ELSEVIER

Global and Planetary Change 19 (1998) 115–135

GLOBAL AND PLANETARY CHANGE

The Project for Intercomparison of Land-surface Parameterization Schemes (PILPS) Phase 2(c) Red–Arkansas River basin experiment: 1. Experiment description and summary intercomparisons

Eric F. Wood [a,*], Dennis P. Lettenmaier [b], Xu Liang [a,1], Dag Lohmann [a], Aaron Boone [c,k], Sam Chang [d], Fei Chen [e,2], Yongjiu Dai [f], Robert E. Dickinson [g], Qingyun Duan [h], Michael Ek [i], Yeugeniy M. Gusev [j], Florence Habets [k], Parviz Irannejad [l], Randy Koster [m], Kenneth E. Mitchel [e], Olga N. Nasonova [j], Joel Noilhan [k], John Schaake [h], Adam Schlosser [n], Yaping Shao [l], Andrey B. Shmakin [o], Diana Verseghy [p], Kirsten Warrach [q], Peter Wetzel [c], Yongkang Xue [r,3], Zong-Liang Yang [g], 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] *Institute of Atmospheric Physics, University of Arizona, Tucson, AZ, USA*
[h] *Office of Hydrology, NOAA / NWS, Silver Spring, MD, USA*
[i] *Oregon State University, Corvallis, OR, USA*
[j] *Institute of Water Problems, Moscow, Russian Federation*
[k] *Meteo-France / CNRM, Toulouse, France*
[l] *Centre for Advanced Numerical Computation in Engineering and Science, The University of New South Wales, Sydney, New South Wales 2052, Australia*
[m] *Hydrological Sciences Branch, NASA / GSFC, Greenbelt, MD, USA*
[n] *NOAA / GFDL, Princeton, NJ, USA*
[o] *Institute of Geography, Moscow, Russian Federation*
[p] *Climate Research Branch, Atmospheric Environment Service, Toronto, Ontario, Canada*
[q] *GKSS Research Center, Geesthacht, Germany*
[r] *Center for Ocean–Land–Atmosphere Studies, Calverton, MD, USA*

Received 8 September 1997; accepted 9 February 1998

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

PII: S0921-8181(98)00044-7

Abstract

Sixteen land-surface schemes participating in the Project for the Intercomparison of Land-surface Schemes (PILPS) Phase 2(c) were run using 10 years (1979–1988) of forcing data for the Red–Arkansas River basins in the Southern Great Plains region of the United States. Forcing data (precipitation, incoming radiation and surface meteorology) and land-surface characteristics (soil and vegetation parameters) were provided to each of the participating schemes. Two groups of runs are presented. (1) Calibration–validation runs, using data from six small catchments distributed across the modeling domain. These runs were designed to test the ability of the schemes to transfer information about model parameters to other catchments and to the computational grid boxes. (2) Base-runs, using data for 1979–1988, designed to evaluate the ability of the schemes to reproduce measured energy and water fluxes over multiple seasonal cycles across a climatically diverse, continental-scale basin. All schemes completed the base-runs but five schemes chose not to calibrate. Observational data (from 1980–1986) including daily river flows and monthly basin total evaporation estimated through an atmospheric budget analysis, were used to evaluate model performance. In general, the results are consistent with earlier PILPS experiments in terms of differences among models in predicted water and energy fluxes. The mean annual net radiation varied between 80 and 105 W m^{-2} (excluding one model). The mean annual Bowen ratio varied from 0.52 to 1.73 (also excluding one model) as compared to the data-estimated value of 0.92. The run-off ratios varied from a low of 0.02 to a high of 0.41, as compared to an observed value of 0.15. In general, those schemes that did not calibrate performed worse, not only on the validation catchments, but also at the scale of the entire modeling domain. This suggests that further PILPS experiments on the value of calibration need to be carried out.

Keywords: PILPS; land-surface parameterization; continental river basin modeling; energy/water balance; calibration of land-surface schemes; Red–Arkansas River basin

1. Introduction

The Project for Intercomparison of Land-surface Parameterization Schemes (PILPS) is a joint research activity sponsored by the Global Energy and Water Cycle Experiment (GEWEX) and the Working Group on Numerical Experimentation of the World Climate Research Program (Henderson-Sellers et al., 1995). Its goal is to improve the parameterization of the land-surface schemes used in climate and numerical weather prediction models, especially the parameterizations of hydrological, energy, and momentum exchanges. Its approach is to facilitate comparisons between models, and between models and observations, to diagnose shortcomings for model improvements. The PILPS philosophy is outlined by Henderson-Sellers et al. (1993, 1995).

PILPS was initiated in 1992, and consists of four phases. In Phase 1 (Pitman et al., 1993), 1 year of atmospheric forcings, generated from NCAR's general circulation model CCM1-Oz, were provided for grid cells in a tropical forest and northern grassland location. The experiments were designed in such a manner that the 1-year forcings were used repeatedly to iterate the model state variables to reach an equilibrium. The surface fluxes and state variables predicted by the 23 participating models were compared among themselves, with particular attention given to the partitioning of net radiation into latent and sensible heat fluxes, and of precipitation into evaporation and run-off.

In an attempt to understand the large scatter shown by the PILPS Phase 1(a) results for the different models, changes in the experimental design were made to assure that the models were physically self-consistent. The resulting experiments, referred to as PILPS Phase 1(c), included consistency checks on the convergence to a steady state, the balance of water and energy in their annual means, and the use of correct forcings. Additional supplementary experiments were also carried out with 100% vegetation cover and all combinations of prescribed albedo, prescribed aerodynamic resistance and saturated surface. The 'perpetual swamp' experiment was performed to check the evaporation consistency within each model when the surface was provided with sufficient water. Model forcing data were from the CCM1-Oz general circulation model for the tropical forest and grassland sites of Phase 1(a). Sixteen models passed the 'consistency check' experiments. The performance of the 16-model results was analyzed and inter-compared among themselves (Pitman et al., 1997; Koster and Milly, 1997).

In PILPS Phase 2(a) and Phase 2(b) experiments, the emphasis was expanded from model intercomparisons to evaluations using observed data. In Phase 2(a), point meteorological data for 1987 from Cabauw, The Netherlands, (51°58′N, 4°56′E) were used to force the land-surface schemes. Output from the models was compared with long-term measurements of surface sensible heat fluxes into the atmosphere and ground, with total net radiative fluxes and with latent heat fluxes derived from a surface energy balance. Evaluations on run-off generation could not be performed because the site was artificially drained. Calibration of the schemes with observations was not permitted. Chen et al. (1997) discussed the Cabauw experiment in detail.

In PILPS Phase 2(b) (Shao and Henderson-Sellers, 1995), a subset of the PILPS Phase 1 models participated in a November 1994 workshop at Macquarie University. The model results were compared with observed surface fluxes for a 35-day intensive observation period (IOP) from the HAPEX-MOBILHY experiment carried out in SW France in the summer of 1985, and with soil moisture measurements taken over that year. Streamflow data at the site were not available. Only the comparisons with nearby catchment run-off were conducted.

The Phase 2(a) and Phase 2(b) experiments represented major advances over the first phase of PILPS in that comparisons were made not only among the models themselves, but also with observations. However, there remained two major problems in those comparisons. The first is the mismatch in time between the comparisons of the models' results and the observations. Both the Cabauw and the HAPEX site have only 1 year of meteorological forcings. Thus, each model was required to use the 1 year forcings repeatedly until an equilibrium in the water and energy balances was reached, as was done in Phase 1. Therefore, the comparisons between model results and observations had to be based on the assumption that the inter-annual variability is small, so that the 1 year of observations provide a sufficient basis to evaluate each model at its equilibrium state. In those comparisons, the response of each model to multiple seasonal cycles cannot be studied. The second problem is the mismatch between the scale of the observations and the scale at which land-surface models are designed to be applied. The Cabauw site observations are essentially at a point scale. The HAPEX-MOBILHY Caumont site observations are at a small field scale. This is of concern even though the landscape surrounding these sites is fairly uniform.

The PILPS Phase 2(c) experiment resolved the mismatch in time by removing the assumption of the equilibrium year being similar to the observed year by conducting a 10-year simulation. The spatial scale mismatch was resolved by applying each land-surface scheme to a continental-scale river basin divided into computational units consistent with the grid scale of climate and weather prediction models. Utilizing a river basin as the modeling area permitted the incorporation of river flows as an evaluation variable—a variable which was unavailable in earlier PILPS experiments.

The major goal in the Phase 2(c) experiment is to evaluate the ability of current land-surface schemes to reproduce measured energy and water fluxes over multiple seasonal cycles across a climatically diverse, continental-scale basin. In designing the Phase 2(c) experiment, an additional objective was to test the ability of the schemes to calibrate their parameters using data from small catchments (on the order of 100s to 1000s of km^2) and to transfer this information from the calibration basins to other small catchments, and to the computational grid boxes.

This paper is the first of a three part series that describe the initial results from PILPS Phase 2(c). This paper (Part 1) discusses the overall design of the experiment, provides a description of the Red–Arkansas basins and the data, an overview of the participating models and submitted runs, presents results for the calibration–validation catchment runs, and intercomparison results from annual mean water and energy balance analysis. Part 2 focuses on intercomparisons of the energy fluxes across a range of spatial and temporal scales for the schemes as well as comparisons with regional evaporation estimates, while Part 3 focuses on similar analyses for the water fluxes and water balance including comparisons to observed river flow and regional evaporation.

2. Sources and experiment design

2.1. The Red–Arkansas River basin

The Arkansas and Red River basins are located in the southern Great Plains of the United States (see Fig. 1). The Arkansas River basin has an area of 409,273 km^2 and the Red River basin an area of 156,978 km^2 to give a total area of the combined basins of 566,251 km^2 which is represented for modeling purposes by 61 1° latitude/longitude computational grid boxes. The headwaters of the basins are at the continental divide of the Rocky Mountains; both rivers flow eastward to the Mississippi River. The Arkansas and the Red Rivers join the Mississippi near Little Rock, AR and near Shreveport, LA, respectively. The courses of the rivers are more or less parallel (see Fig. 2), and their climatologies are similar. For this reason they are treated as one water resources region by the US Geological Survey, and we consider them as a single hydrologic unit.

