# Simulations of a Boreal Grassland Hydrology at Valdai, Russia: PILPS Phase 2(d)

C. ADAM SCHLOSSER,* ANDREW G. SLATER,+ ALAN ROBOCK,# ANDREW J. PITMAN,+ KONSTANTIN YA. VINNIKOV,@ ANN HENDERSON-SELLERS,& NINA A. SPERANSKAYA,** KEN MITCHELL,++ AND THE PILPS 2(D) CONTRIBUTORS##

* *Center for Ocean–Land–Atmosphere Studies, Calverton, Maryland*
\+ *School of Earth Sciences, Macquarie University, Sydney, New South Wales, Australia*
\# *Department of Environmental Sciences, Rutgers—The State University of New Jersey, New Brunswick, New Jersey*
@ *Department of Meteorology, University of Maryland, College Park, Maryland*
& *Royal Melbourne Institute of Technology, Melbourne, Victoria, Australia*
** *State Hydrological Institute, St. Petersburg, Russia*
++ *Environmental Modeling Centers, NOAA/NCEP, Camp Springs, Maryland*

## Aaron Boone, Météo-France/CNRM, Toulouse, France; Harald Braden, German Meteorological Service, Braunschweig, Germany; Fei Chen, Environmental Modeling Centers, NOAA/NCEP, Camp Springs, Maryland; Peter Cox, Hadley Centre for Climate Prediction and Research, Berkshire, United Kingdom; Patricia de Rosnay, Laboratoire de Météorologie Dynamique du CNRS, Paris, France; Carl E. Desborough, School of Earth Sciences, Macquarie University, Sydney, New South Wales, Australia; Robert E. Dickenson, Institute of Atmospheric Physics, The University of Arizona, Tucson, Arizona; Yong-Jiu Dai, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China; Qingyun Duan, Office of Hydrology, NOAA, Silver Spring, Maryland; Jared Entin, Department of Meteorology, University of Maryland, College Park, Maryland; Pierre Etchevers, Météo-France/CNRM, Toulouse, France; Nicola Gedney, Meteorology Department, Reading University, Reading, United Kingdom; Yeugeniy M. Gusev, Institute of Water Problems, Moscow, Russia; Florence Habets, Météo-France/CNRM, Toulouse, France; Jinwon Kim, Lawrence Berkeley National Laboratory, Livermore, California; Victor Koren, Office of Hydrology, NOAA, Silver Spring, Maryland; Eva Kowalczyk, Division of Atmospheric Research, CSIRO, Aspendale, Australia; Olga N. Nasonova, Institute of Water Problems, Moscow, Russia; Joel Noilhan, Météo-France/CNRM, Toulouse, France; John Schaake, Office of Hydrology, NOAA, Silver Spring, Maryland; Andrey B. Shmakin, Institute of Geography, Moscow, Russia; Tatiana G. Smirnova, Mesoscale Analysis and Prediction System/NOAA, Boulder, Colorado; Diana Verseghy, Climate Research Branch, Atmospheric Environment Service, Downsview, Ontario, Canada; Peter Wetzel, Mesoscale Dynamics and Precipitation Branch, NASA/GSFC, Greenbelt, Maryland; Yongkang Xue, Department of Geography, University of Maryland, College Park, Maryland; Zong-Liang Yang, Institute of Atmospheric Physics, The University of Arizona, Tucson, Arizona

(Manuscript received 26 August 1998, in final form 15 February 1999)

ABSTRACT

The Project for the Intercomparison of Land-Surface Parameterization Schemes (PILPS) aims to improve understanding and modeling of land surface processes. PILPS phase 2(d) uses a set of meteorological and hydrological data spanning 18 yr (1966–83) from a grassland catchment at the Valdai water-balance research site in Russia. A suite of stand-alone simulations is performed by 21 land surface schemes (LSSs) to explore the LSSs' sensitivity to downward longwave radiative forcing, timescales of simulated hydrologic variability, and biases resulting from single-year simulations that use recursive spinup. These simulations are the first in PILPS to investigate the performance of LSSs at a site with a well-defined seasonal snow cover and frozen soil. Considerable model scatter for the control simulations exists. However, nearly all the LSS scatter in simulated root-zone soil moisture is contained within the spatial variability observed inside the catchment. In addition, all models show a considerable sensitivity to longwave forcing for the simulation of the snowpack, which during the spring melt affects runoff, meltwater infiltration, and subsequent evapotranspiration. A greater sensitivity of the ablation, compared to the accumulation, of the winter snowpack to the choice of snow parameterization is found. Sensitivity simulations starting at prescribed conditions with no spinup demonstrate that the treatment of frozen soil (moisture) processes can affect the long-term variability of the models. The single-year recursive runs show large biases, compared to the corresponding year of the control run, that can persist through the entire year and underscore the importance of performing multiyear simulations.

## 1. Introduction

The principal goal of the Project for the Intercomparison of Land-Surface Parameterization Schemes (PILPS) has been to obtain a greater understanding of land surface processes and their parameterizations [see Henderson-Sellers et al. (1995) for the latest review of PILPS]. Since its inception, PILPS has designed an infrastructure of experimental efforts to achieve this goal. Phases 1 and 2 involve stand-alone simulations (decoupled from host atmospheric models) of land surface schemes (LSSs) driven by synthetic and observed at-

*Corresponding author address:* Dr. C. Adam Schlosser, COLA/IGES, 4041 Powder Mill Rd., Suite 302, Calverton, MD 20705.
E-mail: adam@cola.iges.org

mospheric forcing. Phases 3 and 4 investigate the LSSs' performance in fully coupled simulations with a host regional atmospheric model and general circulation models (GCMs).

In phase 1, synthetically forced grassland and tropical forest simulations of the participating LSSs were analyzed (Pitman et al. 1999). Further investigation (Koster and Milly 1997) illustrated that seasonal variations of liquid soil-water fluxes for a controlled LSS simulation could be essentially reproduced by fitting diagnostic parameters with linear evaporation and runoff functions in a simple monthly water-balance model. Overall, the results exemplified the wide range in the LSSs' simulations under identical forcing and a common vegetation type, and emphasized the role of the interactions between the evaporation and runoff formulations to the disparity of the LSSs.

Phase 2 of PILPS is primarily aimed at comparing and evaluating stand-alone simulations of LSSs against observed data. Phase 2(a) focused on the partitioning between sensible and latent heat fluxes for a grassland site at Cabauw, Netherlands (Chen et al. 1997), and the response of simulated heat fluxes to changes in air temperature (Qu et al. 1998). The analysis revealed that, for the Cabauw site, consideration of the stomatal control and soil-moisture stress formulations among the LSSs was crucial to the interpretation of the results. In phase 2(b) (Shao and Henderson-Sellers 1996), the diversity of the simulation and functionality of soil moisture among the LSSs using the HAPEX-MOBILHY data was demonstrated. Basin-scale simulations of the LSSs for the Red–Arkansas River area were verified against daily river flow observations and evaporation estimates from atmospheric budget analysis in phase 2(c) (Wood et al. 1998). The verification results illustrated that through proper calibration of the LSS, improved simulations of the large-scale quantities can be attained.

Most of the previous PILPS offline model comparisons have been limited to a 1-yr dataset for both the synthetic forcing (Pitman et al. 1999) and the observed forcing (Shao and Henderson-Sellers 1996; Chen et al. 1997), with the only exception being the recent PILPS 2(c) experiments (Wood et al. 1998). This prevents an assessment of the models' performance in the simulation of transient change over many years. In addition, none of the previous phase 2 experiments were able to provide validation of all the simulated soil-water balance components against measured observations, and the analyses excluded any thorough evaluation of simulated snow processes.

A recent comparison of GCMs has shown there to be a large disparity in both duration and extent of global snow cover (Foster et al. 1996). In addition, snow covers about 50% of the land surface of the Northern Hemisphere during winter (Robinson et al. 1993) and frozen soil affects about 20% of the global land surface (Williams and Smith 1989). As such, these processes are of significant importance to LSSs and their coupling with GCMs to simulate the global climate.

Recently, a 31-yr hydrological dataset was obtained from the Valdai water-balance research station in Russia (Vinnikov et al. 1996), located in a boreal forest region. Schlosser et al. (1997) used an 18-yr meteorological dataset from Valdai as forcing and the hydrological data in a pilot study to test their suitability for stand-alone simulations with two LSSs: a simple bucket hydrology model and a more complex biosphere model. The results show that the implementation of the Valdai data was successful and that the model simulations could be used as a useful evaluation tool. Moreover, some of the deficiencies noted in the models' simulations were directly or indirectly related to their snow simulations. This underscores the importance of using the Valdai data for a PILPS exercise in that we can examine the performance of many participating models for a site characterized by a deep seasonal snow cover and frozen soil, which, to date, has never been considered in PILPS phase 2.

For this study, PILPS phase 2(d), we use the observed meteorological and hydrological data spanning 18 yr (1966–83) for the grassland catchment, Usadievskiy, at Valdai. For 21 participating LSSs of varying complexity, a suite of simulations are performed: a control simulation, and five additional simulations designed to address the sensitivity of the LSSs to downward longwave radiative forcing, and the timescales and controls of simulated hydrologic variability. The main goal of this paper is to summarize the models' results and sensitivities. Discussion regarding model disparity and implications of the sensitivities toward general modeling issues are also provided. However, the analyses should be viewed in the context of model validation and development using data from northern climate regions (e.g., BOREAS (Sellers et al. 1997), GSWP (Dirmeyer et al. 1999), and additional Valdai catchments (Vinnikov et al. 1996). Section 2 provides a description of the data used. Section 3 outlines the rationale and methodology of the PILPS 2(d) experiments. The analysis of the model results is given in section 4. A summary and implications (where applicable) of the results are given in section 5, and closing remarks are made in section 6.

