflow curve

For the BUCK model, two different conceptualizations are in the table.

results during the low-flow period in summer from July to September (see Figs. 8 and 9), although these models did not show the small, observed responses of streamflow to precipitation. The results of BATS for the summer low-flow period were surprisingly good. It has a similar surface-runoff-production function as implemented in ISBA or VIC-3L, but it produced significantly less runoff.

It is rather surprising that the free-drainage lower-boundary condition of most land-surface schemes for subsurface runoff produced reasonable results. The internal delay inferred from a Richards-equation analysis has about the right timing to sustain the peak after high-flow events. However, ALSIS and SPONSOR showed significant amounts of subsurface runoff during the summer dry period and after high precipitation events in fall, resulting in overpredictions of streamflow. The high subsurface runoff produced by ALSIS can be explained by the Broadbridge and White model (Shao and Irannejad, submitted) and, depending on soil type, may give unsaturated hydraulic conductivities hundreds of times higher than those of Clapp and Hornberger.

We suspect that in groundwater-dominated basins in more humid areas, the lower boundary condition of most models would have failed to produce the right runoff timing. Some models include an explicit slope-dependent subsurface runoff parameterization (MOSAIC, PLACE, SWAP), which is sometimes referred to as interflow, to indicate that the scheme considers explicitly lateral discharge. The VIC-3L model also belongs to this group, as its baseflow curve is an empirical description of subsurface-flow processes (Franchini and Pacciani, 1991). Interflow