The basin has a large precipitation gradient with a climatology that ranges from arid and semi-arid in the west to humid in the east. For the years 1980–1986 the mean annual precipitation for the 61 grid cells was 767 mm yr^{-1}, with variations within the basin ranging from about 1400 mm yr^{-1} in the southeast to about 200 mm yr^{-1} in the arid western part of the basin. Precipitation increases at the highest elevations in the west near the continental divide, but the area strongly affected by orography is relatively small. Likewise, while snow processes are important in the headwaters, the area affected is small, and snow has a relatively small climatological and hydrological effect. Hydrologically, the run-off ratio tends to be quite small in the west, and higher in the eastern portion of the basins. Therefore, the hydrologic response of the basin is largely determined by the eastern part. However, the western part is subject to intense convective precipitation events. Vegetation generally ranges from grassland in the drier western parts of the basin to deciduous forest in the east, although, a large portion of the eastern region is cultivated.

The Red–Arkansas River basin has extensive data collection networks for meteorological and hydrological data. For this reason, these basins were the first large scale areas studied under the GEWEX Continental Scale International Project (GCIP). In addi-

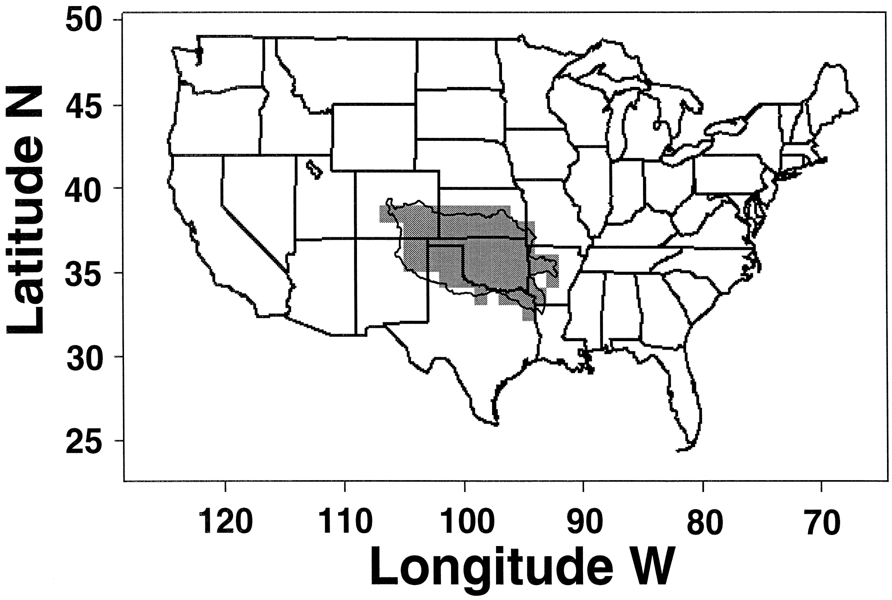

Fig. 1. Location within the USA of the Red–Arkansas River basin and the 61 computational 1° latitude/longitude grid cells.

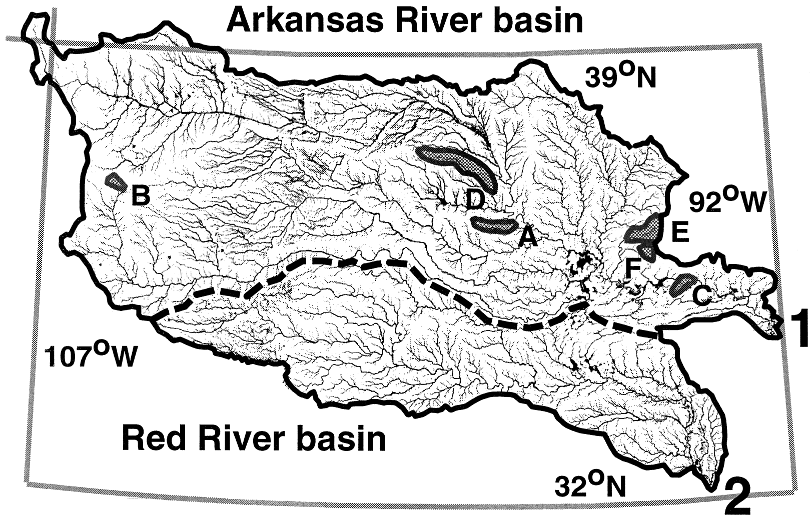

Fig. 2. River network and location of the six validation and calibration catchments within the Red–Arkansas River basin. Calibration catchments: 'A': Black Bear Creek (1491 km^2), 'B': Canadian River (593 km^2), 'C': Mulberry River (966 km^2). Validation catchments: 'D': Chikaskia River (4814 km^2), 'E': Illinois River (2483 km^2), 'F': Lee Creek (1103 km^2). '1' is the outlet of the Arkansas River basin and '2' is the outlet of the Red River basin.

tion, the basins contain a number of US Department of Agriculture, Agriculture Research Service (ARS) experimental catchments, the US Department of Energy Atmospheric Radiation Monitoring Cloud and Radiation Testbed (ARM-CART) site, and the recently established Cooperative Atmosphere-Surface Exchange Study (CASES) boundary layer facility.

2.2. Data sources

Three different types of data were provided to the Phase 2(c) participants: atmospheric forcing data, vegetation related parameters, and soil property parameters.

2.2.1. Atmospheric forcing data

The atmospheric forcing data include 10 years (1979 to 1988) of precipitation, air temperature, wind speed, surface pressure, relative humidity, incoming solar radiation and downward longwave radiation. The daily precipitation data were assembled by Abdulla (1995), from National Climatic Data Center (NCDC) cooperator stations and aggregated to $1 \times 1°$ grids by simple averaging. In most cases, there were two precipitation stations per grid cell. Daily precipitation was then adjusted to an hourly time step uniformly. As discussed in Section 4, this decision regarding the hourly distribution of the daily precipitation raised some concerns by the participants, and resulted in reruns of the original experiments which are reported in Section 4.2.

Air temperature data, at an hourly time step, were assembled from 26 NCDC Surface Airways stations across the basins. The data were interpolated from the Surface Airways stations to the center of each grid cell using least-distance squared weighting and were adjusted for elevation differences between the stations and the mean elevation of the grid cell using pseudo-adiabatic lapse rates. The data were linearly interpolated to 1/2 h for those models that ran at that time step. Surface pressure, relative humidity and wind speed were adjusted to the topographic mean elevation, interpolated from NCDC Surface Airways stations (hourly time step) to the center of each grid cell using least-distance squared weighting and interpolated to modeled time steps that only differed from hourly.

To estimate solar and longwave radiation, cloud cover (available as the hourly opaque fraction) was taken from Surface Airways stations and interpolated in a similar manner as for the other Surface Airways variables. Downward solar and longwave radiation were computed as follows. For shortwave radiation, clear sky surface radiation was first computed using standard relationships that account for the day of year, time of day, and an optical depth computed from surface humidity (see TVA (1972)). The clear sky radiation was subsequently used to compute surface solar radiation using a cloud cover attenuation formula. Downward longwave was estimated using surface air temperature, cloud cover, and humidity (TVA, 1972).

To meet the time step requirements of all participating schemes, the hourly precipitation and wind speed observations were interpolated into 30 min and aggregated into 3-h time steps uniformly.

2.2.2. Vegetation and soil parameters

The vegetation classifications and related parameters were obtained from the ISLSCP global 1° data

Table 1
List of participating models of PILPS Phase 2(c)

Model	Contact	Calibration[a]	Apply calibration knowledge[b]	Input time step	Output time step	Submission information[c]
ALSIS(A)	P. Iranejad Y. Shao	yes	yes	30 min	hourly	resubmitted
BASE(B)	C. Desborough A. Pitman	no	no	30 min	hourly	before workshop
BATS(C)	Z. Yang R. Dickinson	yes	yes[d]	30 min	hourly	resubmitted
BUCK(D)	A. Schlosser	no	no	3 h	3 h	before workshop
CAPS(E)	S. Chang M. Ek	yes	yes	30 min	hourly	before workshop
CLASS(F)	D. Verseghy	yes	no	30 min	hourly	resubmitted
IAP94(G)	Y. Dai Q. Zeng	yes	no	3 h	3 h	before workshop
ISBA(H)	J. Noilhan F. Habets	yes	no	30 min	hourly	before workshop
MOSAIC(I)	R. Koster	yes	yes	20 min	hourly	resubmitted
NCEP(J)	Q. Duan F. Chen K. Mitchell J. Schaake	yes	yes	30 min	hourly	resubmitted
PLACE(K)	A. Boone P. Wetzel	yes	yes	30 min	hourly	resubmitted
SEWAB(L)	K. Warrach H. Mengelkamp	no	no	30 min	hourly	resubmitted
SPONSOR(M)	A. Shmakin	no	no	3 h	3 h	before workshop
SSiB(N)	Y. Xue J. Wang	yes	yes	hourly	hourly	before workshop
SWAP(O)	Y. Gusev O. Nasonova	yes	yes	3 h	3 h	before workshop
VIC-3L(P)	X. Liang E. Wood D. Lettenmaier	yes	yes	hourly	hourly	before workshop

[a] Calibration: The ‘yes’ indicates that a model calibrated its model parameters based on the three catchments provided.
[b] Apply calibration knowledge: The ‘yes’ indicates that a model’s calibration knowledge is transferred to other grids in the base-runs. The ‘no’ indicates that a model’s calibration knowledge is not transferred to other grids in the base-runs, although the model did its calibration exercises for the three provided catchments.
[c] Submission information: The schemes with ‘before workshop’ indicate those that submitted their results before the workshop where initial intercomparison results were presented. The schemes with ‘resubmitted’ indicate those that resubmitted their 10-year base-run results after the workshop.
[d] BATS: Yes for the pre-workshop runs, partially for the resubmitted base-runs, see Section 4.

sets (Meesen et al., 1995). The parameters include: leaf area index, fraction cover of vegetation, greenness, roughness length, zero plane displacement height, and albedo. The monthly averages of the ISLSCP data were used for the Phase 2(c) modeling period (1979–1988), so there is an implicit assumption that the inter-annual variation in monthly vegetation parameters is negligible.