## 2. Data

Fedorov (1977), Vinnikov et al. (1996), and Schlosser et al. (1997) describe the continuous 18 yr (1966–83) of atmospheric forcing and hydrologic data in detail. An overview of the data is given below, and further details concerning the data can be found in the references cited. The data were obtained from the Valdai water-balance station (57°58′N, 33°14′E), located in a boreal forest region. The Usadievskiy catchment, an experimental catchment at Valdai where long-term hydrological measurements were taken, is about 0.36 km$^2$ in areal extent and is covered with a grassland meadow. The atmospheric data were measured at a grassland plot near the Usadievskiy catchment with temperature, pres-

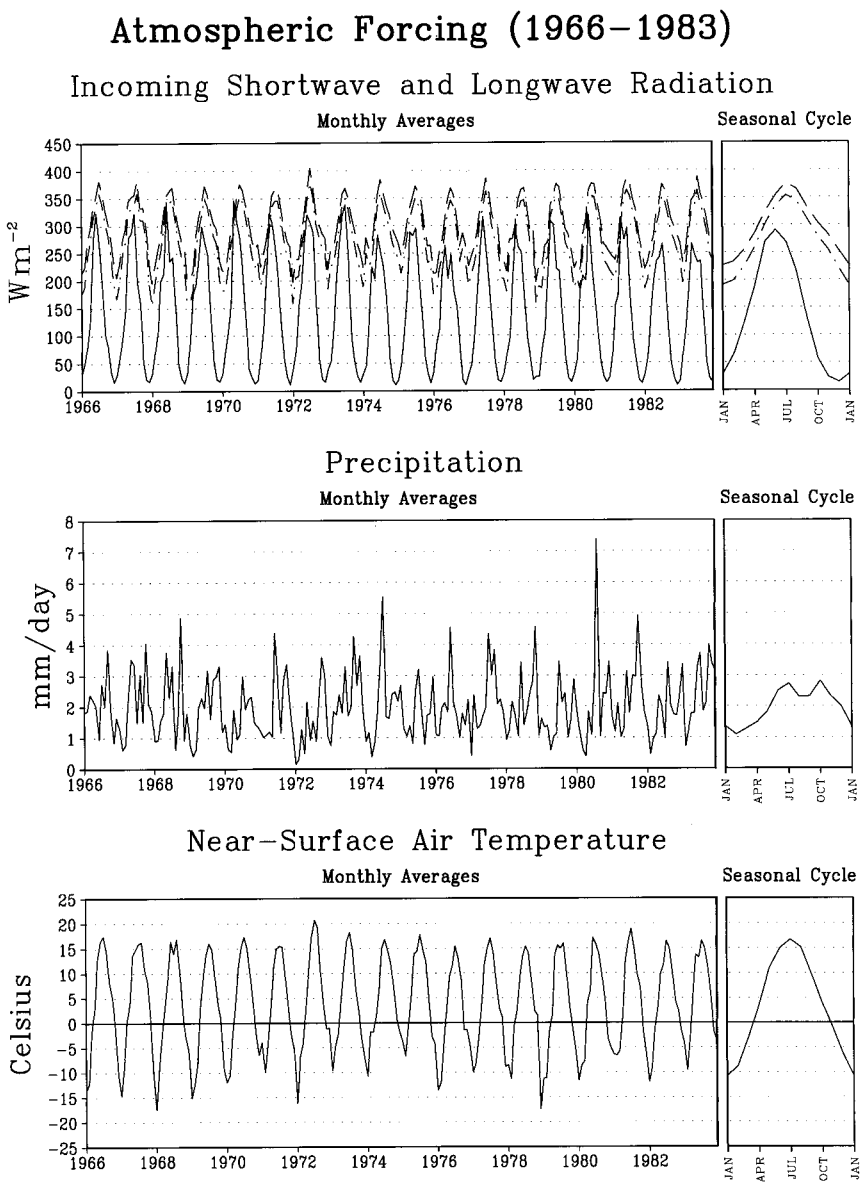


FIG. 1. Monthly averages and seasonal cycles (based on the monthly averages) of solar radiation based on the M. E. Berlyand (1956) and T. G. Berlyand (1961) algorithm: downward longwave based on the Idso (1981) and Brutsaert (1975) algorithms (shown as dashed and dashed–dotted lines, respectively), precipitation, and near-surface air temperature used as forcing for the years 1966–83.

sure, and humidity recorded at a height of 2 m and wind speed logged at 10 m.

The atmospheric data were originally sampled at 3-h intervals. However, in order for all of the participating models to utilize the meteorological data as forcing, the data were interpolated to 30 min (and in some cases 5 min) intervals using a cubic-spline interpolation procedure similar to that used by Chen et al. (1997). The notable difference of the interpolation procedure used for the Valdai forcing was that precipitation was interpolated linearly in the fall and winter [rather than the cubic spline applied year-round as in Chen et al. (1997)]. This was done to better represent the synoptic nature of precipitation during that period of the year. An illustration of the precipitation and air temperature data is given in Fig. 1, which gives the monthly averages and seasonal cycles. Overall, the climate at Valdai is highly seasonal with an annual temperature range of 35°C and an annual average precipitation of 730 mm with the majority falling in the summer and autumn months. Temperatures fall below −10°C in winter and snow typically covers the surface between November and April.

The atmospheric dataset did not contain all the required forcing variables for the LSSs. Shortwave radiation was simulated (Fig. 1) following the M. E. Berlyand (1956) and T. G. Berlyand (1961) algorithm. This algorithm is based on an extensive collection of Russian observations and was successfully tested for use as forcing with two LSSs (Schlosser et al. 1997). In addition, Schlosser (1995) repeated the two LSS simulations of Robock et al. (1995) for six Russian sites, except the downward shortwave forcing (originally obtained from observations) was replaced with the Berlyand algorithm. The details of the results are beyond the scope of this paper. However, the findings revealed that the use of the shortwave algorithm resulted in very small biases of the simulated hydrology (less than 5% in the winter and on average 10% in the warmer months) and that the simulated monthly variations of the hydrologic budget were essentially reproduced. Observed downward longwave radiation (LWD) was also not available, and following previous studies (Yang et al. 1997; Slater et al. 1998a) an empirical algorithm (Idso 1981) was chosen as an effective simulation of the LWD forcing (Fig. 1). An additional LWD algorithm by Brutsaert (1975) was also used (Fig. 1) to test the sensitivity of the Valdai LSS simulations to LWD (the details of which are described in section 3c).

The observed hydrological data for model evaluation consisted of total soil moisture in the top 1 m, snow water equivalent depth (SWE), evaporation, runoff, and water-table depth. Runoff was measured each month by a stream gauge at the catchment outflow site. To assure a more consistent comparison of the observed catchment runoff to modeled runoff from the root-active zone (one of the requested outputs from the simulations), the observed catchment runoff was modified (Schlosser et al. 1997) according to monthly variations in the observed catchment averaged water-table depth (measured at numerous wells within the catchment). At the end of every month, total soil moisture in the top 1 m of soil was sampled, using a thermostat weight technique (Robock et al. 1995) at 11 point measurement sites within the catchment. As shown by Robock et al. (1995), the approximate error in measurement is about $\pm 1$ cm. Monthly evaporation observations from lysimeter measurements during the summer (and estimates for the winter months) for the years 1966–73 were taken from Fedorov (1977). In addition, monthly evaporation estimates were calculated as a residual of the monthly water balance in the top 1 m of soil. Schlosser et al. (1997) illustrates that while the residual estimates of evaporation are prone to significant errors in the winter and disagreement in the monthly trends against the lysimeter measurements exist, their seasonal cycles (averaged over the period for which the lysimeter measurements are available) are in good agreement. Snow measurements, at 44 sites within the catchment, were made at least every month during the winter, and more frequently (at intervals of days) during the spring snow melt. Illustrations of all the hydrological data taken within the Usadievskiy catchment (including a detailed map of the catchment) are given by Vinnikov et al. (1996) and Schlosser et al. (1997).

For the observations of total soil moisture and water equivalent snow depth, the average of all the site measurements made within the catchment (i.e., the catchment average) is used for comparison with the models. However, the scatter of the values and trends among the observation sites within the catchment, or the degree of spatial variability, must also be considered. All the model parameters (described in the next section) are prescribed to represent the catchment as spatially homogeneous. In reality, subcatchment variations of spatial features such as topography, soil properties, and vegetation characteristics affect the spatial and temporal variability of the measured hydrologic quantities. Therefore, any LSS result that falls outside the range of error in the catchment average (mentioned in the preceding paragraph) but lies within the observed scatter of the measurement sites should not necessarily be considered erroneous, since the models do not incorporate the factors that caused the subcatchment spatial variability in the observations. Nonetheless, given that the LSSs are designed to simulate an ''averaged'' hydrology over a given area, the most desirable simulation for these schemes is to lie close to the catchment average (i.e., near the center of the observed spatial scatter). For our analysis, the 11 measurement sites for total soil moisture were digitized and available, but only the catchment averaged values of water equivalent snow depth were able to be obtained.

Overall, the atmospheric data and hydrologic record provide a coherent set of forcing and validation data for stand-alone LSS simulations. In addition, the predominant grassland vegetation within the catchment and at the neighboring meteorological observation site requires grassland vegetation parameters to be prescribed to the LSSs for their simulations. In the next section, a description of the suite of model simulations is given, and the prescribed parameters as well as certain parameterization constraints and freedoms for the simulations are detailed.

## 3. Experiment design

### *a. Parameters and model specifications*

Model parameters were specified in an attempt to eliminate, as much as possible, their influence on the intermodel differences and the resulting analysis. For the participating models (Table 1), a set of consistent parameters, based whenever possible on observed properties of the Usadievskiy catchment, were supplied for the simulations. In the event that no direct observation of a particular parameter was available, a value was set based on the overall vegetation and soil properties observed at Usadievskiy (Schlosser et al. 1997). Those

TABLE 1. List of the models, the primary contacts and contributors, and the corresponding institutions that have participated in PILPS phase 2(d). The numbers that are assigned to the models are used as references to the model results shown in Fig. 4. Shown also for each model is the history of their PILPS participation in previous phase 1 and 2 experiments.