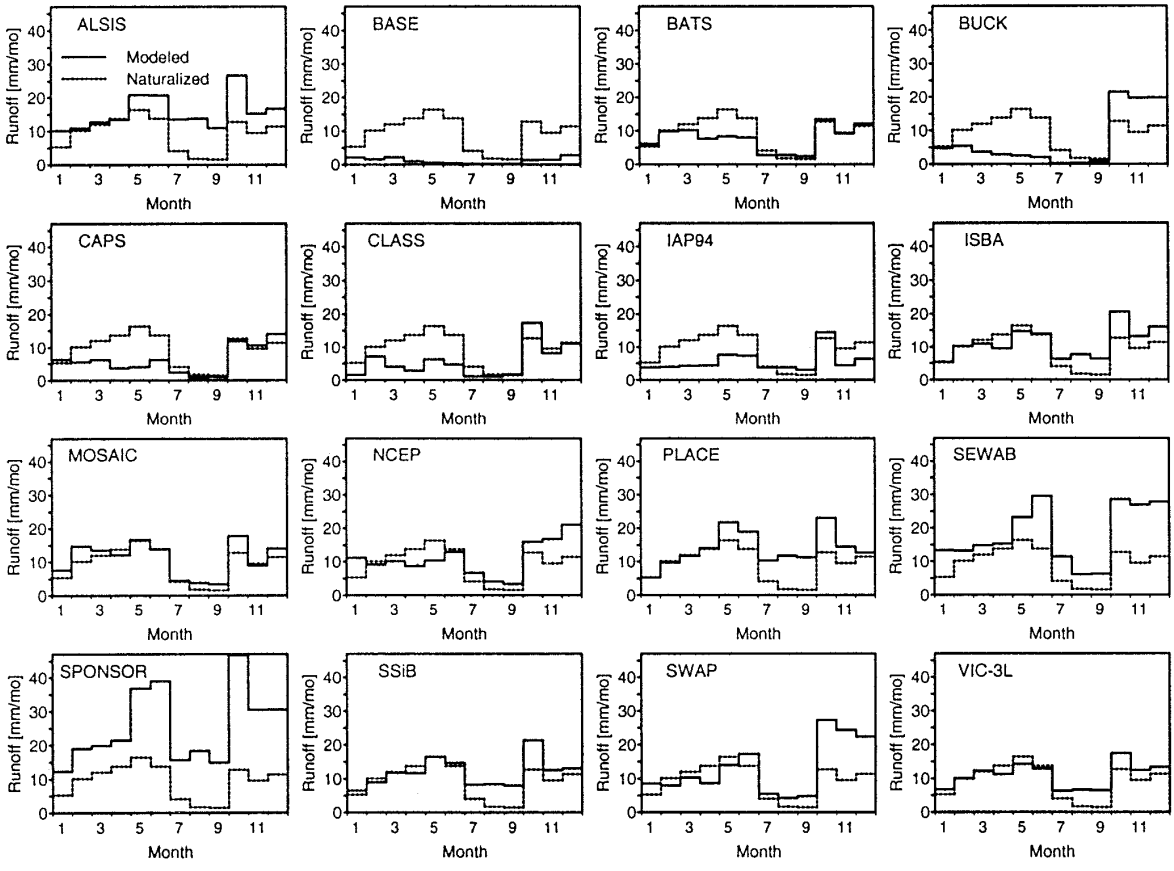

Fig. 10. Sum of Arkansas and Red River monthly mean (1980–1986) runoff routed to respective basin outlets, compared to observed monthly mean naturalized streamflows summed at basin outlets.

was added to subsurface flow in PILPS Phase 2(c) and, therefore, cannot be analyzed separately.

Fig. 10 shows the aggregated average monthly runoff for the years 1980–1986 for the entire catchment. The 16 schemes produced quite different mean seasonal runoff patterns but, in general, had the minimum runoff occurring in the summer. The majority of the schemes had peaks in the spring and fall, but some had anomalous late fall–winter high runoff (BUCK, SEWAB and SPONSOR). Quite a few schemes produced too much summer runoff (ALSIS, ISBA, PLACE, SEWAB, SPONSOR, SSiB and VIC-3L), suggesting that the infiltration-runoff processes under dry conditions can be improved. These schemes also tended to overpredict runoff for the arid catchments listed in Table 2.

Table 2 summarizes the statistics for the years 1980–1986 for the Arkansas River (Ark), the Red River (Red), and the subcatchments at Ralston (Ral), Eufaula (Euf), Ft. Gibson (Gib), and Texoma (Tex) (see Fig. 1). Ralston and Eufaula represent the arid part of the basin, while Ft. Gibson represents the humid part. Texoma represents an arid location in the Red River. In all cases, performance was evaluated by errors in model predictions of mean annual runoff volume and the Pearson correlation coefficient R^2 between daily modeled and naturalized streamflow. Only five schemes were within 25% of the total volume of the naturalized streamflow (BATS, ISBA, MOSAIC, NCEP and VIC-3L). Seven were within 50% (BUCK, CAPS, CLASS, IAP94, PLACE, SSiB, SWAP). The four schemes with subgrid runoff production conceptualizations (BATS, ISBA, MOSAIC, and VIC-3L) had higher R^2 values for the large (Red–Arkansas) river basin than did the other models. For the smaller catchments, the results were not as conclusive.

4.3. Evapotranspiration comparison

Fig. 11 shows the monthly mean (1980–1986) estimated atmospheric budget evapotranspiration plotted against the model-based evapotranspiration from the 16 schemes. The dashed line is the 1:1 line, and the solid line results from a linear regression of modeled and observed streamflow. The slope of the regression line is generally less than one, indicating that in general all the models tend to overpredict low (winter) evapotranspiration and underpredict high (e.g., summertime) evapotranspiration. It should be noted that the atmospheric budget evapotranspiration

Table 2
Runoff statistics for the Arkansas and Red rivers for the years 1980–1986