Model parameters related to soil properties were either obtained or estimated using soil information from the US Soil Conservation Service State Soil Geographic data base (STATSGO, 1994) Data from the STATSGO data base included soil texture (percent sand and clay), residual soil moisture and total soil depth, including the depths to the A and B horizons. The derived parameters included saturated hydraulic conductivity, saturated matric potential, porosity, soil moisture at field capacity, and wilting point. The derived saturated hydraulic conductivity, and saturated matric potential based on Cosby et al. (1984) and Rawls and Brakensiek (1985), were compared with literature values. On the basis of this comparison, final estimates were derived using the empirical equation of Cosby et al. (1984). For the Clapp–Hornberger '*b*' parameter, the average of the two empirical methods was used since this gives estimates close to the median values for a wide range of soils found in the basins. The STATSGO data are available at a 1-km spatial resolution, therefore, the derived soil parameters were computed at the 1-km scale and then averaged up to the 1° computational grids. The resulting soil parameters were either used by the models directly or were used by the modelers to derive model-specific parameters.

2.3. Experimental design and model runs

To evaluate the goals and objectives of PILPS 2(c), two different types of model runs are presented and analyzed. These are the following.

(1) Calibration and validation runs. For the calibration catchments, model forcing data along with observed streamflow data were provided to each modeling group (scheme). For the three additional validation catchments (see Fig. 2) only model forcing data were provided.

(2) Base-runs. For these, the models used 10 years (1979–1988) of forcing data to simulate the energy and moisture fluxes for each of the 1° grid cells. Some of the model parameters were provided (and fixed) as described in Section 2.2. Other parameters were either based on the calibration runs or were independently specified by the modeling group (see Section 3.2).

In addition to these runs, sensitivity runs were carried out for three 1° grid cells with different hydrologic characteristics. These sensitivity runs are not discussed here. Table 1 lists the sixteen PILPS Phase 2(c) participating schemes and their related base-run submission information. The model runs were to be submitted prior to a workshop held at Princeton University (October 28–31, 1996, hereafter referred to as the Princeton Workshop) where primary intercomparisons among schemes, and between schemes and observations were presented. Table 2 summarizes the variables submitted by the participants.

Table 2
Variables submitted by each model

No.	Variable	Unit
1	Year	–
2	Month	–
3	Day	–
4	Hourly-index	–
5	Precipitation	mm h^{-1}
6	Total evapotranspiration	mm h^{-1}
7	Surface run-off	mm h^{-1}
8	Subsurface run-off and/or baseflow	mm h^{-1}
9	Total canopy interception	mm
10	Root zone soil moisture	mm
11	Soil moisture in top 0.1 m	mm
12	Soil moisture in total soil column	mm
13	Surface radiative (effective) temperature	K
14	Absorbed solar radiation	W m^{-2}
15	Net radiation	W m^{-2}
16	Total surface latent heat flux	W m^{-2}
17	Evapotranspiration from canopy interception	W m^{-2}
18	Transpiration from vegetation	W m^{-2}
19	Evaporation from bare soil	W m^{-2}
20	Potential evaporation for two grid cells	W m^{-2}
21	Bare soil potential evaporation (same as above)	mm h^{-1}
22	Surface sensible heat flux	W m^{-2}
23	Surface ground heat flux	W m^{-2}
24	Aerodynamic conductance to vapor (heat) transport	m s^{-1}
25	Surface conductance	m s^{-1}
26	Surface albedo	–

As will be discussed in Section 4, seven schemes resubmitted their base-run results after the workshop for various reasons. The results presented in this paper are based on the most recently submitted runs for these models. For the other nine schemes, the analysis is based on the runs submitted before the workshop (with the exception of additional runs described in Section 4 designed to test the effect of the diurnal precipitation pattern). The schemes whose base-run results were resubmitted after the workshop are ALSIS, BATS, CLASS, MOSAIC, NCEP, PLACE, and SEWAB. The reasons for the resubmissions are briefly summarized in Section 4 where comparisons between the original and resubmitted runs are made for the mean annual water and energy balance, and mean monthly run-off and evaporation.

3. Analysis and results

3.1. Calibration–validation results

The purpose of the calibration–validation runs was to test the ability of schemes to calibrate their parameters using data from smaller catchments and to transfer this information to other similarly sized catchments and to larger computational grids. The runs are intended to provide insight as to whether such parameter calibration, widely used in hydrological modeling, would improve the performance of the land-surface schemes, even when the land-surface schemes use 'physically-based' parameters that in theory can be estimated from land cover characteristics such as soil or vegetation data. These calibration–validation runs are a first attempt within PILPS to address this issue.

The calibration and validation runs consisted of two parts. In the calibration runs, model forcing data for three catchments were provided along with observed streamflow data. These catchments, shown in Fig. 2, were Black Bear Creek (1491 km^2, designated by 'A' in Fig. 2, the Canadian River (593 km^2, designated 'B') and the Mulberry River (966 km^2, designated 'C'). For these catchments, the models were first run using their 'standard' parameter values. Then, using the streamflow data, adjustments were allowed to specific model parameters to improve the comparison between predicted and observed streamflow. The modeling groups provided

Table 3
Changes made by models for application to verification catchments based on calibration catchment results

Model	Calibration
ALSIS	a constant pre-infiltration run-off coefficient of 0.15 was introduced to the surface effective precipitation for all catchments
BASE	no calibration
BATS	reduced interception capacity multiplier from 0.2 to 0.01 for all catchments; increased lower layer K_{sat} to match surface K_{sat}
BUCKET	no calibration
CAPS	changed roughness ratio Z_0 h/Z_0 m (heat/momentum) from 0.1 to 10^{-5}
CLASS	no calibration
IAP94	K_{sat} (bottom layer) reduced; canopy interception capacity lowered
ISBA	infiltration shape parameter adjusted (increased in Mulberry River to produce more surface run-off; decreased in Black Bear and Canadian River to produce less surface run-off)
MOSAIC	diurnal variability of precipitation was imposed for all catchments; upper layer water holding capacity was adjusted for Black Bear and Canadian River (but not changed for verification catchments)
NCEP	changed K_{dt} (upper layer infiltration capacity) catchment-by-catchment
PLACE	increased pre-infiltration run-off ratio (direct diversion of precipitation) to 0.2 for all catchments
SEWAB	no calibration
SPONSOR	no calibration
SSiB	precipitation fraction coverage (convective vs. frontal; controls subgrid precipitation coverage) adjusted, baseflow parameter adjusted
SWAP	adjusted Manning's n and depth to water table
VIC-3L	changed non-linear baseflow parameters D_m and W_s to reduce 'fast' drainage from lower layer

the organizers with the model-derived streamflow time series before and after calibration.

For three additional validation catchments only forcing data were provided. These catchments were the Chikaskia River (4814 km^2, designated 'D' in Fig. 2), the Illinois River (2483 km^2, designated 'E') and Lee Creek (1103 km^2, designated 'F'), which served as the validation catchments. The modeling groups were asked to summarize what model parameters they adjusted in the calibration process. It should be noted that some groups varied parameters that should have remained fixed, so the potential for improvements due to calibration may be somewhat different than indicated in the results. Table 3 indicates how each model's parameters were adjusted during the calibration runs.

The range of approaches to parameter calibration can be grouped into four classes.

(1) Six models (ALSIS, ISBA, NCEP, PLACE, SSiB and VIC-3L) empirically adjusted the model parameters so as to fit the calibration data.

(2) Six models (BATS, CAPS, IAP94, MOSAIC, SSiB, SWAP) varied their internal representation of various processes, using their knowledge about model sensitivities, so as to fit the calibration data.

(3) One model (MOSAIC) varied the precipitation forcing diurnal pattern, which was assumed uniform in the original data set (see Section 4.2), by imposing a new daily cycle.

(4) Five models (BASE, BUCK, CLASS, SEWAB, SPONSOR) did not calibrate.

SSiB appears in two groups because it adjusted its baseflow parameter (group 1) and adjusted the parameter which describes the spatial heterogeneity of precipitation (group 2).

The data were run at an hourly time step and the results aggregated to a monthly time interval for the analysis. This aggregation captures the seasonal cycle of the water balance dynamics without having to consider the short-term dynamics of routing water through the catchments. Fig. 3 shows the effect of calibration on model performance for the calibration

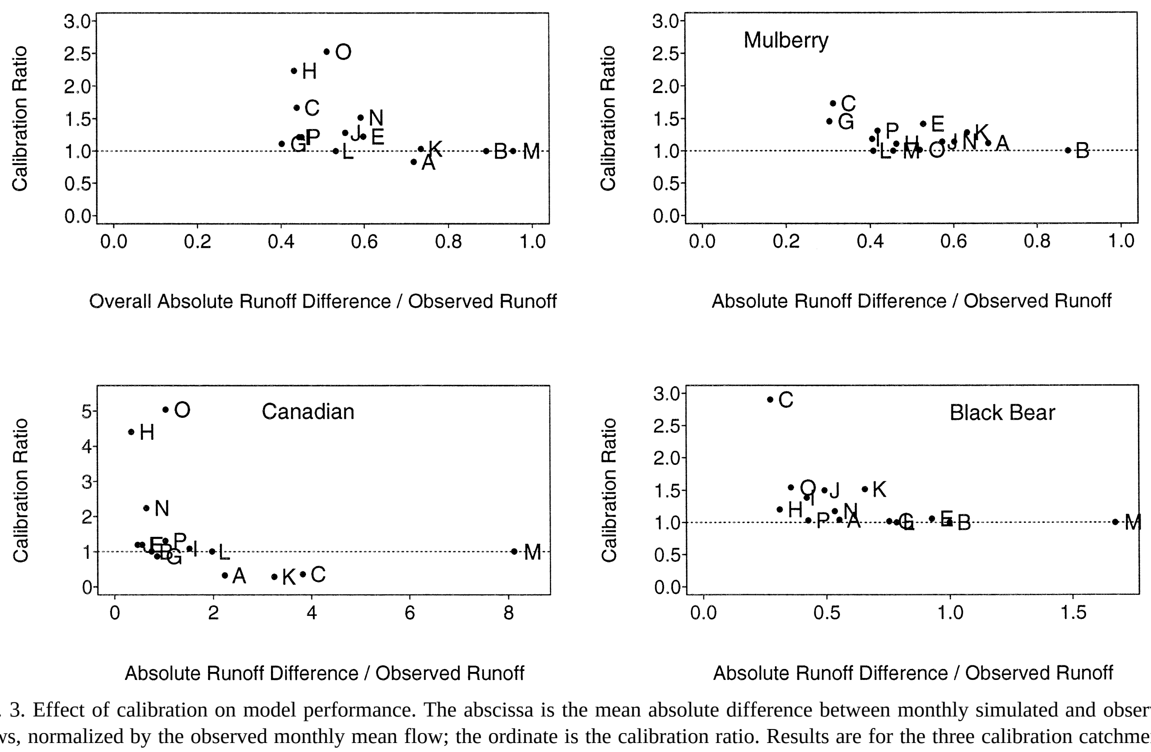

Fig. 3. Effect of calibration on model performance. The abscissa is the mean absolute difference between monthly simulated and observed flows, normalized by the observed monthly mean flow; the ordinate is the calibration ratio. Results are for the three calibration catchments, and the average over the three calibration catchments (upper left).