| Model | | PILPS participation | Contact(s) | Institution | Reference |
|---|---|---|---|---|---|
| 1 | AMBETI | 2d | H. Braden | German Meteorological Service, Germany | Braden (1995) |
| 2 | BASE | 1, 2a–d | A. Slater, C. Desborough, A. Pitman | Macquarie University, Australia | Desborough and Pitman (1998) |
| 3 | BATS | 1, 2a–d | Z. L. Yang, R. E. Dickinson | University of Arizona, USA | Yang et al. (1997) |
| 4 | BUCK | 1, 2a, 2c, 2d | C. A. Schlosser | COLA/IGES, USA | Robock et al. (1995) |
| 5 | CLASS | 1, 2a–d | D. Verseghy | Atmos. Environment Service, Canada | Verseghy et al. (1991) |
| 6 | CROCUS | 2d | P. Etchevers | CNRM, France | Brun et al. (1992) |
| 7 | CSIRO | 1, 2a, 2b, 2d | E. Kowalczyk | CSIRO, Australia | Kowalczyk (1999, manuscript submitted to *J. Climate*) |
| 8 | IAP94 | 2a, 2c, 2d | D. YongJiu | Inst. of Atmos. Physics, China | Dai and Zeng (1997) |
| 9 | ISBA | 1, 2a–d | F. Habets, J. Noilhan | CNRM, France | Noilhan and Mahfouf (1996) |
| 10 | MAPS | 2d | T. Smirnova | Mesoscale Anal. & Pred. System/NOAA, USA | Smirnova et al. (1997) |
| 11 | MOSES | 2d | P. Cox | Hadley Centre, UK | Cox et al. (1999) |
| 12 | NCEP | 2a, 2c, 2d | K. Mitchell, Q. Duan | NCEP, USA | Chen et al. (1996) |
| 13 | PLACE | 1, 2a–d | A. Boone, P. Wetzel | NASA/GSFC, USA | Wetzel and Boone (1995) |
| 14 | SECHIBA | 1, 2a–d | P. de Rosnay, J. Polcher | LMD, France | de Rosnay and Polcher (1998) |
| 15 | SLAM | 2d | C. Desborough | Macquarie University, Australia | Desborough (1998) |
| 16 | SPS | 2a, 2d | J. Kim | LBNL, USA | Kim and Ek (1995) |
| 17 | SPONSOR | 1, 2a, 2c, 2d | A. B. Shmakin | Institute of Geography, Russia | Shmakin (1998) |
| 18 | SSiB | 1, 2a–d | Y. Xue, C. A. Schlosser | University of Maryland, USA & COLA/IGES, USA | Xue et al. (1996) |
| 19 | SWAP | 2a, 2c, 2d | Y. M. Gusev, O. N. Nasonova | Institute of Water Problems, Russia | Gusev and Nasonova (1998) |
| 20 | UGAMP | 1, 2a, 2d | N. Gedney | Reading University, UK | Gedney (1995) |
| 21 | UKMO | 1, 2a, 2d | P. Cox | Hadley Centre, UK | Warrilow and Buckley (1989) |

parameter assignments that were identical for all the participating models are given in Table 2. However, as with previous PILPS experiments, the problem of different parameters having a different meaning and functionality across the range of LSSs persists (Polcher et al. 1996). Therefore, a parameter such as this having an identical value between LSSs does not presuppose model agreement. This has been an ongoing problem for LSS comparisons (Chen et al. 1996). Nonetheless, for these model-specific parameters, efforts were made to provide values that not only are consistent with the conditions at Valdai, but also would minimize the model disparity that results.

In addition, consistent partitioning of liquid and frozen precipitation forcing used by the models was assured. In order that all the models would receive the identical amount of liquid and frozen precipitation as forcing, each model was instructed to assume that precipitation fell as snow when the observed air temperature (of the forcing data) was less than or equal to 0°C. The rain–snow temperature threshold used for these simulations is quite consistent to both the observational and computational evidence of Yang et al. (1997).

Due to the presence of a seasonally varying water table at Valdai (Vinnikov et al. 1996), the hydrological soil column was set to a depth of 2 m for all the models. This minimizes the biases that would result between the models, which may or may not account for water-table variations. A climatology of the observed catchment averaged water-table depth (Schlosser et al. 1997) was also provided for those models that required implicit or explicit water-table information.

One important aspect of the simulations is the relative degree of freedom given to the models in their parameterizations of soil thermal and snow processes. Given that the Valdai simulations are the first in PILPS to consider a site with a well-defined seasonal snow cover and frozen soil, we should first explore the sensitivities that result from a variety of model parameterizations that consider these processes. While we acknowledge that any freedom given to the LSSs for the simulations will likely lead to different results, the parameterizations and assigned parameter values for each LSS are chosen to achieve an optimal result. Therefore, this can be regarded as a fair test between the range of parameterizations. Moreover, the insights gained from these sen-

TABLE 2. List of parameter values that were assigned to each of the models for the phase 2(d) simulations. The asterisks denote those parameters whose values were based on direct observations.

| Parameter | Value |
|---|---|
| Soil properties | |
| Soil type composition* | Loam 56%, sandy loam 28%, sandy 16% |
| Wilting level in top 1 m* | 115 mm |
| Field capacity in top 1 m* | 271 mm |
| Total water holding capacity in top 1 m* | 401 mm |
| Hydraulic conductivity at saturation | $2.00 \times 10^{-5}$ ms$^{-1}$ |
| Bare soil roughness length | 0.01 m |
| Volumetric dry-soil heat capacity | $2.00 \times 10^{6}$ J m$^{-3}$ K |
| Clapp and Hornberger ''B'' parameter | 7.12 |
| Vegetation properties | |
| Stem area index (SAI) | 0.8 |
| Maximum value of leaf area index of green | 4.2 |
| Minimum value of leaf area index of green | 0.4 |
| Maximum stomatal resistance | 20,000 s m$^{-1}$ |
| Minimum stomatal resistance | 60 s m$^{-1}$ |
| Canopy roughness length | 0.035 m |
| Interception capacity | 0.0001 (LAIG + SAI) |
| Roughness length | 0.035 m |
| Zero-plane displacement height | 0.25 m |
| Canopy heat capacity | 2,000 J m$^{-3}$ K |
| Rooting depth | 1 m |
| Rooting fraction | 70% in top 0.1 m; 30% in the remaining top 1 m |
| Maximum fractional cover of vegetation | 0.90 |
| Other properties | |
| Thermal emissivity of all surfaces | 1.0 |
| Snow roughness length | 0.0024 m |
| Maximum snow albedo* | 0.75 (0.65 for near IR, 0.85 for visible) |
| Snow-free ground and vegetation albedo* | 0.23 |
| von Kármán's constant | 0.378 |

sitivities can serve as a guide to the direction and importance of the development of these parameterizations. As a result of these freedoms in model complexity, some models have up to 11 soil thermal layers that can extend to a depth of 6 m (Fig. 2), and a wide range of snow schemes are represented ranging from very simple parameterizations to multiple snow-layer schemes that consider snow aging and/or spatial heterogeneity of the cover. For the snow parameters, only snow albedo was given a maximum (i.e., fresh snow) value. The models were asked to use their own snow albedo schemes for the purposes of changing albedo as the snow aged and/or melted. In addition, no specification was made for the fractional extent of snow cover nor the density of snow.

### *b. Control run*

For the control run, the participating LSSs were run for the entire 18 yr of forcing (1966–83). Initialization was not prescribed. Instead, all the models ran the first year (i.e., 1966) of forcing repeatedly until achieving equilibrium, defined as the point at which all the January fluxes between 1 yr of the spinup and its precedent year are within 0.01 W m$^{-2}$ and all temperatures were within 0.01 K. This is essentially the same method employed in previous PILPS experiments (e.g., Chen et al. 1997), but with stricter tolerances on the heat fluxes by an order of magnitude. The tighter controls to achieve equilibrium were assigned as a result of preliminary spinup tests conducted for one of the participating models (the BASE model). The results indicated that simulated frozen soil-water conditions inflict a large thermal inertia into the soil column, and thus small temperature differences in frozen soil (moisture) can subsequently lead to large ground and surface heat flux differences. The tested model had a soil thermal discretization that was intermediate among all the participating models and explicitly accounted for frozen soil moisture processes (as did over half of the participating models—see Fig. 2). As a result, the stricter heat flux tolerance was employed to ensure proper spinup for all the LSSs, especially for those that considered frozen soil moisture processes and had a more complex soil thermal discretization.

Once the model equilibrated, it ran through the remaining 17 yr of forcing (1967–83). For the equilibrated spinup year (1966) and the remainder of the simulation period (1967–83), the models were requested to report daily averaged values of 24 output variables (Table 3). The identical set of output variables was also requested for the sensitivity runs, which are summarized in Table 4 and described in the next section.

### *c. Sensitivity runs*

Additional experiments were carried out in which models ran a single year (i.e., 1 January–31 December)

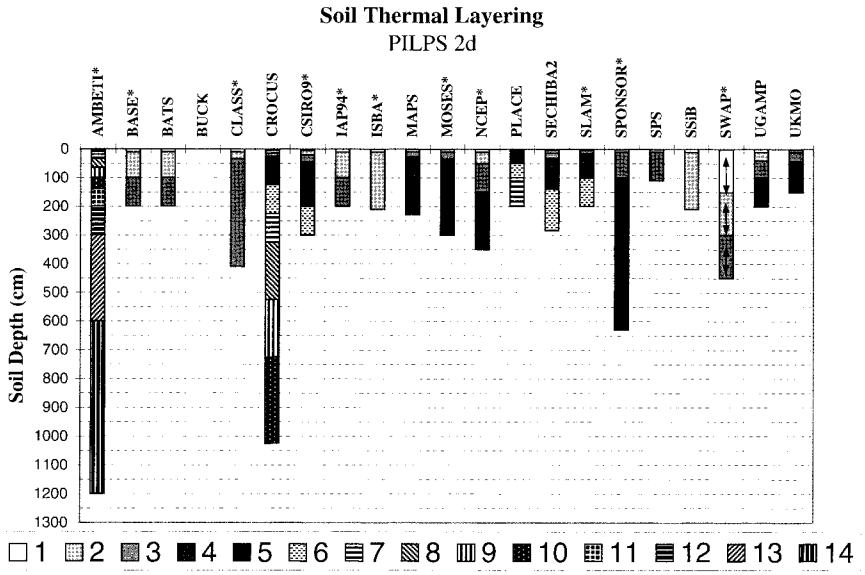


FIG. 2. Discretization of the thermal soil layers for all the participating models. The different shadings refer to each successively deeper and discrete layer of soil the model considers. A zero thickness for layer 1 indicates that the model calculates a surface skin temperature. The arrows drawn in SWAP's three soil layers indicate that the thickness of its soil layers vary according to the depth of soil freezing in the winter and the progression of the melting front in the spring. In addition, model acronyms followed by a ''*'' indicate that the LSS contains an explicit frozen soil moisture storage term and considers its effects on the soil heat budget and hydraulics.

of the forcing recursively until equilibrium. This is identical to what was done for the spinup year, 1966, in the control run. These recursive year runs were performed for the years 1972 (a year with a dry summer), 1974 (a year with a wet summer), and 1983 (the last year of the forcing and a typical hydrologic year). We refer to these test simulations as RY72, RY74, and RY83, respectively. The purpose of these experiments is to illustrate the potential effects of recursive spinup procedures used for previous (and future) stand-alone simulations with only 1 yr of forcing data. This will be achieved by comparing the RY72, RY74, and RY83 runs to the corresponding years of the control runs. Given the range of hydrologic extremes that occur within the three years tested, these simulations can also determine whether hydrologic events influence the biases that result from running only 1 yr of forcing.