	Ark (mm)	R^2	Red (mm)	R^2	Ral (mm)	R^2	Euf (mm)	R^2	Gib (mm)	R^2	Tex (mm)	R^2	Mean (mm)
Observed	98.4		147.5		37.4		43.8		285.5		63.8		112.0
ALSIS	162.8	0.53	244.0	0.48	117.5	0.15	133.1	0.17	227.6	0.46	190.2	0.41	185.3
BASE	10.4	0.26	21.7	0.14	0.93	0.00	5.1	0.04	14.2	0.09	3.4	0.20	13.5
BATS	76.4	0.72	136.7	0.68	29.8	0.16	46.2	0.17	135.8	0.60	71.5	0.39	93.1
BUCK	69.4	0.43	119.7	0.32	18.6	0.14	36.6	0.14	150.8	0.35	55.1	0.24	83.4
CAPS	56.5	0.39	113.5	0.33	12.1	0.05	20.0	0.10	103.9	0.24	38.9	0.12	72.3
CLASS	53.4	0.50	103.1	0.39	9.0	0.25	39.6	0.20	82.4	0.40	20.7	0.13	67.2
IAP94	47.1	0.40	119.0	0.30	12.9	0.03	20.8	0.09	113.8	0.37	53.1	0.29	67.0
ISBA	115.4	0.66	184.1	0.68	60.0	0.21	83.5	0.19	195.2	0.60	111.5	0.39	134.4
MOSAIC	106.5	0.62	192.5	0.65	41.2	0.13	60.6	0.24	242.4	0.53	75.9	0.40	130.3
NCEP	100.5	0.52	205.8	0.47	24.6	0.18	51.2	0.21	228.7	0.40	77.6	0.29	129.7
PLACE	136.6	0.49	237.6	0.48	46.0	0.22	59.0	0.25	354.4	0.44	87.3	0.40	164.6
SEWAB	178.4	0.58	313.7	0.55	69.3	0.12	88.0	0.19	424.4	0.39	153.2	0.22	215.9
SPONSOR	269.0	0.59	397.2	0.72	150.7	0.23	205.6	0.16	464.2	0.54	253.3	0.41	304.5
SSiB	113.4	0.58	215.7	0.54	43.8	0.11	72.8	0.27	232.8	0.51	126.5	0.39	141.8
SWAP	118.0	0.43	251.9	0.43	22.0	0.09	77.3	0.14	223.0	0.27	114.7	0.20	155.1
VIC-3L	114.4	0.73	171.1	0.71	57.3	0.25	78.7	0.32	155.3	0.61	109.4	0.42	130.1

R^2: Pearson correlation coefficient; Ral: Ralston subcatchment; Euf: Eufaula subcatchment; Gib: Ft. Gibson subcatchment; Tex: Texoma subcatchment.