catchments. The upper left panel shows results averaged over the three calibration catchments, while the other three panels show results for the individual calibration catchments. The horizontal axis indicates the average monthly absolute deviation, normalized by observed streamflow, after calibration. The vertical axis, labeled the calibration ratio, is the ratio of the mean absolute error before and after calibration, and indicates the degree of improvement in model performance due to calibration. Values less than 1.0 indicate poorer performance after calibration. For some models the improvement is by a factor of almost 5 (SWAP and ISBA). The improvement was not consistent across the three catchments. For example, calibration significantly improved BATS' performance on the Black Bear catchment, improved it somewhat on the Mulberry, and degraded its performance on the Canadian. For reference, those models that did not calibrate but submitted uncalibrated runs are shown (BASE, 'B', and SPONSOR, 'M') and plotted on the horizontal line, 1.0. Overall, calibration improved the estimated streamflow from SWAP, ISBA and BATS the most.

Fig. 4 shows the model results for the three validation catchments. Here the vertical axis is the ratio of the modeled streamflow to observed streamflow. Ratios larger than 1.0 occur when models over-predict the streamflow and less than 1.0 under-predict. The model parameters applied during the validation runs were as estimated from the calibration catchments. Although the method of transferring information from the calibration to the validation catchments was left to the judgment of the modelers, most opted to apply calibrated parameters from the geographically closest calibration catchment. Calibrated models in general performed better than uncalibrated models in these validation runs. This result is notable because of the spatial scale difference (generally more than one order of magnitude) between the calibration catchments and the area of the 1° grid cells. Thus, the results suggest the desirability of model calibration and improved performance of

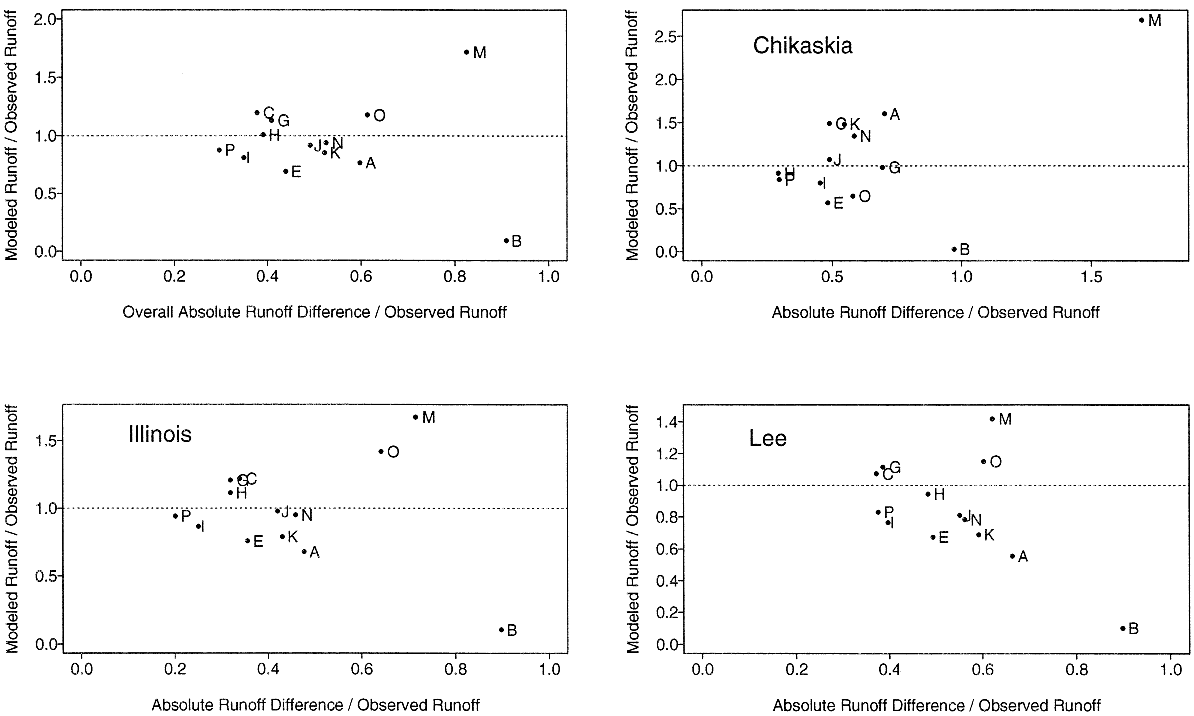

Fig. 4. Simulation results for the three validation catchments. The abscissa is the mean absolute difference between monthly simulated and observed flows, normalized by the observed monthly mean flow; the ordinate is the modeled runoff divided by the observed.

land-surface schemes. It is recognized that these results are not definitive and that PILPS should consider organizing more extensive calibration–validation experiments.

3.2. Approaches to transferring information from catchments to grid-scale

Each of the modeling groups was asked to summarize how information from the test catchments was transferred to the grid scale. The responses are summarized in Table 4. Again, it should be noted that during the calibration runs some modelers varied parameters that should have remained fixed according to the guidelines provided by the experiment organizers. It is unclear how these varied parameters, when transferred to the 61 grid boxes, influenced their results. In addition, some modelers (e.g., ISBA) chose not to use the calibrated parameters in the 10-year base-runs. A number of important issues remain to be resolved with respect to calibration of land-surface models-most importantly, how many calibration basins are necessary, what objective functions should be used for model calibration, and how to transfer calibration information between scales (i.e., from intermediate scale catchment to the region).

3.3. Subsection intercomparisons over the 10-year base-runs

In conducting the base-runs, each model initialized its soil moisture at half of its soil saturation for each soil moisture layer. The canopy interception was initialized at zero for all models. The soil and surface temperatures were initialized to the air temperature at the first modeling time step. Each model was run using the 10-year forcing data set with the same initial condition. The required output variables at each hourly time step for each of the 61 grid cells are listed in Table 2, which include model-computed fluxes, state variables, and selected forcings. For the schemes which do not have an easy way to output the required variables as listed in Table 2, −999 was reported instead.

Ten years of forcing data (1979–1988) were used to conduct the base-run simulations. The 1st year, 1979, was eliminated from the analysis to remove any initialization effects. In addition, there were concerns about some of the data used for the atmo-

Table 4
Parameter changes and information transfer for the base-runs using calibration–validation run results

Model	Changes applied to regional scale
ALSIS	none
BASE	did not calibrate
BATS	none in pre-workshop runs, used interception capacity = 0.2 mm in the base-run resubmission
BUCKET	did not calibrate
CAPS	changed roughness ratio (momentum to heat) everywhere same as calibration catchments; set zero plane displacement height to zero everywhere
CLASS	did not calibrate
IAP94	none
ISBA	none
MOSAIC	areal storm fraction from the calibration catchments
NCEP	changed K_{dt} (upper layer infiltration coefficient) by classifying grid cells (according to precipitation climatology) as (a) arid/semi-arid; (b) semi-humid; (c) humid; applying calibration changes from Canadian River; Black Bear Creek, Mulberry River to a, b, c, respectively
PLACE	none
SEWAB	did not calibrate
SPONSOR	did not calibrate
SSiB	applied change in proportion of convective vs. large scale precipitation uniformly to entire region
SWAP	linearly interpolated depth to water table from calibration catchments; applied geometric mean of Manning's n from calibration catchments uniformly to entire region
VIC-3L	scale soil properties; adjusted W_s based on grid scale field capacities

spheric budget computations which provided the regional evapotranspiration estimates. Therefore, only the estimates from 1980 to 1986 from each scheme were used in the analyses presented here.