TABLE 3. List of daily output variables (and their units) that were requested for all of the simulations of PILPS phase 2(d).

| |
|---|
| Precipitation (mm day$^{-1}$) |
| Total evapotranspiration (mm day$^{-1}$) |
| Transpiration (mm day$^{-1}$) |
| Bare-soil evaporation (mm day$^{-1}$) |
| Canopy evaporation (mm day$^{-1}$) |
| Evaporation from snow surface (mm day$^{-1}$) |
| Drainage of soil water from root-active zone (mm day$^{-1}$) |
| Surface runoff (mm day$^{-1}$) |
| Lateral flow from root-active zone (mm day$^{-1}$) |
| Snow melt (mm day$^{-1}$) |
| Total canopy-water storage (mm) |
| Total root-zone soil moisture (mm) |
| Water equivalent snow depth (mm) |
| Snow depth (mm) |
| Radiative temperature (K) |
| Canopy temperature (K) |
| Depth-averaged temperature of the top 0.1 m of soil (K) |
| Depth-averaged temperature of the root-active zone (K) |
| Soil temperature of layer closest to 2-m depth (K) |
| Absorbed solar radiation (W m$^{-2}$) |
| Net radiation (W m$^{-2}$) |
| Latent heat flux (W m$^{-2}$) |
| Sensible heat flux (W m$^{-2}$) |
| Surface albedo |

As previously mentioned, the LWD forcing for the control run was estimated using the Idso (1981) algorithm. In previous studies that used high-latitude, Russian data, namely, Yang et al. (1997), Slater et al. (1998a), and Schlosser et al. (1997) the sensitivity of their LSS's winter simulations to LWD was examined. The specific findings of these studies are beyond the scope of this paper. However, their testing underscores the importance of assessing the LWD sensitivity for high-latitude simulations of LSSs. Among the aforementioned studies, four LWD algorithms were considered: the Idso (1981), Satterlund (1979), Monteith (1973), and Brutsaert (1975) schemes. In light of this, the sensitivity of LWD was examined for the participating PILPS 2(d) LSSs by repeating the control run but using the Brutsaert (1975) algorithm for LWD forcing (hereafter referred to as the LNGW run). The Brutsaert (1975) algorithm was selected because, paired with the Idso (1981) scheme, the two algorithms spanned the

TABLE 4. List of sensitivity experiments run for PILPS 2(d). The left column provides the acronym of each sensitivity run used in the discussion of the results. The right column provides a brief explanation of each experimental run.

| Run | Description |
|---|---|
| NOSU (no spinup run) | As in the control run, except that no recursive spinup is performed for the first year (1966) of the simulation. The initial values of all the prognostic variables of the models are prescribed at intermediate values. |
| LNGW (longwave sensitivity run) | As in the control run, except the longwave algorithm of Monteith (1975) is used to provide the models with downward longwave forcing. |
| RY72, RY74, and RY83 (the recursive year runs) | Using the forcing data and parameter assignments of the control run, the years 1972, 1974, and 1983 (respectively) are run individually. The 1 yr of forcing is used recursively until equilibrium. |

range of disparity of estimated LWD obtained from each of the aforementioned algorithms for Valdai during the winter. On average, the Brutsaert (1975) scheme produces values of LWD that are approximately 20% lower in the winter and 10% lower in the summer than the Idso (1981) scheme (Fig. 1). The range of difference between the Idso and Brutsaert schemes is also quite comparable to the range in the uncertainty of observed LWD, which is about $\pm 10\%$ during the winter and as low as $\pm 5\%$ in the summer (F. Miskolczi 1998, personal communication; Miskolczi and Guzzi 1993). As such, the LNGW run together with the control run can also serve as an envelope of the model sensitivities to random errors in instrumentation when using observed longwave radiation as forcing.

An additional run in which models ran from a precise specification of initial conditions with no recursive first-year spinup was also conducted (hereafter referred to as the NOSU run). The prescribed initial conditions were: the total of liquid and frozen soil-water storages set to half capacity, all snow and canopy storages set to zero, and all prognostic temperatures equal to the air temperature of the first time step of the forcing. The simulation is constructed on the premise that we have no information for initialization, and we must use arbitrary initial conditions. In phase 1(a) of PILPS, Yang et al. (1995) explored the spinup processes of the participating LSSs and found a large range of spinup timescales, from a year up to 31 yr, that primarily depends on total soil moisture holding capacity and the initial value of the soil moisture stores. However, these tests were not conducted for conditions of deep snow cover and frozen soil over a range of parameterizations that account for these processes. As mentioned in section 3a and illustrated in Table 2, we fix the porosity and the total hydrological soil column among all the models. As such, the total water holding capacity and all the initial liquid and frozen moisture storages are equal among all models for the NOSU run. Therefore, we can use the NOSU run to determine if the presence of snow and frozen soil and the variety in the treatment of these processes within LSSs affect spinup.

## 4. Analysis

In this analysis, we identify general aspects of the model simulations in the context of overall model development and improvement by comparing the control runs of the models to each other, the observations, and the sensitivity runs. Given that the observations at Valdai focus on the hydrology of the Usadievskiy catchment, our analysis will focus on the hydrological outputs of the model simulations that were requested (Table 3).

### *a. Annual results*

#### 1) CONTROL RUN

The models' control simulation of annual averaged root-zone total soil moisture falls within the total range of the catchment observations (Fig. 3a), with the slight exception of one model in 1972. The range of model scatter for annual root-zone soil moisture is found to be approximately 100 mm (Fig. 3a). While the model scatter for the annual anomalies decreases to approximately 25 mm on average, most of the models are unable to completely capture the observed variability of the anomalies (Fig. 3c). This is most evident in 1975, when 17 of the models fall outside the observed range of dry anomalies (all of the models have a wet bias). In addition, some models show unique anomaly variations for isolated years within the simulation.

Averaged over the entire simulation period (1966–83), most of the models are in qualitative agreement with the observations in the partitioning of water fluxes (Fig. 4, with the observed evaporation obtained by the residual calculation). With the exception of one model, simulated evaporation is greater than total runoff, although there exists a scatter of 0.5 mm $day^{-1}$ among the models in the partitioning for the control run. The observed partitioning of the water fluxes lies very close to the middle of the model scatter. The models show a lower degree of interannual variability for both evapotranspiration and total runoff than the observations suggest. The interannual variability in both modeled and observed water fluxes are also bounded by the interannual variability of annual precipitation.

Not surprisingly, the models show a large scatter in their simulations of annual averaged SWE and its variability (Figs. 5 and 6). Compared with observations, there appears to be no overall tendency for the models to overestimate (as seen for 11 of the models) or underestimate (as seen for 10 of the models) SWE for the control runs (Fig. 5). The model scatter seen in Fig. 5

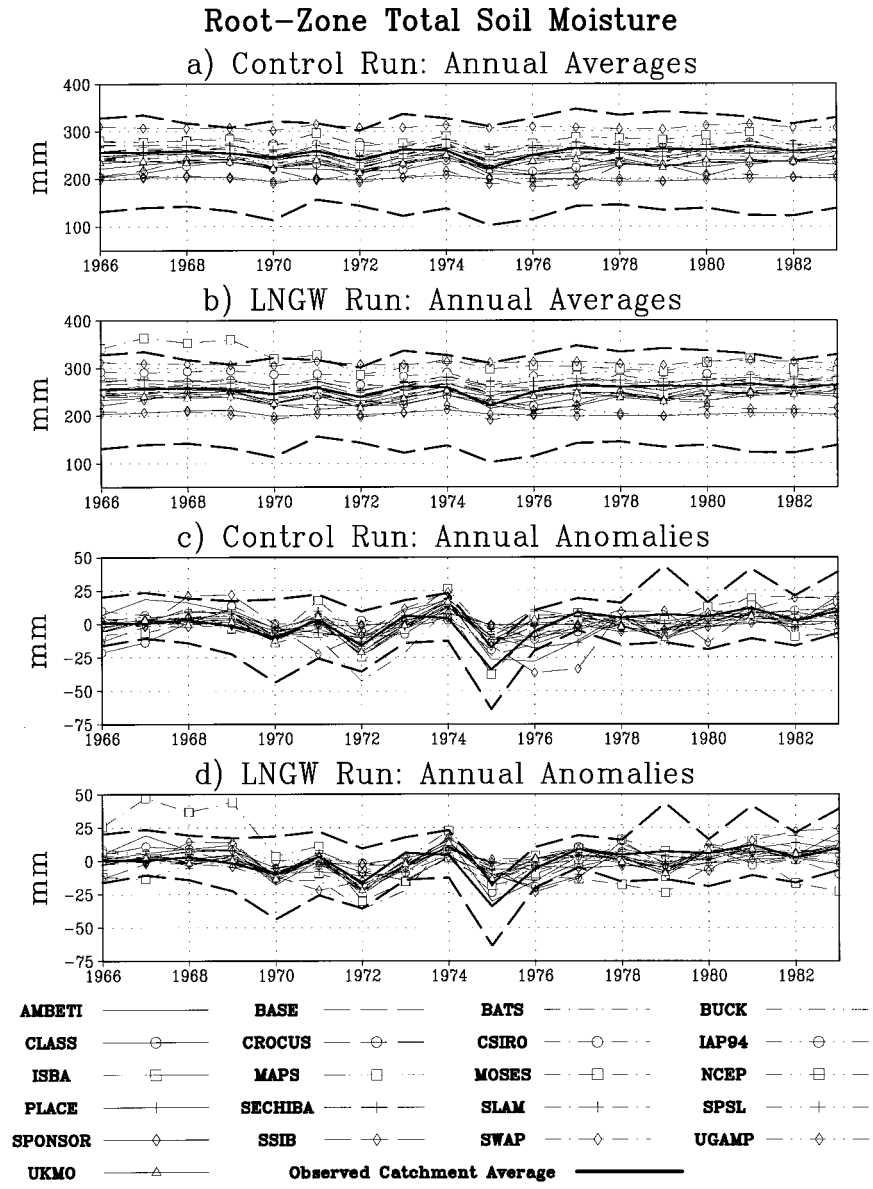


FIG. 3. (a) Control run results of simulated and observed (catchment averaged) annual averages of root-zone total soil moisture for the years 1966–83. Shown also are the highest and lowest observed values among the 11 observation sites within the catchment (thick dashed lines); (b) as in (a) but for the LNGW run; (c) as in (a) but for the annual averaged anomalies of root-zone soil moisture; (d) as in (c) but for the LNGW run.

is somewhat affected by the small sample size used for averaging. To consistently compare to observations, the daily modeled values were sampled only for the days when observations were made, about 12 times per winter. When sampled for all days during the modeled snowpack (about 168 times for each winter), the intermodel scatter of annual averaged SWE and the interannual variability SWE decreases considerably (Fig. 6).