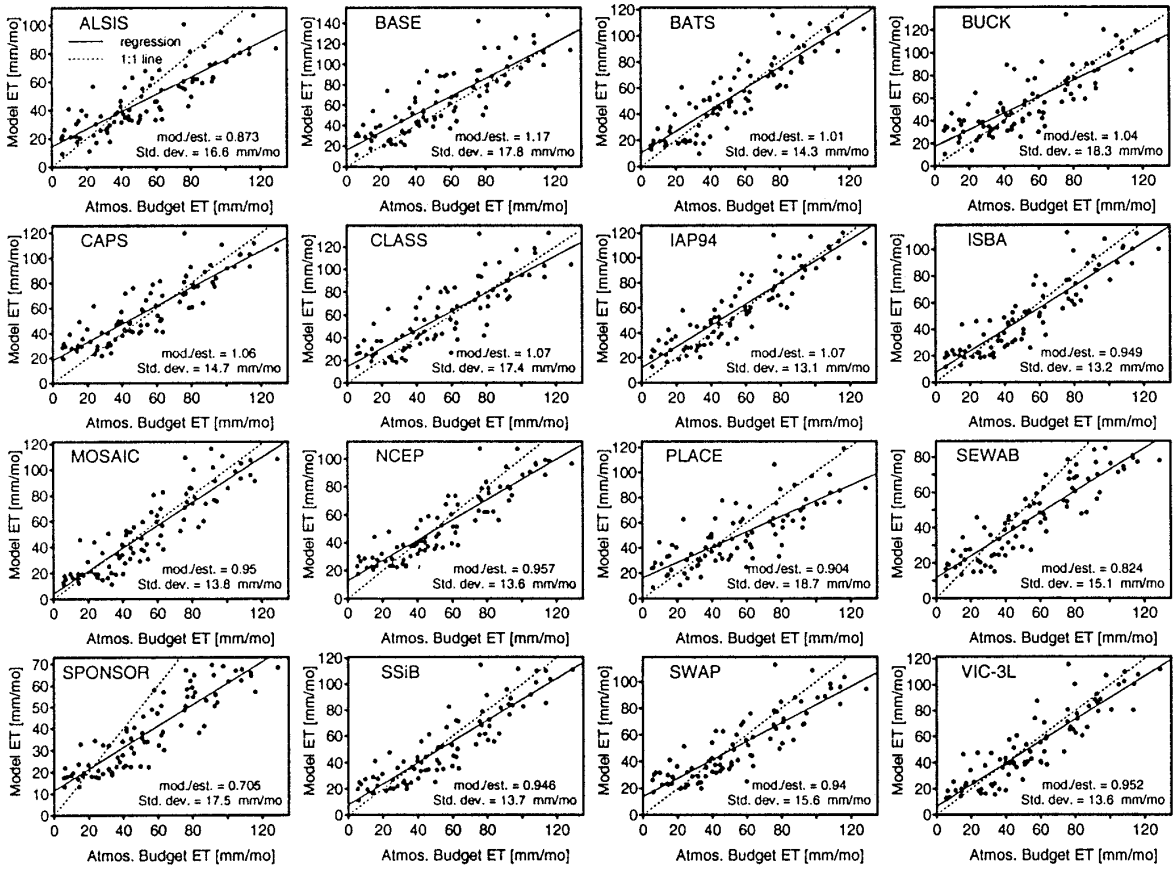

Fig. 11. Scatterplot of mean monthly (1980–1986) basin evapotranspiration estimated from atmospheric budget vs. modeled evapotranspiration. The solid line is a regression of the modeled evapotranspiration, the dashed line is the 1:1 line.

estimates are more error-prone (in terms of fractional errors) in winter than in summer, so less significance should be attached to the apparent winter overpredictions than to the summer underpredictions. Owing to longterm closure of the surface water balance, schemes whose evapotranspiration compared poorly to the atmospheric budget were the same as those with poor runoff performance.

Fig. 12 shows the mean monthly evapotranspiration for the 16 schemes for the period 1980–1986. In general, all the schemes followed the same evapotranspiration seasonal cycle as that derived from the atmospheric budget. SPONSOR significantly underestimated evaporation, but the seasonal cycle was approximately correct. ALSIS and PLACE underestimated evapotranspiration, but they (and SWAP) appear to have a pronounced July underestimation. Many schemes underestimated July and August evapotranspiration and overestimated evapotranspiration in October. As shown in Section 4.4, evapotranspiration during summer exceeds precipitation on average, requiring extraction of moisture from the soil column. On the other hand, October is the month with the greatest average soil-moisture recharge. Therefore, the results suggest that most of the schemes could be improved by refining the parameterizations of soil–evapotranspiration interactions.

4.4. Basin water-balance inference

Fig. 13 compares the predicted water balance of the Red–Arkansas River basin among schemes. The