3.4. Water and energy balance checks

Based on surface flux outputs from each scheme, an energy balance check was conducted. The results are shown in Fig. 5, in which the abscissa gives the mean annual ground heat flux in W m^{-2} for the period 1980–1986, aggregated over the 61 1° grid cells. The ordinate shows the mean annual energy residuals for each scheme over the same period. The residuals are expressed by δ_1,

$$\delta_1 = \overline{R}_n - \overline{E} - \overline{H} - \overline{G} \quad (1)$$

where $\overline{R}_n$, $\overline{E}$, $\overline{H}$, and $\overline{G}$ represent the 7-year mean annual net radiation, latent, sensible, and ground heat fluxes, respectively. If a scheme conserves energy, then its energy residuals should be on the zero (dotted) line in Fig. 5. From Fig. 5, it is seen that all of the schemes have residuals less than ± 3 W m^{-2} which is the criterion used for the consistency checks in earlier PILPS experiments. It is expected that the mean ground heat flux over the seven years (1980–1986) should be close to zero. This occurs for most schemes with the exception of SWAP, IAP94, SPONSOR, ISBA and VIC-3L. The first three have mean annual ground heat fluxes greater than 4 W m^{-2} with SWAP having the highest annual mean ground heat flux at 6.05 W m^{-2}. ISBA loses ground heat with a mean annual ground heat flux of -3.08 W m^{-2}. VIC-3L has a mean ground heat flux of 1.21 W m^{-2}, probably due to using the same soil temperature as the lower boundary for all 61 grid cells, even though the model can specify different temperatures as the lower boundary for different grid

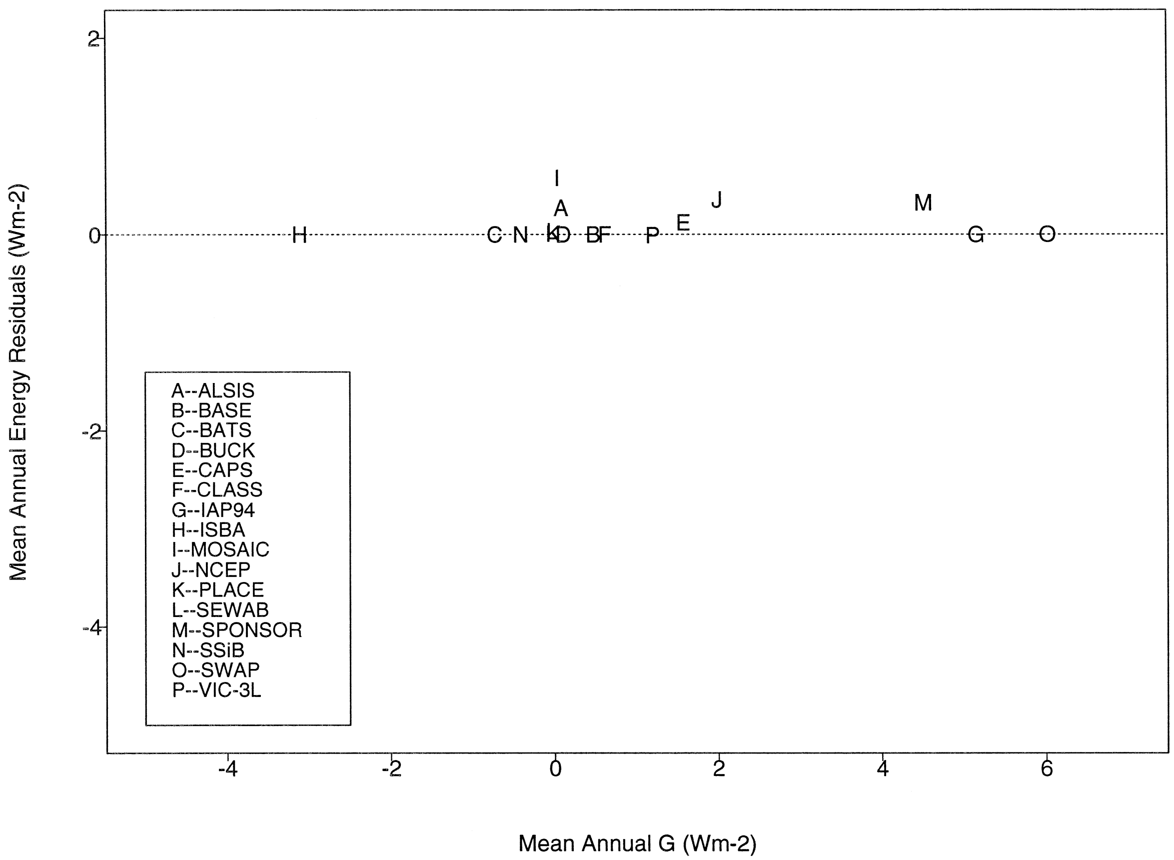

Fig. 5. Mean energy balance residual for the years 1980–1986. The abscissa shows the mean ground heat flux in W m^{-2}.

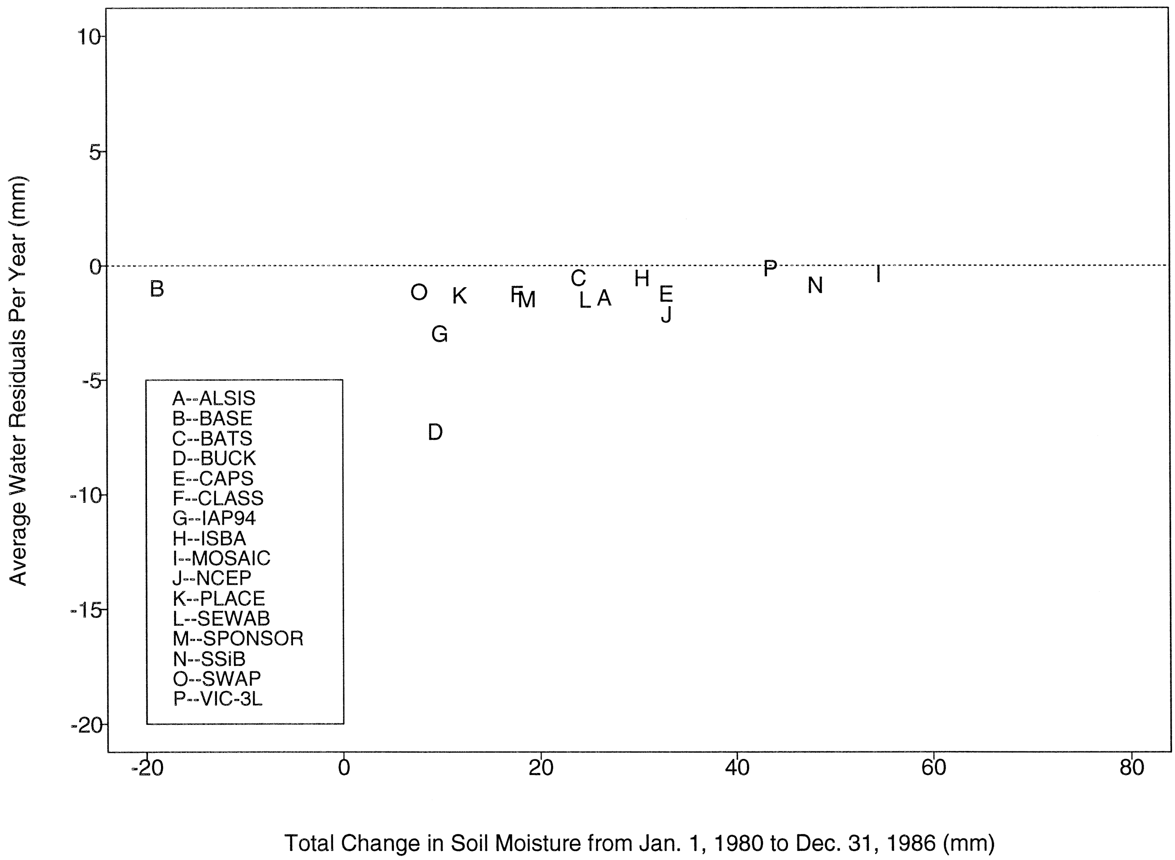

Fig. 6. Average annual water balance residual for the years 1980–1986. The abscissa shows the moisture change between 12/31/1986 and 1/1/1980 in millimeter.

cells. Some of the schemes, e.g., MOSAIC, ALSIS, NCEP and SPONSOR, have slightly positive annual energy residuals of about 0.5 W m^{-2}. These may arise from phase-change processes related to snow melt and frozen soils which are not completely accounted for in Eq. (1). Additional discussion of this issue can be found in the energy flux companion paper (Liang et al., this issue).

Fig. 6 shows the conservation of water for the period of 1980–1986 for all of the schemes. The *x*-axis gives the total change in soil moisture between 1980 and 1986, and the *y*-axis shows the mean annual water residual, δ_2 which is computed using

$$\delta_2 = \sum P - \sum E - \sum R_{\mathrm{surf}} - \sum R_{\mathrm{sub}} - (W_{86} - W_{80}) \quad (2)$$

where $\sum P$, $\sum E$, $\sum R_{\mathrm{surf}}$, $\sum R_{\mathrm{sub}}$ represent the total precipitation, evaporation, surface run-off, and subsurface run-off for the 7 years, respectively; $W_{86} - W_{80}$ represent the total change in soil moisture storage over the period 1980 to 1986. Fig. 6 shows that all the models, except for BUCK, satisfy the water balance consistency criterion of having residuals less than 3 mm yr^{-1}, which was used in the PILPS Phase 2(a) (Cabauw) and Phase 2(b) (HAPEX) experiments. For BUCK, the residuals exceeded 7 mm yr^{-1}. However, BUCK is still within the PILPS Phase 1(c) water balance criterion of 0.1 kg m^{-2} day^{-1} which is equivalent to 0.1 mm day^{-1} or 36.5 mm yr^{-1}. Thus, using the water and energy balance criterion established for earlier PILPS experiments, all of the participating schemes conserve energy and water over Phase 2(c) experiment period and domain.

3.5. Mean annual energy and water balance intercomparisons

The initial analysis of Phase 2(c) model runs, after checking that the schemes conserve energy and water, focuses on the long-term components of the balances. For the energy balance, this is the mean annual sensible heat and latent heat which is balanced by the net radiation assuming a mean annual ground heat flux of zero and an energy conserving scheme. For the water balance, the mean annual precipitation is balanced by the mean annual run-off and evapotranspiration, again assuming no long term change in soil moisture storage and a water conserving scheme.

Figs. 7 and 8 present results for the schemes, averaged over the analysis period (1980–1986) and aggregated over the Red–Arkansas basin. In Fig. 7, the mean annual sensible heat is plotted against the mean annual latent heat. In Fig. 8, the mean annual run-off is plotted against the mean annual evapotranspiration. In addition, observational data are also plotted on Figs. 7 and 8. The observational data consist of river discharge data for the outlets of the and Red and Arkansas Rivers and an estimate of total basin evaporation. The basin evapotranspiration estimates are based on an atmospheric budget analysis using radiosonde data for the convergence and change in atmospheric moisture storage terms. Over the 7-year analysis period, we feel confident that this yields an accurate estimate of the basin evapotranspiration. In fact, given that the run-off and evapotranspiration observations, each established independently, sum to the observed precipitation to within 0.5% supports the reliability of these estimates.