Another important aspect of the modeled snow processes is the timing of the end of the spring snow melt. In terms of the hydrologic cycle, the frozen water stored in the snowpack has been melted and partitioned between runoff and soil-water recharge (the latter then available for evapotranspiration from growing plants and bare soil). In terms of the surface energy budget, the effects of higher albedo, soil insulation, and near-surface atmospheric cooling by the snowpack has been eliminated. In this analysis, we define the end of the snowmelt as the day in which the well-defined winter snowpack has been completely ablated. A 40-day scatter exists among the models in the average timing of the total ablation of the winter snowpack for the control

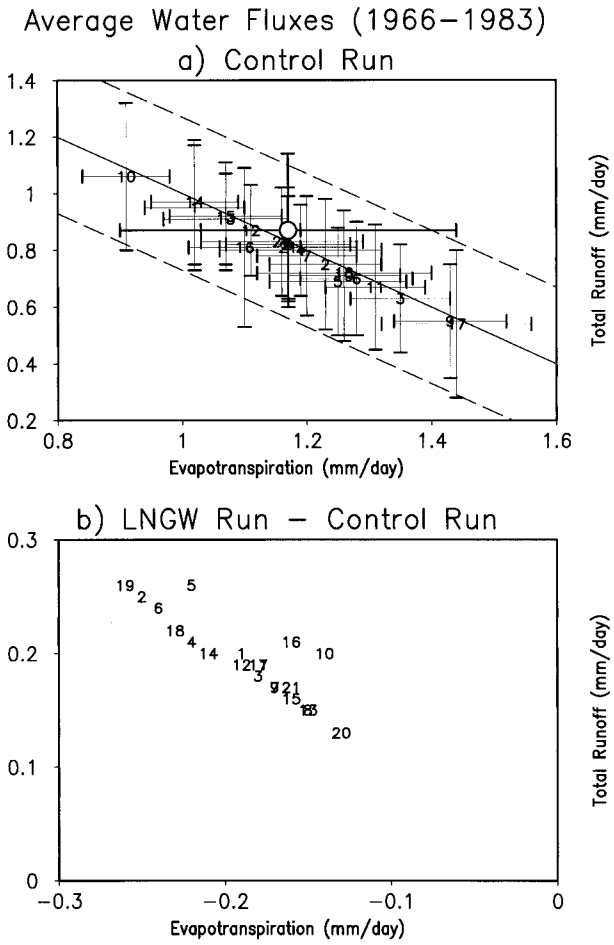


FIG. 4. (a) Total runoff from the root-active zone vs evapotranspiration averaged for the simulation period (1966–83). Model results are indicated by numbers that correspond to the list of models given in Table 1. Shown also is the observation point indicated by an open circle. The standard deviations of the annual averages for both the models and observations are indicated by crossbars. The solid straight line represents the range of possible partitioning points of total runoff and evaporation as given by the precipitation forcing (assuming no trends in soil and canopy soil-water storages and snow cover). The dashed lines have a distance in the $x$ and $y$ directions to the solid line of one standard deviation of annual average precipitation. (b) Difference between the control and LNGW simulations of total runoff from the root-active zone vs evapotranspiration averaged for the simulation period (1966–83).

runs (Fig. 7), with the earliest completion in late March and the latest in early May. The models also differ in their interannual variability. Some models vary the completion of snow melt by as little as 10 days from year to year, while for others the total range is as large as a month.

#### 2) LNGW RUN

While the LNGW runs produce slight increases in the simulation averages of the models' soil-water storages, the interannual variability of the collection of models is, for the most part, unchanged with respect to the control run (Fig. 3). The only notable exception is one model, which produces what appears to be a long-term decreasing trend. Further analysis indicates that for the LNGW run this particular model's deep soil moisture layer underwent a gradual thaw over the first 9 yr of the simulation, causing a slow drainage of soil water from its initial state. The model scatter of simulated water fluxes (Fig. 4b) shifts toward higher rates of runoff (by about 0.2 mm $d^{-1}$ on average) consistent with the lower amounts of incoming energy for evapotranspiration and more saturated soil moisture stores (Fig. 3). As a result, the observations are on the higher evaporation/lower runoff end of the LNGW model scatter.

The models' SWE simulation is quite sensitive to the choice of LWD forcing, affecting both the average value for the simulation and the interannual variability. For all the models, averaged SWE increases (Figs. 5 and 6) in response to the decreased LWD (Fig. 1), but the increases among the models range from 20 to 75 mm. In addition, all models complete the melting of the winter snowpack (Fig. 7) about 10–20 days later in the spring (compared to the control run), which is likely a combined response of the deeper snowpack and less incoming energy available to melt the snow. For most of the models, the absolute value of the change in average SWE seen in the LNGW runs could span the bias seen in the control run results. One additional aspect that needs to be included when comparing these results to observations is that snow was prescribed to fall when the air temperature was 0°C or less, but it is known for snow to fall at warmer air temperatures (Auer 1974; Yang et al. 1997). Yet a higher (i.e., above freezing) snow–rain temperature threshold for the control and LNGW runs could then cause extremely high SWE biases for some of the models. However, the question as to whether snowfall then accumulates or merely melts upon contact with the surface, and whether a model explicitly accounts for the process, complicates this issue. In addition, precipitation considered as snowfall during the spring melt, rather than rainfall, with a higher temperature threshold would contribute to snow accumulation rather than ablation and potentially cause a delay in the completion of the snow melt. These considerations regarding the SWE simulations and sensitivities will be explored in a future paper (Slater et al. 1999, manuscript submitted to *Climate Dyn.*).

### *b. Seasonal cycles*

#### 1) CONTROL RUN

Qualitatively speaking, nearly all of the models are successful at capturing the general features of the observed seasonality of root-zone soil moisture (Figs. 8a,b). Root-zone total soil moisture is depleted in the summer and recharged in the fall and spring. During the spring snow melt, catchment averaged root-zone to-

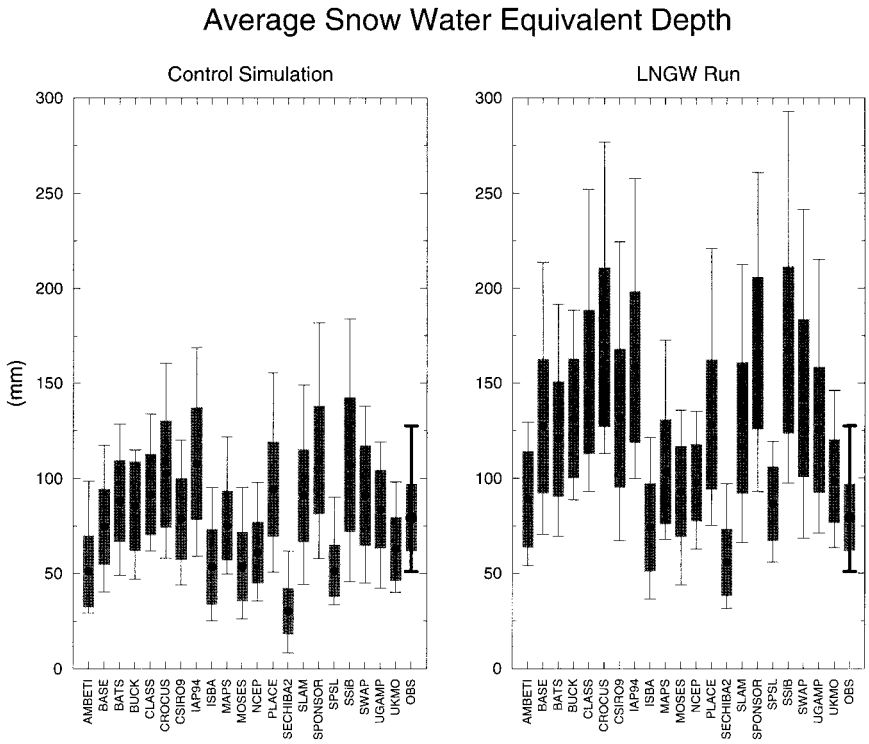


FIG. 5. Model simulations from the control and LNGW runs and observations of averaged water equivalent snow depth (indicated by the solid circle) for the simulation period (1966–83). The model averages are based on daily values that are sampled only for the day in which observations were taken. The standard deviations of the averages for the winter, defined as each period of the well-defined snowpack, are also indicated by the shaded region. The highest and lowest average snow depth among all the winters are indicated by the crossbars.

tal soil moisture is observed to rise above field capacity (=271 mm). Although some models do not simulate the spring soil-water recharge above field capacity, they are still within the observed spatial variability. In addition, most models are able to reasonably simulate the magnitude of the seasonal cycle of root-zone total soil moisture. A few outliers to these seasonal features exist even in light of the sensitivity test results (described in the next section) and the observed spatial variability of the catchment. The range of model scatter is approximately 100 mm for monthly root-zone soil moisture and 50 mm for the monthly anomalies, with larger scatter during the spring. Evapotranspiration (Fig. 8c) peaks in early to midsummer and is near zero from November to March. During the summer, however, the observed evapotranspiration (as estimated from the residual calculation) lies in the center of the largest model scatter during the year (spanning approximately 2 mm day$^{-1}$). Two relative peaks in total runoff occur (Fig. 8d). The larger peak, concurrent with the largest model scatter of 5 mm day$^{-1}$, is associated with the spring snow melt, and the smaller peak is associated with the fall recharge of soil-water storage. The observations suggest that low amounts of runoff, on average, occur during the summer. Though most models reduce total runoff during the summer, five of the models recharge the root-active soil layers from deeper soil-water stores (shown as negative total runoff in Fig. 8d). The two models that produce the strongest recharge are the only two models that required information regarding water-table variations (section 3a).

Though all of the models begin snow accumulation in late fall and have completely ablated the snowpack by late spring, notable differences between the models' SWE simulations exist. The most significant feature, in the context of overall model scatter, is seen during the spring snow ablation. The model scatter of SWE during the spring is at its largest and more than double the model scatter at any time preceding the onset of the spring snow melt (Fig. 8e). The largest model scatter in simulated total runoff is also coincident with the spring snowmelt scatter. The scatter in snow ablation rates is not surprising since the parameterizations of snow processes within all of the participating LSSs were given freedom of complexity (and only a fresh-snow albedo was prescribed). Nevertheless, the model scatter of spring snow ablation is notable given that identical atmospheric forcing was imposed on all the LSSs.