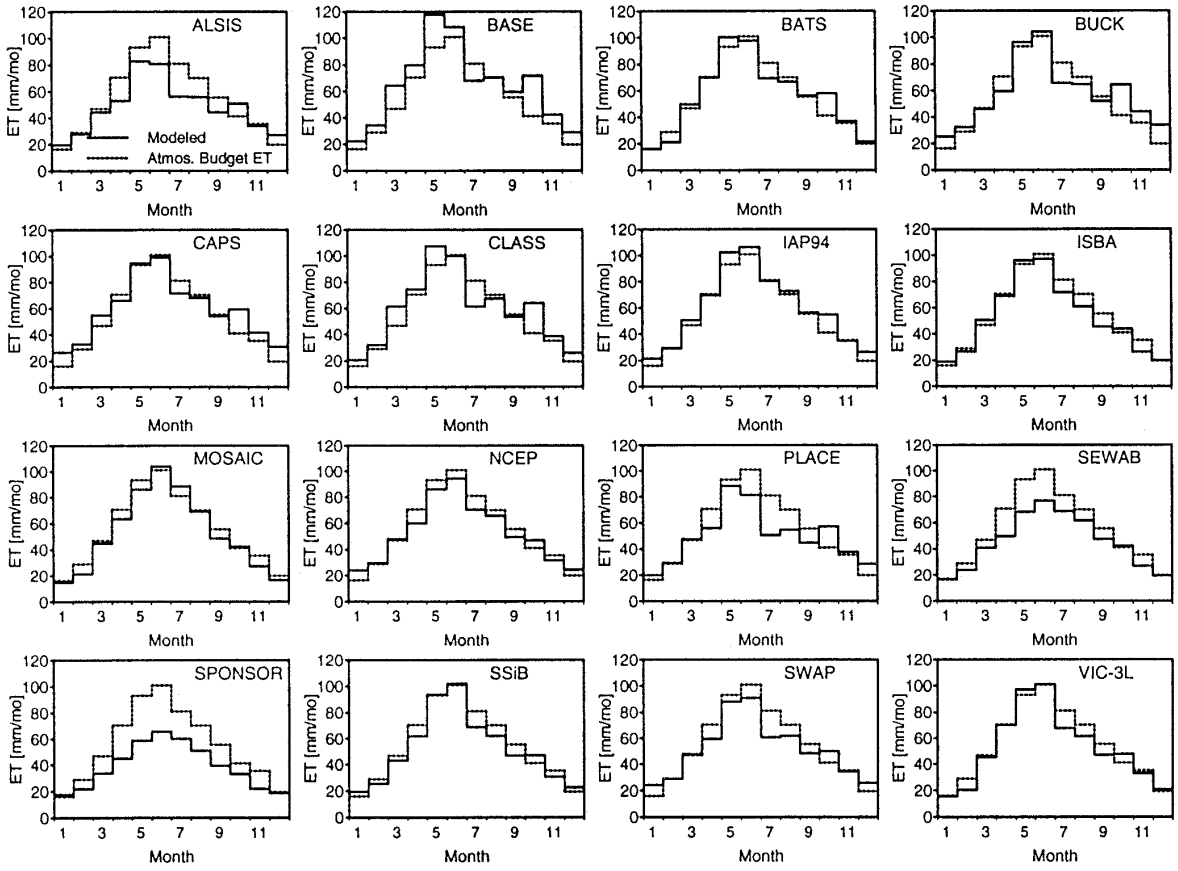

Fig. 12. Mean monthly basin evapotranspiration for all the models for the years 1980–1986.

estimated soil-moisture storage change (shown as the dashed line) was calculated as the residual from Eq. (7), using observed precipitation, the atmospheric budget calculation, and naturalized streamflow. To represent a storage change which is closer to that computed by the models, naturalized streamflow was deconvoluted to represent an instantaneous value using Eq. (11). The resulting monthly P^{eff} (not shown) had a small phase shift from the naturalized streamflow, with a maximum difference of 4 mm/month.

Deviations in the monthly water balance among the schemes are shown by the difference of the triangles from the solid, black curve. They result from canopy-storage effects (which are almost negligible) and snow processes (winter months only), which are not included in the estimated storage term.

From Fig. 13, it can be seen that the different schemes predicted a wide range in the soil-moisture storage-change term and in its seasonal cycle. Most schemes captured the seasonal cycle in the storage-change term, with little change during late winter–early spring, depletion during the summer, and recharge during the fall. All schemes captured the small storage changes in January, February, March and September, and negative storage change in the summer (June, July and August). However, the negative storage change suggested in the data for April only shows up in some schemes. All models, except BUCK, had their largest positive storage change in October, but, in general, most schemes underpredicted its magnitude. In the summer, and particularly in July, the evapotranspiration is strongly determined by the availability of soil moisture. In July, MO-