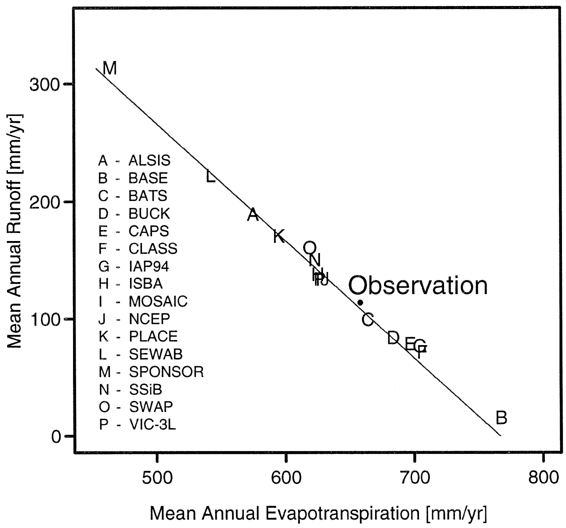

Fig. 8. Mean annual water balance (1980–1986) for the Red–Arkansas River basin.

In Figs. 7 and 8, it is possible to investigate how the land-surface schemes partition net radiation between sensible and latent heat, and precipitation between run-off and evapotranspiration. The former leads to the Bowen ratio (sensible/latent heat) and the latter to the run-off ratio (run-off/precipitation). The results shown in Figs. 7 and 8 are consistent with those found in the PILPS Phase 1(a) and Phase 2(a) results (Pitman et al., 1993; Chen et al., 1997).

In Fig. 7, there is considerable scatter from the net radiation line (i.e., the sum of the mean sensible and latent heat varies from scheme to scheme) suggesting that the computation of the net radiation from each scheme varies considerably. This difference is analyzed in Part 2 of this series (Liang et al., this issue) and is due to differences in albedo and the surface temperature computations. Observations for net radiation are unavailable for the Red–Arkansas basin domain. For reference two lines of equal net radiation are shown, 80 and 105 W m^{-2}. All the

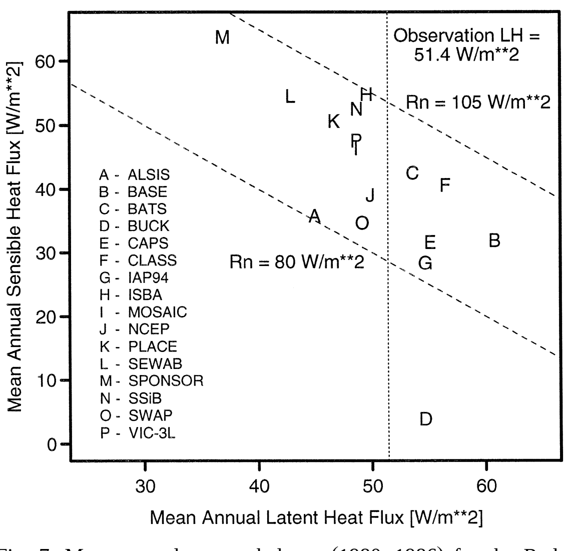

Fig. 7. Mean annual energy balance (1980–1986) for the Red–Arkansas River basin.

schemes except BUCK fall within these limits. As explained more fully by Liang et al. (this issue), BUCK computes very high surface temperatures leading to low net radiation and low sensible heat. Also shown in Fig. 7 is the atmospheric budget-derived mean annual evapotranspiration which is 51.4 W m^{-2}. The majority of the schemes fall within about 5% of the mean latent and sensible heat values. Exceptions are SPONSOR which has the lowest latent heat (36.9 W m^{-2}) which is 14.5 W m^{-2} lower than the basin-wide estimate, BASE which has the largest mean annual latent heat (60.8 W m^{-2}), and BUCK, discussed earlier, which is an outlier among all the schemes with its very low sensible heat of 4.0 W m^{-2}. This result is consistent with BUCK's performance in other PILPS experiments. Although in the earlier experiments (e.g., Cabauw Phase 2(a), see the work of Chen et al. (1997)) the net radiation computed by BUCK was consistent with other schemes, in the PILPS Phase 2(c) results, the net radiation computed by BUCK was significantly lower than that of the other schemes for reasons that are unresolved.

ALSIS, CAPS, IAP94, NCEP and SWAP compute net radiation that is about 10 to 15% lower than the all-model average. These schemes have latent heat estimates close to the basin-wide estimated value, so this deviation appears as a deficit in the sensible heat term. This suggests that differences in net radiation may not be partitioned equally between latent and sensible heat fluxes but will appear in one term or the other, depending on land-surface conditions. This supports similar conclusions in the work of Peters-Lidard et al. (1997) based on analysis using data from the First ISLSCP Field Experiment (FIFE), and warrants further comparisons to data and analyses of the performance of land-surface schemes over diverse climates.

Table 5 presents the Bowen ratios from the schemes based on the mean annual sensible and latent heats. As can be seen from Table 5, there is wide variation in these ratios, suggesting that the schemes predict quite different energy balance climatologies for the region.

Fig. 8 shows the mean annual water balance for all schemes. Small deviations from the line ($P = R + E$) are due to schemes having small annual changes in soil moisture as shown in Fig. 6. Also plotted in Fig. 8 are the observed basin run-off and derived evaporation. Eleven schemes fall within 5% of the observed values, with ALSIS, BASE, PLACE, SEWAB and SPONSOR falling outside. As for the energy balance, SPONSOR has the lowest evapotranspiration (464.3 mm yr^{-1}) which results in its water balance as having high run-off (313.8 mm yr^{-1})—which is about 2.8 times higher than the observed run-off of 112.0 mm yr^{-1}. This result contrasts with the results in the Cabauw PILPS Phase 2(a) experiment in which SPONSOR had higher than average evapotranspiration and lower run-off.

Table 5
Mean annual (1980–1986) Bowen ratio and run-off ratio averaged over the 61 1° latitude/longitude grid cells for all models

Model	Bowen ratio	Run-off ratio
ALSIS	0.800	0.247
BASE	0.525	0.020
BATS	0.793	0.130
BUCK	0.072	0.109
CAPS	0.573	0.103
CLASS	0.718	0.093
IAP94	0.519	0.100
ISBA	1.104	0.180
MOSAIC	0.952	0.175
NCEP	0.783	0.175
PLACE	1.082	0.223
SEWAB	1.271	0.290
SPONSOR	1.726	0.409
SSiB	1.078	0.196
SWAP	0.705	0.209
VIC-3L	0.978	0.174

These differences were explained in part by Koster and Milly (1997) who found that SPONSOR has a low rate of interception loss. This reduced evaporation from interception storage has a larger effect in the Red–Arkansas experiment than in Cabauw, which was a grassland site with relatively low precipitation rates. In addition, the current version of SPONSOR seems to produce too much run-off, resulting in dry soils and low soil evaporation (Shmakin, personal communication). The linkage of the energy and water balances through evapotranspiration demonstrates that weaknesses in the parameterization of one process affects the flux estimates throughout the scheme.

BASE produced almost no run-off (15.7 mm yr^{-1}), recycling almost 100% of the mean annual precipitation. Table 5 also lists the run-off ratios (R/P) for the various schemes. Additional analysis regarding the run-off estimates and water balance from the schemes is presented in the Part 3 companion paper of Lohmann et al. (this issue).

4. Post-workshop model re-runs

During the workshop, two issues arose that resulted in some of the participants resubmitting the base-runs. The first issue related to inconsistencies between the experiment's protocols and the submitted runs. In addition, some participants submitted output with inadvertent errors, found coding errors in their models, and/or felt that the parameters from the calibration–validation were inappropriate for their schemes. These resubmissions are referred to as the resubmitted base-runs (see Section 4.1). The second issue that arose in the workshop was the use, in the original forcing data, of uniform (daily average) precipitation throughout a rain-day, and how this could influence the results. Post-workshop model runs using a precipitation data set that includes the observed hourly pattern are referred to as the disaggregated precipitation reruns (see Section 4.2). In Sections 4.1 and 4.2, we compare the pre- and post-workshop submissions.

4.1. Resubmitted base-runs

Seven schemes resubmitted the 10-year base-runs after the workshop: ALSIS, BATS, CLASS, MOSAIC, NCEP, PLACE and SEWAB. The reasons for the resubmission are briefly summarized below. ALSIS resubmitted due to bad soil parameters retained from test runs, resulting in inconsistent results in the earlier submission. This modeling group did not attend the Princeton Workshop.

BATS resubmitted due to deviations of their procedure from the workshop instructions with respect to modification of model parameters during the calibration runs, the effect of which was that the interception capacity multiplier was changed from 0.2 mm to 0.01 mm. BATS resubmitted using 0.2 mm interception capacity in order to have the same basis for intercomparisons of the base-runs.

CLASS resubmitted twice. CLASS felt that the first base-run version (with adjusted parameters from the calibration run) did not correspond to any published version of CLASS and preferred to resubmit results using its standard parameter values; i.e., an uncalibrated run. The second resubmission was owing to the discovery of a recently introduced bug in the surface mixed layer formulation.

MOSAIC originally submitted using (with the permission of the organizers) a different diurnal precipitation pattern. In order to have the same basis for intercomparisons among schemes, MOSAIC resubmitted the 10-year base-run using uniform (daily average) precipitation. In the resubmitted base-runs, instead of using the calibration data to adjust the temporal partitioning of precipitation, MOSAIC used these data (and these data only) to calibrate their areal storm wetting fraction. The calibration–validation analysis for MOSAIC (see Section 3.1) is based on the original runs using the non-uniform precipitation.

NCEP resubmitted due to a coding error as a result of code transfer and due to the adjustment of a constant related to the calculation of potential evapotranspiration.

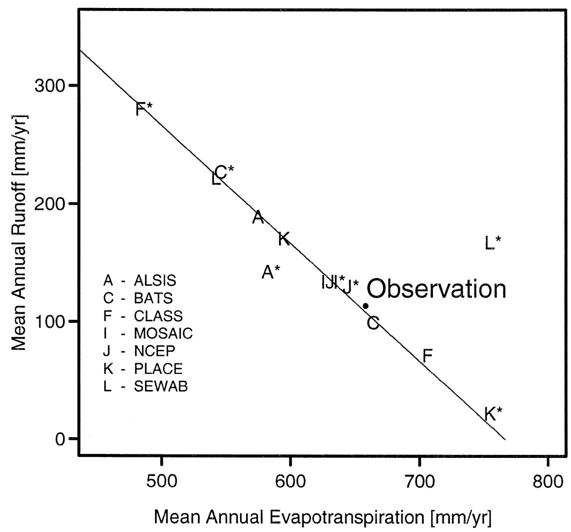

Fig. 9. Mean annual water balance (1980–1986) for the Red–Arkansas River basin for the seven models which resubmitted base-runs after the workshop. The old results are shown with asterisks.