### 2) LNGW RUN

A central component of the model sensitivities to LWD is the snow response. As a result of the lower

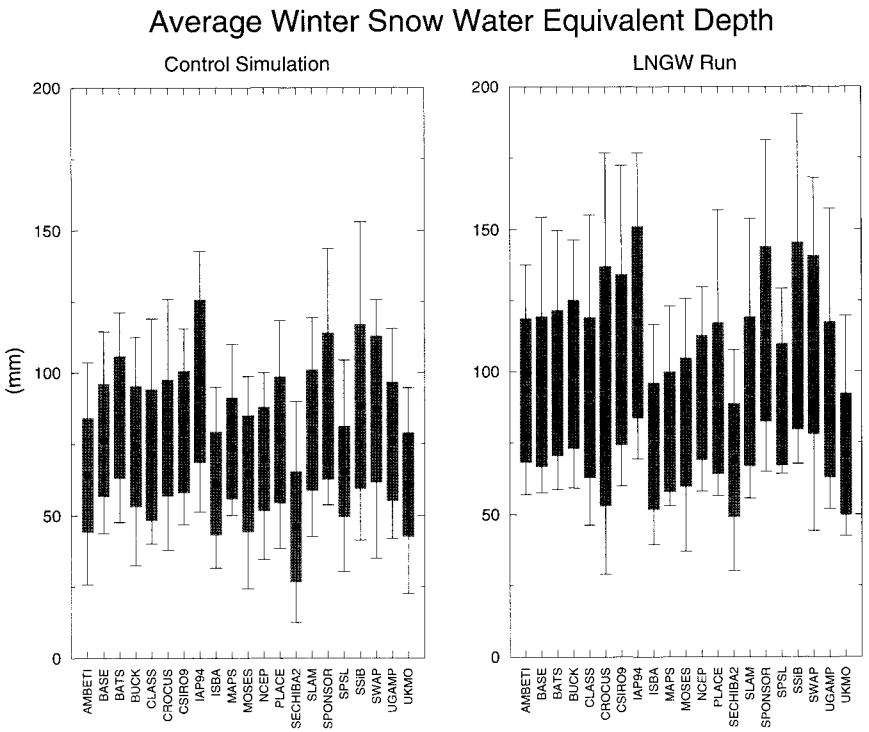


FIG. 6. As in Fig. 5 but averages are based on all of the daily averaged values of model output during the well-defined snowpack (and are therefore not compared to observations).

amounts of LWD for the LNGW runs (Fig. 1), the simulation of the winter snowpack, for all the LSSs, increases in depth (Fig. 9d). During the spring melt, the decreases in total runoff seen in March and April (Fig. 9c) imply lower amounts of meltwater and thus a weaker snowmelt for these months. However, due to the deeper snowpacks, and thus a larger source of meltwater, the subsequent increase of total runoff resulting from the delayed snowmelt is larger than the preceding negative counterpart. As a result of the larger increase in runoff (i.e., less meltwater infiltration), less water is available for evapotranspiration (Fig. 9b). Although not shown, the decrease in the models' evapotranspiration in April and May is largely (about 75%) composed of depletions in bare-soil evaporation and transpiration (and not evaporation from the snow surface).

### *c. Hydrologic events and the recursive year runs*

Two of the recursive year sensitivity runs, RY72 and RY74, were chosen for the hydrologic extremes that occurred within the years of the test runs (1972 and 1974, respectively). In 1972, a period of below-average rainfall was observed from June to August, and in 1974 during July, monthly precipitation was well above average (Figs. 10a,c, respectively; Schlosser et al. 1997). The choice of performing a recursive run for 1983 was based on the fact that it is the last year of the simulation period and relatively normal conditions, averaged for the catchment, persist through the year. The results of each of the recursive year runs were found to be quite similar, and as a result, the results for 1983 are not shown. However, due to the extreme hydrologic events that occur in 1972 and 1974, these results for the control and recursive year tests are discussed below.

#### 1) CONTROL RUN

The control simulations of the models for 1972 and 1974 are summarized in Figs. 10–12. In 1972, nearly all of the models tend to underestimate the observed trend of root-zone total soil moisture in July (Fig. 10a). At the beginning of September when precipitation returns to normal (Schlosser et al. 1997), the range of accumulated summer drying by the models is large but spans the entire range of the observed spatial variability. All but three of the models underestimate the accumulated catchment average drying. All the models use excess evaporation as the primary mechanism to dry the soil and most produce small amounts of total runoff from the root-active zone consistent with observations (Figs. 11a,c). In addition, most models agree with the lysimeter observations in producing the highest rates of evaporation in June. Some of the models recharge the depleted root-active soil layers by upward transport of soil water (shown as negative total runoff from the root-active zone in Fig. 11c), but direct measurements of this process were not available for validation. However, the

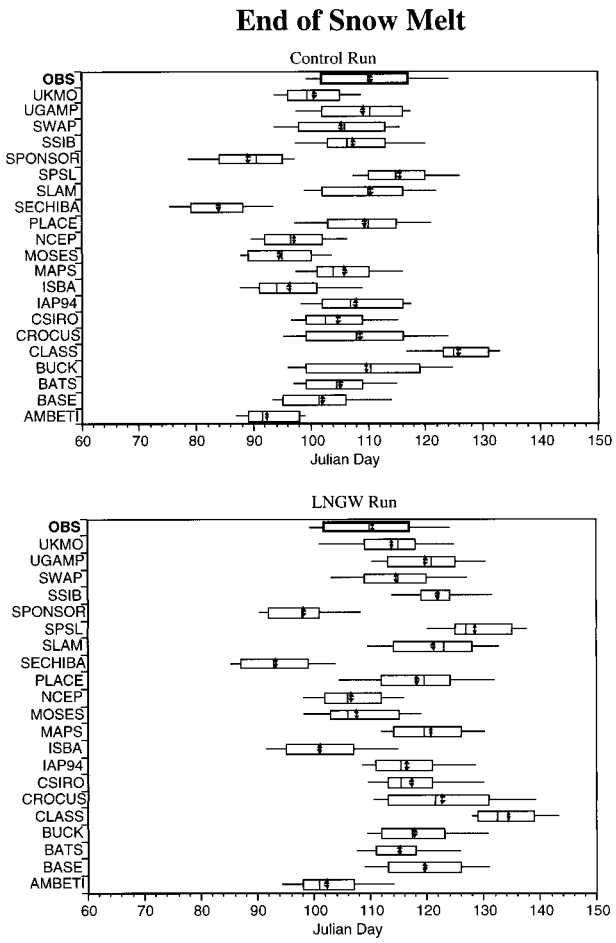


FIG. 7. Day of the completion of the spring snow melt (given as Julian day) as simulated by the models for the control and LNGW runs and seen in the observations. The completion of the snow melt is defined as the day in which total ablation of the well-defined winter snowpack is achieved. For each model and the observations, the results are given by the median value (vertical doubled-arrow line), the average value (vertical solid line), the upper and lower quartile (boxed region), and period along which 95% of the snow melts were completed (horizontal solid line) for the 18 yr of the simulation (1966–83).

modified streamflow measurements (Schlosser et al. 1997) imply little, if any, recharge of root-zone soil water.

The control simulation results of the wet summer of 1974 are qualitatively similar to the dry summer of 1972 in that most of the models capture the anomalous trends in root-zone total soil moisture within the range of the observed spatial variability (Fig. 10c). Most models simulate an anomalous increase in runoff during July that is qualitatively consistent with the observations (Fig. 12c). However, the range of total runoff simulated by the models is 2 mm day$^{-1}$, which is larger than the total runoff that is observed. In addition, a large portion of the model scatter lies below the observed value of total runoff during July, which is consistent with most of the models simulating higher amounts of evaporation (Fig. 12a).

#### 2) RECURSIVE RUNS

The results of the RY72 and RY74 runs are summarized in Figs. 10–12 (as previously mentioned, similar results are found for the RY83 simulations). The impact of using precipitation recursively is most evident in the SWE differences (bottom panels of Figs. 11 and 12). Large differences between the control and recursive runs are seen at the beginning of the year due to the different accumulation of frozen precipitation in the preceding months. This leads to large differences in total runoff during the spring as a result of different magnitudes of meltwater (Figs. 11d and 12d). By the summer, differences in total runoff, as well as evaporation (Figs. 11b and 12b), for most of the models are much smaller. However for a few of the models, the differences in total runoff and evaporation persist, but these models also show very long timescales of convergence from the analysis of their NOSU runs (described in the next section). It is more difficult to generalize the features of the recursive runs for root-zone soil moisture (Figs. 10b,d). However, an overall seasonal feature can be discerned. During the early part of the year up to the snow melt, the highest biases with respect to the control run for root-zone total soil moisture occur. These biases then significantly decrease (with the exception of four of the models) in the summer as a result of soil saturation that typically occurs during the spring snow melt (Fig. 8a). Comparing the collection of model results between all three of the recursive runs, there is no single recursive run that stands out as having the lowest or highest overall biases. Therefore whether the recursive run, starting on 1 January, was performed for the year with a dry, wet, or normal summer does not appear to be important in the biases that result.

### *d. Sensitivity to initial conditions*

The impact of the prescribed initial conditions with no recursive spinup performed in the first year (i.e., the NOSU simulation) is assessed through a convergence test with the control run. We define convergence as the year into the simulation after which all differences between the NOSU run and the control run for monthly averaged temperatures, soil-water stores, and heat fluxes (Table 3) are within 0.01 K, 0.1 mm, and 0.1 W m$^{-2}$, respectively. Any result of an analysis such as this is sensitive to the definition and threshold of convergence (Yang et al. 1995). However, our convergence requirements are identical to the equilibrium criteria used by Chen et al. (1997) for the PILPS 2(a) Cabauw experiments for consistency. In addition, for the spinup tests of Yang et al. (1995), the identical heat flux equilibrium constraint was applied to the annual averaged quantities.