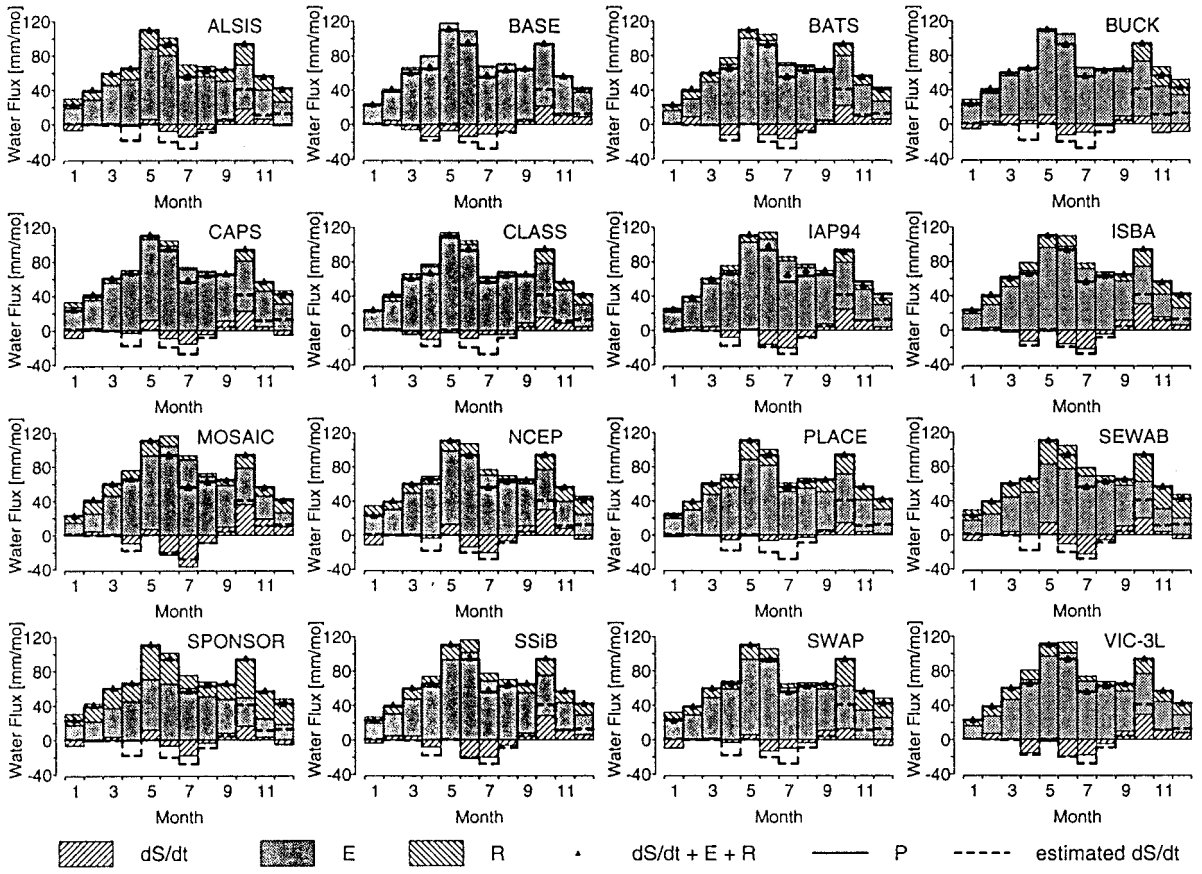

Fig. 13. Mean monthly (1980–1986) water balances. The dS/dt of the land surface is estimated from the effective precipitation and the atmospheric budget. The triangles indicate whether the models balance their water budget on a monthly time scale, without the inclusion of the snow and canopy storage.

SAIC had a storage change of about 38 mm on average, while CLASS and PLACE only had about 5 mm. The opposite occurred in October, when the soil water is replenished.

Most land-surface schemes (with the exception of BUCK and NCEP) used the soil data provided by the PILPS Phase 2(c) organizers. The total soil depth and the rooting depth are arguably among the most important parameters in land surface modeling, as they determine how much water can be stored and made available for evapotranspiration during the dry season (Milly and Dunne, 1994). Therefore, it is rather interesting that the schemes predicted a wide range in moisture storage, as shown in Fig. 13. There is, of course, a question of how representative the parameters used in the PILPS Phase 2(c) are of the large-scale features of the basin. And, it must be noted, that the root-zone depth, although given to the modelers, varied among the models, mainly because of the different model structures. Also, it should be emphasized that a budget analysis, such as that reported here, cannot infer an absolute soil moisture; It can only estimate storage changes. Therefore, even models with similar changes in soil water storage can show quite different soil-moisture values. This is consistent with the results found in PILPS Phases 1 and 2(a) experiments.

It is our belief that the variations in moisture-storage terms among the schemes arise from the interconnected water and energy cycles, so that weaknesses in the parameterization of one process (for example runoff) affect other fluxes and storage terms

(see also Koster and Milly, 1997). What is unresolved among the parameterization of land surface schemes is the sensitivity between processes. Further analyses of Phase 2(c) experiment results are expected to shed more light on such questions.

5. Conclusions

The water balances of the 16 PILPS Phase 2(c) land surface schemes were evaluated by comparisons of predicted and observed streamflow, predicted evapotranspiration and evapotranspiration inferred from an atmospheric moisture budget analysis, and soil moisture storage changes inferred from a surface-water budget analysis. The comparisons were based on a seven-year period (1980–1986). This period was assumed to be representative of the longterm climatology, and therefore the water balance of the validation data was assumed to be closed. This was confirmed by comparison of the seven-year mean annual naturalized streamflow and the water–vapor convergence for the same period over an area which represents the basin; the two matched closely. With the help of these data, inferences can be made about the water budget. Although driven with the same forcing data, the models showed significant differences in their water-balance components.