PLACE resubmitted due to not using their calibration information in the initial submission. The resubmission is more consistent by applying knowledge from the calibration procedure.

SEWAB resubmitted due to coding errors found in their canopy evaporation parameterization.

Fig. 9 shows the mean annual water balance and the partitioning of precipitation into evapotranspiration and run-off for the original and resubmitted runs of the seven schemes. The asterisk refers to the pre-workshop model results. Two of the models did not conserve water before the resubmission but did so afterwards (ALSIS, SEWAB). Three models significantly improved their water balance results (BATS, CLASS and PLACE). PLACE had increased run-off while BATS and CLASS decreased their run-off. Two models (MOSAIC, NCEP) had only small changes in their water balance results. Fig. 10 shows the mean monthly run-off and evaporation for the resubmitted runs along with the observations. In the resubmitted run, PLACE captured the seasonality in the run-off even though they over-predict in the summer and fall periods. The over-prediction is most likely due to the limitations imposed on the calibration runs and the fact that PLACE turned off its heterogeneity modules (soil moisture, texture and surface heterogeneity) to be consistent with their earlier PILPS runs. With these conditions, PLACE had no adjustable parameters to calibrate. In the resubmitted runs, PLACE developed a procedure that extrapolated the information gained from the three calibration catchments across the entire 61 grids with a focus on improving the surface run-off to take account of sub-grid heterogeneity.

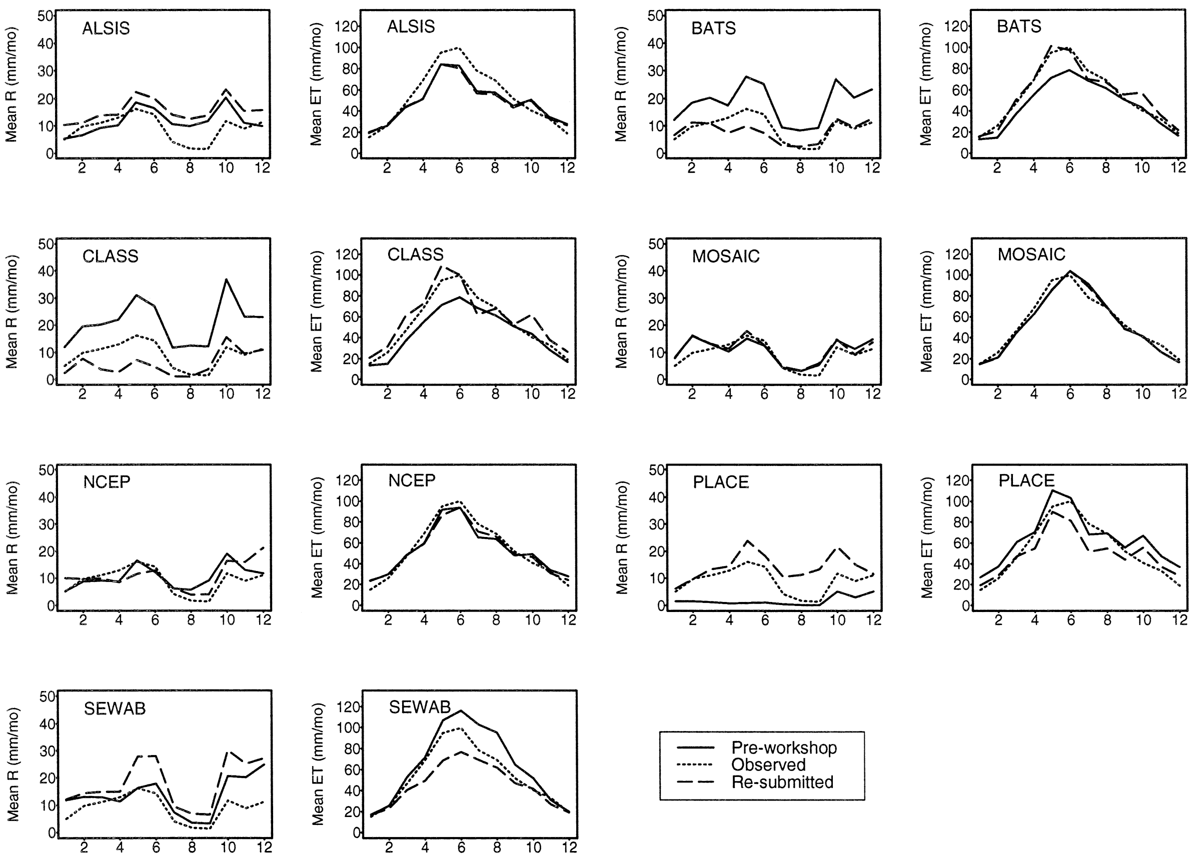

Fig. 10. Mean monthly runoff (R) and evapotranspiration (ET) (1980–1986) for the Red–Arkansas River basin for the seven models which resubmitted base-runs after the workshop.

BATS and CLASS improved significantly their run-off prediction, especially during the summer, but tended to under-predict spring run-off. ALSIS and SEWAB had poorer run-off results after the resubmission; ALSIS had a small change to its evapotranspiration and SEWAB went from over-predicting evapotranspiration to under-predicting. BATS, CLASS, and PLACE all improved their seasonal evapotranspiration. NCEP and MOSAIC showed little change in their monthly results.

Fig. 11 summarizes the changes in the energy balance. The results show that one of the schemes (NCEP) which previously deviated from the $R_n = E + LH$ line was much closer following resubmission. Initial analysis of the NCEP results suggested that energy was not conserved. After the workshop, an error in the ground heat was found and the runs resubmitted. Thus, in Fig. 11 NCEP has increased net radiation with the decreased ground heat flux appearing as sensible heat. The remaining schemes show little change to their computed net radiation. Furthermore, most schemes that resubmitted did not lie closer to the observed latent heat after post-workshop resubmission.

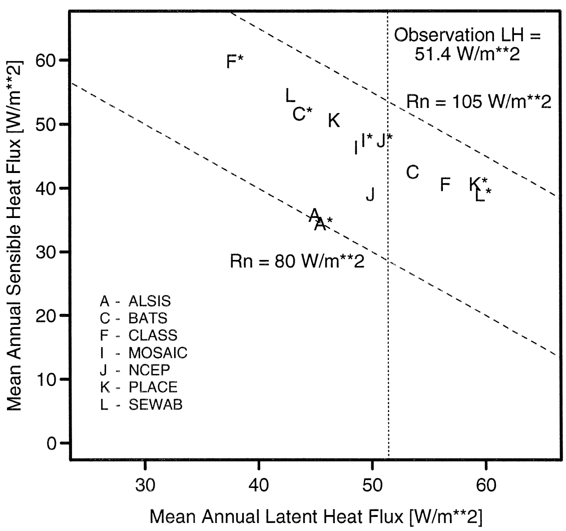

Fig. 11. Mean annual energy balance (1980–1986) for the Red–Arkansas River basin for the seven models which resubmitted base-runs after the workshop. The old results are shown with asterisks.

4.2. Disaggregated precipitation re-runs

At the Princeton Workshop there was discussion about the assumption of daily uniform precipitation used for the original forcing data, and how this might have affected the results. This discussion prompted the resubmission of the MOSAIC base-runs, as discussed above. At the request of the Princeton workshop participants, a data set with hourly disaggregated precipitation and fractional precipitation coverage (of the 1° grids) was created. Hourly manually digitized radar (MDR) data (see the work of Baeck and Smith (1995) for detailed descriptions of MDR data) were used to examine the space–time characteristics of the precipitation events. MDR grid spacing is approximately 40 km in the Red–Arkansas River basins. Radar reflectivities are coded into 6 VIP (video integrator and processor) levels which were translated into rainfall rates using the Marshall–Palmer *Z–R* relationship (Austin, 1987; Doviak and Zrnic, 1993).

In order to have comparable precipitation totals to the original gage-based Phase 2(c) precipitation, the MDR data were used only to provide the hourly rainfall pattern, the relative hourly depths, and the rainfall fractional coverage for the 1° grids. The daily totals were then temporally distributed on this basis and adjusted for fractional coverage so as to maintain the original station-based total precipitation.

Eight schemes submitted re-runs for analysis using the revised precipitation data (ALSIS, BASE, BATS, CAPS, ISBA, MOSAIC, PLACE, SSiB). Fig. 12 shows the mean annual water balance for the original runs which are denoted with asterisks. The results show that seven of the schemes had an increase in the run-off ratio, with the increases ranging from 0.002 (MOSAIC) to 0.048 (BATS). Only SSiB had a decrease in the run-off ratio (−0.014). In SSiB's rerun, an adjustment for spatial precipitation (fractional coverage, based on a convective parameterization used in the original runs) was turned off when the new precipitation data set was used. This probably explains the difference in SSiB's response compared to the other schemes.

The largest run-off ratio change (by BATS) represents a change of about 35% in its run-off ratio corresponding to an increase in mean annual run-off of 37 mm, which corresponds to a 5.4% decrease in

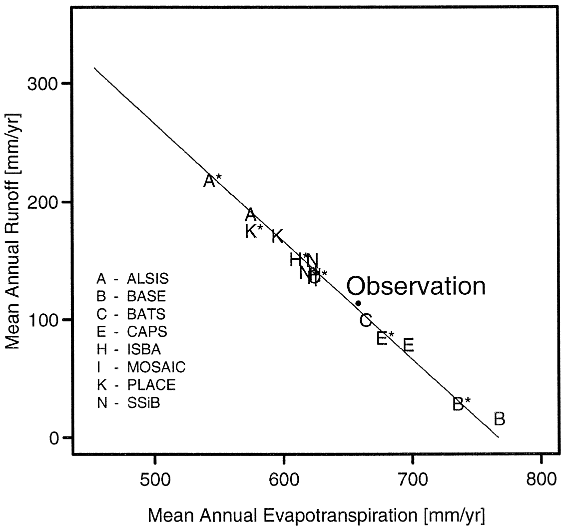

Fig. 12. Comparison of mean annual water balance results (1980–1986) for the Red–Arkansas River Basin, using the original base-run results and the resubmitted disaggregated precipitation rerun results, shown with asterisks.

mean annual evapotranspiration. NCEP, which did not submit the disaggregated precipitation re-runs in time to be analyzed, reported comparable changes to BATS. NCEP found an increase in their annual run-off ratio of 29.6% (from 0.17 to 0.22) and an even larger run-off ratio increase in summer of 61.9% (from 0.104 to 0.168) (Duan, personal communication.)