The convergence of the NOSU run to the control

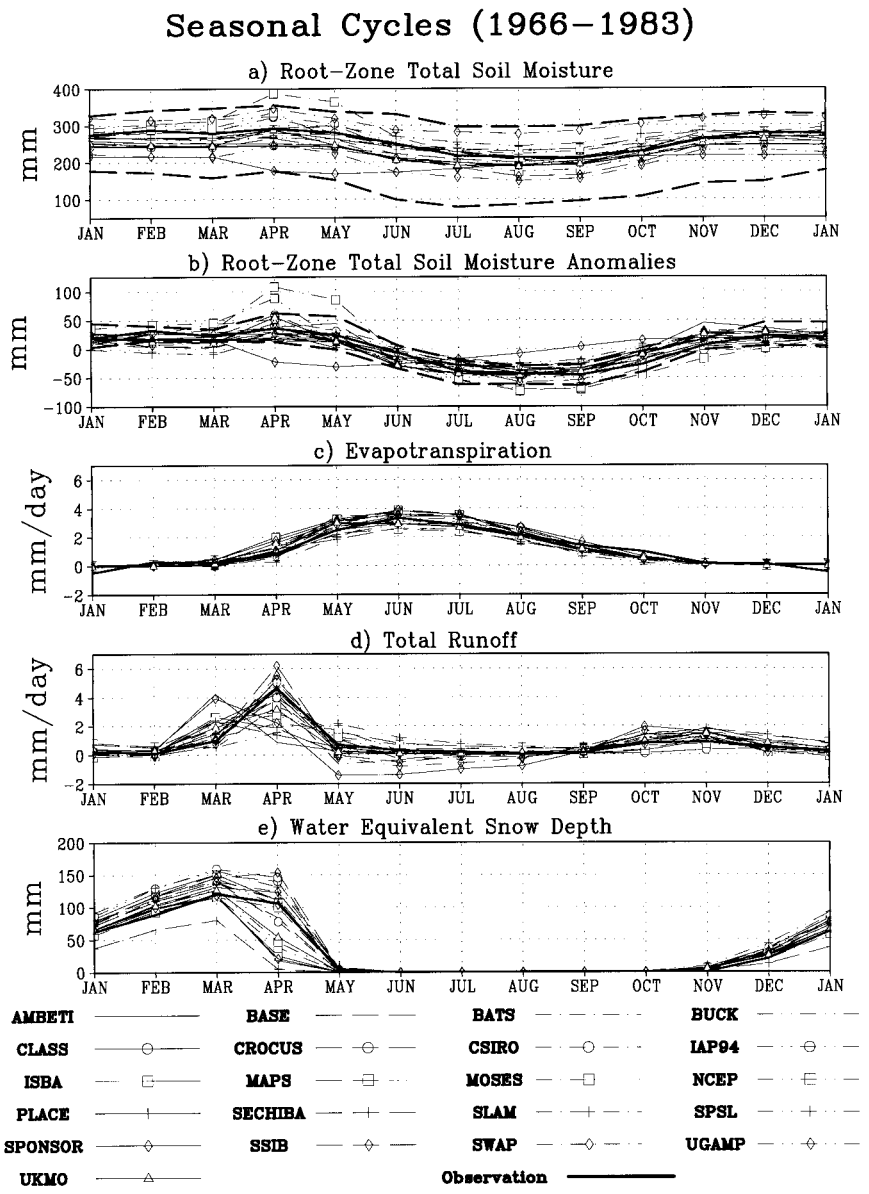


FIG. 8. Seasonal cycles of (a) root-zone total soil moisture, (b) root-zone total soil moisture anomalies (with respect to the simulation mean), (c) evapotranspiration, (d) total runoff from the root-active zone (negative values indicating upward transport of soil water from deeper soil layers), and (e) water equivalent snow depth from the control-run model simulations and observations. The fluxes are based on monthly averages and root-zone total soil moisture and water equivalent snow depth based on daily values at the beginning of each month for the simulation period, 1966–83. Observed evapotranspiration is determined by the residual calculation. For observed total soil moisture, the catchment averaged value is given as a thick solid line, and the highest and lowest values of the seasonal cycles for the 11 observation sites within the catchment are indicated by the thick dashed lines.

simulation is achieved within the first year of the simulation for five models (Fig. 13) of which one is the only model with no soil thermal layers (Fig. 2). For the remaining models, the points of convergence span the entire simulation period. Five of the models do not achieve convergence until the final year of the simulation, and one model is unable to reach convergence for the simulation period. The results of Yang et al. (1995) suggest that the timescales of thermal and hydrologic equilibrium in LSSs are controlled by total water holding capacity and soil moisture initialization. However, a large range in the convergence of the NOSU runs exists even though the soil-water column and total water holding capacity is the same for all models (sec-

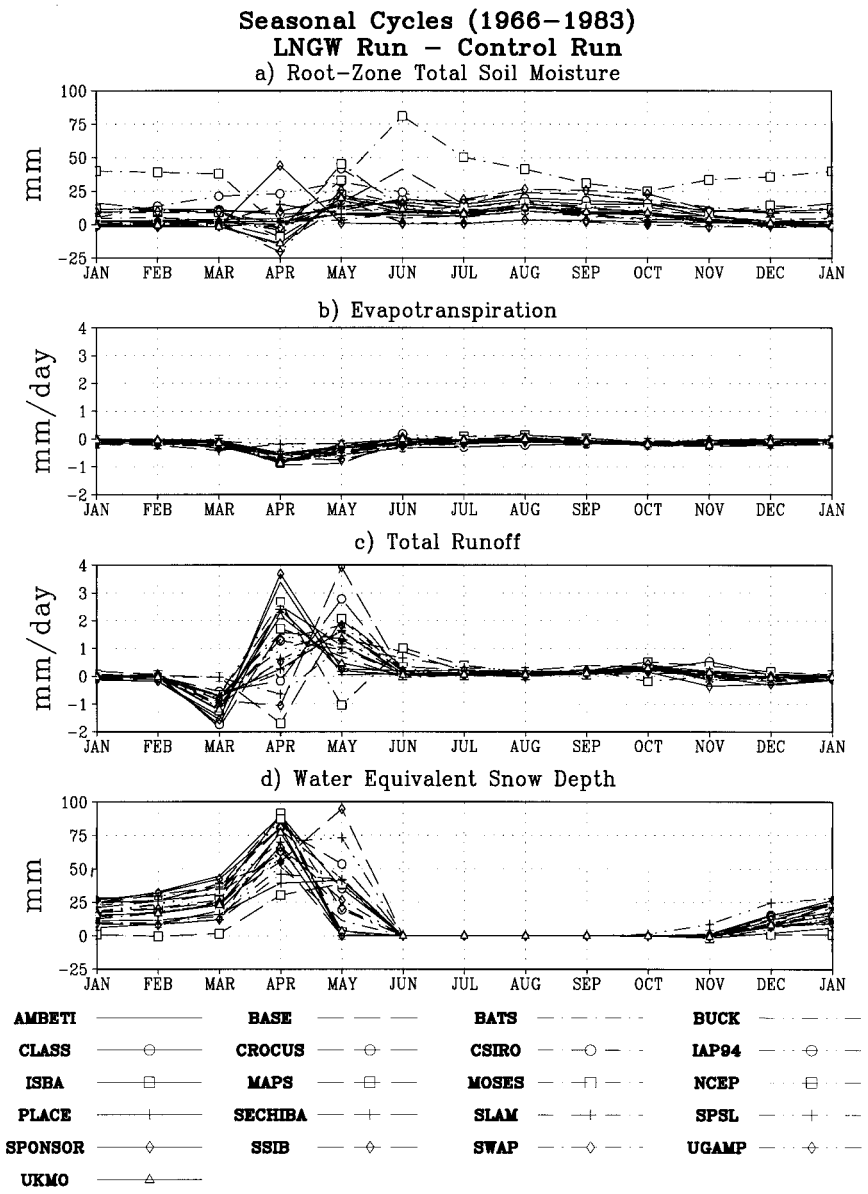


FIG. 9. Difference between the LNGW and control run of the seasonal cycles (as given in Fig. 8) of root-zone total soil moisture, evapotranspiration, total runoff from the root-active zone, and water equivalent snow depth.

tion 3a), and the models' soil-water stores are initialized with an adequate water supply.

These results warrant a more thorough analysis, which is beyond the scope of this summary paper. Nevertheless, further testing reveals that even after multiplying the convergence criteria (given above) by a factor of 100, a notable model scatter in convergence still exists. Six of the models require more than 5 yr to achieve convergence and the remaining models converge within the first 3 yr. Moreover, the findings of Vinnikov et al. (1996) indicate that the timescale of soil-moisture autocorrelation at Usadievskiy is about 3 months (i.e., less than a year). With respect to these convergence tests, the results do not differ significantly (i.e., model scatter still exists) for the case where only the constraints for modeled root-zone total soil moisture are applied. Given the wide range of soil thermal discretization (Fig. 2) coupled with the large disparity in the treatment of frozen soil (moisture) processes (Table 4), their effects on the model scatter of these convergence tests are, most likely, significant.

## 5. Discussion

### *a. Summary of the simulation results*

After completion of the prescribed PILPS 2(d) simulations, model scatter is still evident. We have analyzed

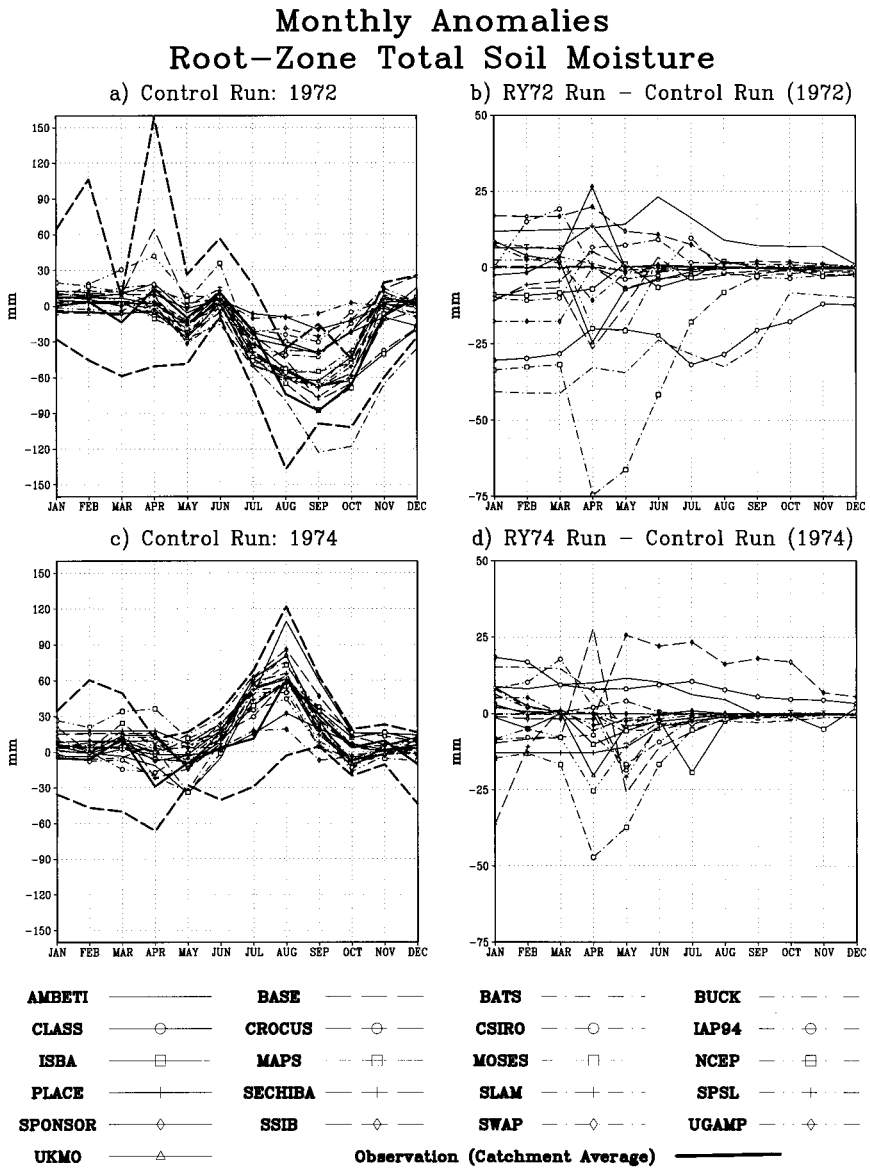


FIG. 10. Monthly anomalies (seasonal cycle removed) of root-zone total soil moisture for the model simulations and observations for the years (a) 1972 and (c) 1974. The differences in simulated soil moisture between the corresponding year of the control run and the (b) RY72 and (d) RY74 runs are also shown. Tick marks refer to the first day of the month. For observed total soil moisture, the catchment averaged value is given as a thick solid line, and the highest and lowest values of the seasonal cycles for the 11 observation sites within the catchment are indicated by the thick dashed lines.