The evaluations of model-generated streamflow, evaporation, and soil-moisture changes support the following conclusions.

(1) The mean annual runoff of all models followed, at least generally, the strong climatic East–West gradient of precipitation. The spatial variability of runoff, expressed as the spatial standard deviation of the annual means, had about the same magnitude as the mean runoff itself, and this feature was roughly captured by all of the models. The most common variation from this characteristic of the observations was the prediction, by most of the models, of too much runoff in the dry part of the basin and, hence, underprediction of the spatial variability in the runoff fraction. Also, there were considerable differences in the spatial pattern of annual mean runoff (and, therefore, also runoff ratio and evapotranspiration). Some models had fairly smooth spatial variations, but a number of models had much stronger spatial variations and, hence, less coherent spatial patterns.

(2) The conceptualizations used by the different models for generation of surface and subsurface runoff vary greatly, and these differences strongly affected the character of the runoff hydrographs generated by the difference models. Only two models showed a bucket-type runoff-generation process, which implies a threshold behavior in surface runoff. Nonetheless, among the remaining 14 models, the response ranged from models which responded to all summer precipitation events with surface runoff, to one that produced almost no runoff in summer. Those models whose runoff production is subsurface dominated generally produced the most reasonable results during the summer low-flow period.

(3) Comparison of mean seasonal cycles of changes in soil-moisture storage with storage changes inferred from observations indicates that the seasonal cycles were qualitatively similar for most models. For instance, all schemes captured the small storage changes in January, February, March and September and negative storage change in the summer (June, July and August). However, the negative storage change suggested in the data for April were missed by most schemes, and most schemes underpredicted the large, positive storage change in October that was present in the observations. The differences of the mean basin average storage change among the models were quite large, and accounted for up to 25% of the modeled evapotranspiration in July.

(4) The use of soil water in the summer and the refilling of the storage in autumn strongly determined evapotranspiration in April, June, July and October. In general, the models followed the seasonal cycle of the atmospheric budget-derived evapotranspiration quite well, although most underestimated the high summertime (July and August) evapotranspiration, and overestimated the low evapotranspiration in October. The results suggest that most of the schemes could be improved by refining the parameterizations of soil–evapotranspiration interactions. It also raises the question of uncertainties in the model parameters (e.g., root and soil depth).

Acknowledgements

The results presented in this paper are based on the PILPS Phase 2(c) workshop, which was held

from October 28–31, 1996, at Princeton University. The PILPS Phase 2(c) activities at Princeton University were supported by NSF Grant EAR-9318896 and by the NOAA (Office of Global Programs) Grant NA56GP0249. The PILPS Phase 2(c) activities at University of Washington were supported by NSF Grant EAR-9318898 and by NOAA/OGP Grant NA67RJ0155.

References

Abdulla, F., 1995. Regionalization of a macroscale hydrological model. PhD thesis, Department of Civil Engineering, University of Washington, USA.

Beven, K., Kirby, M., 1979. A physically based, variable contributing area model of basin hydrology. Hydrol. Sci. Bull. 24, 43–69.

Box, G., Jenkins, G., Reinsel, G., 1994. Time Series Analysis, Forecasting and Control, 3rd edn. Prentice-Hall, Englewood Cliffs, NJ.

Cogley, J., 1991. GGHYDRO-global hydrographic data release 2.0. Trent Climate Note 91-1. Trent University.

Dooge, J., 1991. Deterministic input–output models. In: Lloyd, E., O'Donell, T., Wilkinson, J. (Eds.), The Mathematics of Hydrology and Water Resources. Academic Press, pp. 1–37.

Duband, D., Obled, C., Rodriguez, J., 1993. Unit hydrograph revisited: an alternate iterative approach to UH and effective precipitation identification. J. Hydrol. 150, 115–149.

Franchini, M., Pacciani, M., 1991. Comparative analysis of several conceptual rainfall-runoff models. J. Hydrol. 122, 161–219.