For reference, the observed run-off ratio is 0.15 and the schemes had a range in run-off ratio from 0.02 to 0.41. The largest changes in the run-off ratios (BATS and NCEP) due to the hourly precipitation pattern appear large in percentage terms but are much less than the variation in run-off ratios among the schemes. Conclusions regarding the hourly rainfall pattern are mixed. For many schemes, the effect of the precipitation diurnal pattern can be significant with respect to its run-off ratio. On the other hand, these large percentage differences in run-off ratio have a small effect on the total mean annual partitioning of water and energy. This is due to the small run-off ratio which exists in the Red–Arkansas basin. An important implication from these reruns is that the total precipitation volume and the parameterization of the hydrologic processes within each scheme often dominate the results. Overall, it is clear that further work is needed to improve the parameterization of run-off in land-surface schemes.

5. Discussion and conclusions

The main goal of PILPS Phase 2(c) was to evaluate the ability of current land-surface schemes to reproduce measured energy and water fluxes over multiple seasonal cycles across a climatically diverse, continental-scale basin. In addition, the experiment was intended to test the ability of the schemes to calibrate their parameters using data from smaller catchments and transferring this information to other basins and to the 1° computational grid cells.

We found that a number of schemes benefitted from parameter calibration, with SWAP and ISBA and BATS showing the most improvement. In general, improvements in predictions of energy and water fluxes as a result of calibration carried over to the validation basins where all of the schemes that were calibrated performed well (Fig. 4). Schemes that did not calibrate (BASE, SEWAB, SPONSOR) generally did not perform as well. As shown in Fig. 4, BASE and SPONSOR had the poorest performance for the validation catchments. SEWAB did not submit runs for the validation basins and BUCK did not submit either calibration basin or validation basin runs. BASE, SEWAB, SPONSOR had the poorest performance with respect to the mean annual water balance (see Fig. 8) and arguably for the mean annual energy balance (see Fig. 7). Given that the mean annual net radiation for these schemes was close to the all-model average, it appears that their overall performance could be improved through calibration.

SWAP's improvement in estimating run-off as a result of calibration (Fig. 3), appears to have been carried through into the water balance (Fig. 8), which shows an accurate mean annual evaporation estimate. SWAP's poor performance with respect to the energy balance (Fig. 7) is perhaps related to its surface temperature calculation. In fact, the models with low net radiation (ALSIS, IAP94, CAPS, SWAP and NCEP) generally performed well for the validation runs, suggesting that their land-surface water balance parameterizations are reasonable.

The results discussed here strongly suggest that there is value in using catchment data to calibrate the parameters of land-surface schemes. One possible implication for global implementation is the desirability of establishing a global set of calibration catchments that could be used by land-surface schemes for parameter estimation.

There were significant differences among the schemes with respect to partitioning water and energy on an average annual basis. For example, the mean annual Bowen ratio varied from 1.73 (SPONSOR) to 0.52 (BASE, IAP94) to the anomalously low 0.07 of BUCK. Based on data analysis, we believe that the regional mean annual Bowen ratio is about 0.92. Similarly the run-off ratios varied from a low of 0.02 (BASE) to a high of 0.41 (SPONSOR) as compared to the observed regional run-off ratio of about 0.15. Further detailed analysis of the energy fluxes and water fluxes are given in Parts 2 and 3 of this paper.

The sensitivity of the schemes to changes in the diurnal pattern of precipitation (Fig. 12) can be significant but is much smaller than the differences among schemes in partitioning run-off and evaporation. The most sensitive scheme, BATS, had a 35% increase in its mean annual run-off ratio due to the diurnal pattern and NCEP reported (Duan, personal communication) similar annual and larger summertime sensitivities. Some schemes are more sensitive to changes in forcings than others: models originally developed as surface hydrology models or for climate applications might show less sensitivity than other schemes in that they have been developed and tested extensively using climate time and spatial scales. These schemes are more likely to include implicit parameterizations to represent the heterogeneities in surface run-off response and energy balances, and are likely to be more amenable to grid-averaged forcings at large spatial scales.

Acknowledgements

E. Wood and D. Lettenmaier would like to recognize the tireless efforts of the third author, Dr. Xu Liang, in preparing the Phase 2(c) data and organizing the runs. 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 through NSF grant ERA-9318896 and NOAA (Office of Global Programs) grant NA56GP0249. The PILPS Phase 2(c) activities at University of Washington were supported through NSF grant ERA-9318898 and NOAA/OGP Grant NA67RJ0155. This support is gratefully acknowledged.

References

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

Austin, P., 1987. Relation between measured radar reflectivity and surface rainfall. Mon. Wea. Rev. 115, 1053–1070.

Baeck, M., Smith, J., 1995. Climatological analysis of manually digitized radar data for the United States east of the Rocky Mountains. Water Resour. Res. 31 (12), 3033–3049.

Chen, T., Henderson-Sellers, A., Milly, P., Pitman, A., Beljaars, A., Abramopoulos, F., Boone, A., Chang, S., Chen, F., Dai, Y., Desborough, C., Dickinson, R., Dümenil, L., Ek, M., Garratt, J., Gedney, N., Gusev, Y., Kim, J., Koster, R., Kowalczyk, E., Laval, K., Lean, J., Lettenmaier, D., Liang, X., Mahfouf, J., Mengelkamp, H.-T., Mitchell, K., Nasonova, O., Noilhan, J., Polcher, J., Robock, A., Rosenzweig, C., Schaake, J., Schlosser, C., Schulz, J.-P., Shao, Y., Shmakin, A., Verseghy, D., Wetzel, P., Wood, E., Xue, Y., Yang, Z.-L., Zeng, Q., 1997. Cabauw experimental results from the Project for Intercomparison of Land-surface Parameterization Schemes (PILPS). J. Clim. 10, 1194–1215.

Cosby, B., Hornberger, G., Clapp, R., Ginn, T., 1984. A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. Water Resour. Res. 20 (6), 682–690.

Doviak, R., Zrnic, D., 1993. Doppler Radar and Weather Observations, 2nd edn. Academic Press, New York.

Henderson-Sellers, A., Yang, Z.-L., Dickinson, R., 1993. The Project of Intercomparison of Land-surface Parameterization Schemes. Bull. Am. Meteorol. Soc. 74, 1335–1349.

Henderson-Sellers, A., Pitman, A., Love, P., Irannejad, P., Chen, T., 1995. The Project of Intercomparison of Land-surface Parameterization Schemes (PILPS): Phases 2 and 3. Bull. Am. Meteorol. Soc. 94, 489–503.

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

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., 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: 2. Spatial and temporal analysis of energy fluxes. Global Planet. Change.

Lohmann, D., Lettenmaier, D., Liang, X., Wood, E., 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., 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: 3. Spatial and temporal analysis of water fluxes. Global Planet. Change.

Meesen, B., Corprew, F.E., McManus, J.M.P., Myers, D.M., Closs, J.W., Sun, K.J., Sunday, J., Sellers, P.J., 1995. ISLSCP Initiative I—Global Data Sets for Land–Atmosphere Models, 1987–1988, Vols. 1–5. CD-ROM, NASA.

Peters-Lidard, C., Zion, M., Wood, E.F., 1997. A soil–vegetation–atmosphere transfer scheme for modeling spatially variable water and transfer scheme for modeling spatially variable water and energy balance processes. J. Geophys. Res. 102 (D2), 4303–4324.

Pitman, J., Henderson-Sellers, A., Yang, Z.-L., Abramopoulos, F., Avissar, R., Bonan, G., Boone, A., Cogley, J., Dickinson, R., Ek, M., Entekhabi, D., Famiglietti, J., Garrat, J., Frech, M., Hahmann, A., Koster, R., Kowalczyk, E., Laval, K., Lean, L., Lee, T., Lettenmaier, D., Liang, X., Mahfouf, J.-F., Mahrt, L., Milly, C., Mitchell, K., de Noblet, N., Noihan, J., Pan, H., Pielke, R., Robock, A., Rosenzweig, C., Running, S., Schlosser, A., Scott, R., Suarez, M., Thompson, S., Verseghy, P., Wetzel, P., Wood, E., Xue, Y., Zhang, L., 1993. Results from the off-line control simulation phase of the Project for Intercomparison of Land-surface Parameterization Schemes (PILPS). Technical Report 7. IGPO Publication Series.

Pitman, J., Henderson-Sellers, A., Yang, Z.-L., Abramopoulos, F., Boone, A., Desborough, C., Dickinson, R., Garrat, J., Gedney, N., Koster, R., Kowalczyk, E., Lettenmaier, D., Liang, X., Mahfouf, J.-F., Noihan, J., Polcher, J., Qu, W., Robock, A., Rosenzweig, C., Schlosser, A., Shmakin, A., Smith, J., Suarez, M., Verseghy, D., Wetzel, P., Wood, E., Xue, Y., 1997. Key results and implications from Phase 1(c) of the Project for Intercomparison of Land-surface Parameterization Schemes. Clim. Dyn. Submitted.

Rawls, W., Brakensiek, D., 1985. Prediction of Soil Water Properties for Hydraulic Modeling. ASCE.

Shao, Y., Henderson-Sellers, A., 1995. Validation of soil moisture simulation in land-surface parameterization schemes with HAPEX data. Global Planet. Change 13, 11–46.

STATSGO, 1994. State soil geographic (STATSGO) data base: data use information. Technical Report 1492. US Dept. Agric.

TVA, 1972. Heat and mass transfer between a water surface and the atmosphere. Water Resources Report No. 0-6803 14. Tennessee Valley Authority.