Henderson-Sellers, A., Yang, Z.-L., Dickinson, R., 1993. The project for intercomparison of land-surface parameterization schemes. Bull. Am. Meteor. Soc. 74, 1335–1349.

Henderson-Sellers, A., Pitman, A., Love, P., Irrannejad, P., Chen, T., 1995. The project for intercomparison of land-surface parameterization schemes (PILPS): Phases 2 and 3. Bull. Am. Meteor. Soc. 94, 489–503.

Hillel, D., 1982. Introduction to Soil Physics. Academic Press.

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

Lettenmaier, D., Wood, E., 1993. In: Handbook of Hydrology: Hydrologic Forecasting, Chap. 26. McGraw-Hill.

Liang, X., Wood, E., Lettenmaier, D., Lohmann, D., Boone, A., Chang, S., Chen, F., Dai, Y., Desborough, C., Dickinson, R., Duan, Q., Ek, M., Gusev, Y., Habets, F., Irannejad, P., Koster, R., Mitchell, K., Nasonova, O., Noilhan, J., Schaake, J., Schlosser, A., Shao, Y., Shamkin, A., Verseghy, D., Wang, J., Warrach, K., Wetzel, P., Xue, Y., Yang, Z., Zeng, Q., this issue. The Project for Intercomparison of Land-surface Parameterization Schemes (PILPS) Phase 2(c): Red–Arkansas River basin experiment: 2. Spatial and temporal analysis of energy fluxes: global and planetary change.

Linsley, R., Kohler, M., Pauhlus, J., 1975. Hydrology for Engineers, 2nd edn. McGraw-Hill, New York.

Lohmann, D., Nolte-Holube, R., Raschke, E., 1996. A large scale horizontal routing model to be coupled to land surface parameterization schemes. Tellus 48 A, 708–721.

Milly, P., Dunne, K., 1994. Sensitivity of the global water cycle to the water-holding capacity of land. J. Clim. 7, 506–526.

Peixoto, J., 1973. Atmospheric vapor flux computations for hydrological purposes. Reports on WMO/IHD Projects 20, World Meteorological Organization.

Richards, L., 1931. Capillary conduction of liquids through porous mediums. Physics 1, 318–333.

Rodriguez, J., 1989. Modélisation pluie-débit par la méthode DPFT. Technical Report Thése de doctorat, Grenoble, France.

Shao, Y., Irannejad, P., submitted. On the choice of soil hydraulic models in land surface schemes. Boundary-Layer Meteorol.

Singh, V., Baniukiewiez, A., Ram, R., 1982. Some empirical models of determining the unit hydrograph. In: Singh, V.P. (Ed.), Rainfall–Runoff Relationship. Water Resources Publications, pp. 67–90.

Starr, V., Peixoto, J., 1958. On the global balance of water vapor and the hydrology of deserts. Tellus 10, 189–194.

Todini, E., 1991. Hydraulic and hydrologic food routing schemes. In: Bowles, D.S., O'Connell, P.E. (Eds.), Recent Advances in the Modeling of Hydrologic Systems. NATO ASI Series C, Vol. 345, pp. 389–406.

Todini, E., 1996. AFORISM: a comprehensive forecasting system for flood risk mitigation and control, technical report. Final report to the Commission of European Communities.

Wetzel, S., 1994. A hydrological model for predicting the effects of climate change. Technical report, BS thesis, Princeton University, Department of Civil Engineering and Operations Research.

Wood, E., Lettenmaier, D., Liang, X., Lohmann, D., Boone, A., Chang, S., Chen, F., Dai, Y., Dickinson, R., Duan, Q., Ek, M., Gusev, Y., Habets, F., Irranejad, P., Koster, R., Mitchell, K., Nasonova, O., Noilhan, J., Schaake, J., Schlosser, A., Shao, Y., Shmakin, A., Verseghy, D., Wang, J., Warrach, K., Wetzel, P., Xue, Y., Yang, Z., Zeng, Q., this issue. The Project for Intercomparison of Land-Surface Parameterization Schemes (PILPS) Phase 2(c) Red–Arkansas River basin experiment: 1. Experiment description and summary intercomparisons: global and planetary change.

Zhao, W., 1997. Diagnostic studies and numerical modeling of heavy rainfall in the central plains. PhD thesis, Princeton University, Department of Civil Engineering and Operations Research.