the simulations of root-zone soil moisture in light of the observed spatial variability among 11 measurement sites within the catchment. The framework of the participating LSSs does not allow us to explicitly account for the processes that affect the intracatchment variability, and therefore this source of disagreement should be taken into consideration for the evaluation of the LSSs' performance. The resulting analysis indicates that in nearly all cases, the models' root-zone soil moisture fall within the observed spatial variability. However, notable model disagreements (i.e., simulations that fall outside the observed scatter) are seen in the anomalies of root-zone soil moisture.

Averaged over the entire simulation, nearly all of the models partition a larger portion of incident precipitation to evaporation rather than runoff, similar to the observations. While the ratio of evaporation to runoff partitioning varies by about 40% (Fig. 4a), efforts have already been successful at significantly improving the results of the extreme outliers. While the results of these

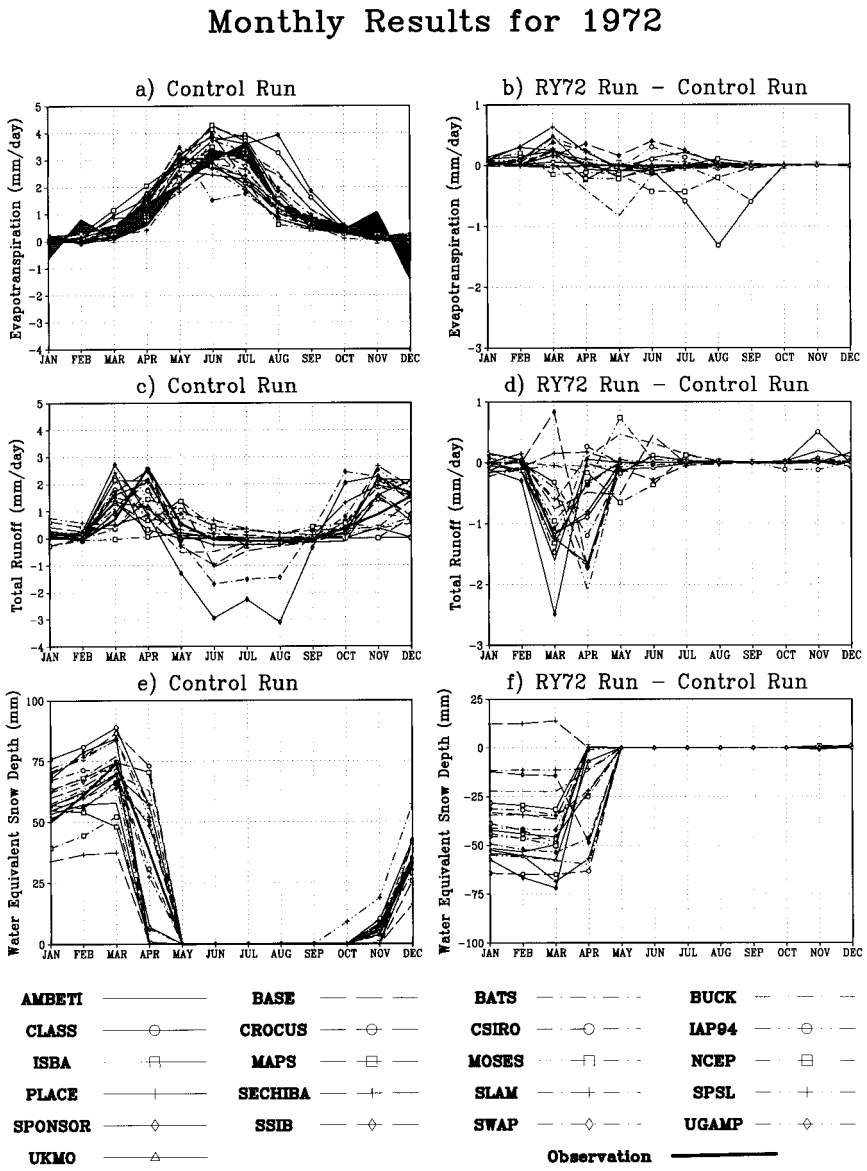


FIG. 11. Model calculations and observations for 1972 of the control run and the modeled differences between 1972 of the control run and the corresponding RY72 run. (a) and (b) Monthly averaged evapotranspiration. Observations include both the residual estimates (thick solid line) and lysimeter measurements (thick dashed line) with their differences highlighted by the gray shaded area (differences in monthly trends of the observations are denoted by a crossing of the gray shaded regions). (c) and (d) Monthly averaged total runoff from the root-active zone. (e) and (f) Water equivalent snow depth sampled at the first day of each month.

specific model refinements are beyond the scope of this paper (and will likely be included in future publications that document a particular LSS's development), in nearly all cases the improvements were not obtained through major changes to their model framework, but rather slight code modifications.

Over the averaged annual cycle, all the models produce the largest amounts of runoff in response to the spring snow melt, which is qualitatively consistent with the direct observations of runoff. The models differ, however, in the phase and magnitude of the spring runoff peak, which can be primarily attributed to differences in snow ablation rates (discussed below) as well as differences in meltwater partitioning between runoff and infiltration. Model disparity in meltwater partitioning is likely a result of the interactions between the various parameterizations of soil thermal discretization (Fig. 2) and runoff criteria (e.g., Robock et al. 1998) employed by the LSSs. These interactions are complex and require further attention that is beyond the scope of this paper.

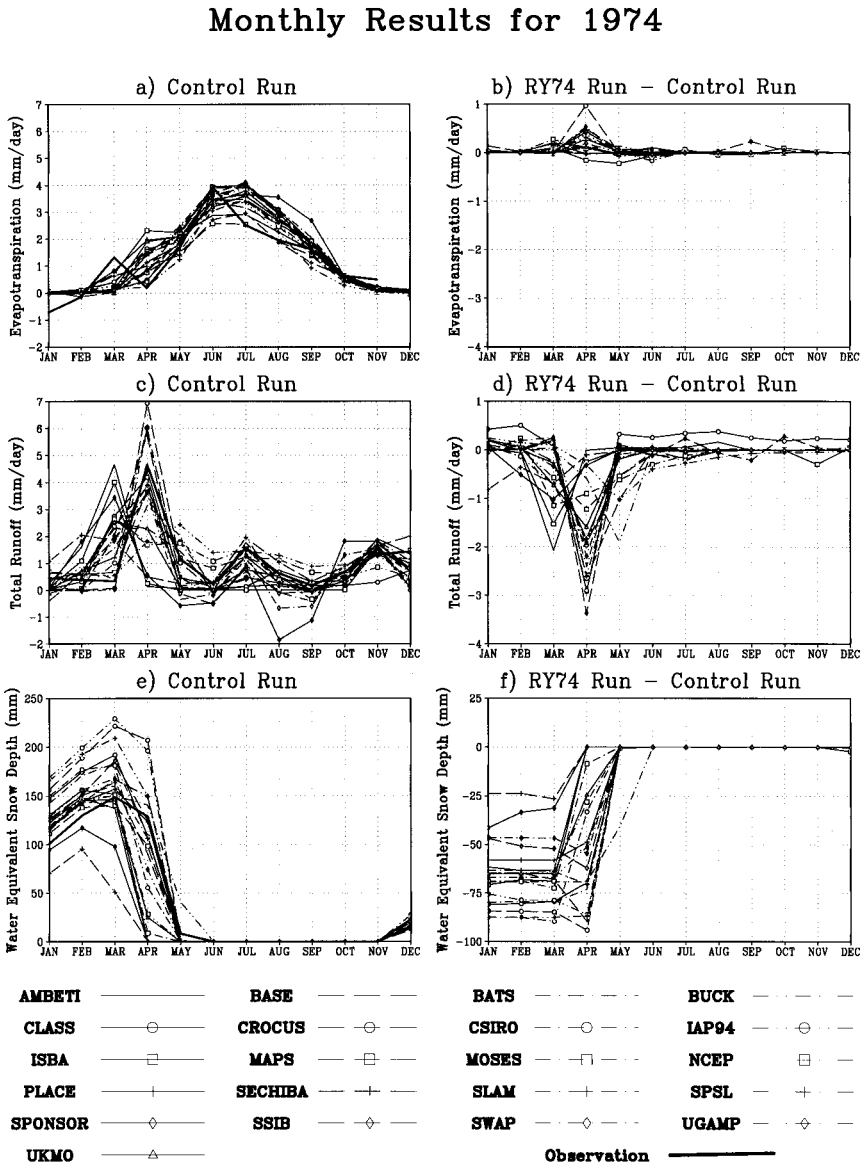


FIG. 12. As in Fig. 11 but for 1974 of the control run and the corresponding RY74 run and observations of evaporation include only the residual estimates.

Generally speaking, all of the models show some level of skill in simulating the seasonal variation of evapotranspiration. Over the averaged annual cycle, the models peak their evapotranspiration rates during the early and middle summer, which is consistent with the residual evaporation estimates. However, model scatter of $\pm 1$ mm day$^{-1}$ during the warmer months is large. Although special care was taken to prescribe parameter values that were as consistent to the environmental conditions of the site, as well as to the differences in their functionality among the models, the effects of model parameter assignment cannot be discounted as a source for scatter (Polcher et al. 1996). Moreover, differences in the model formulations of potential evaporation are also likely to contribute (McKenney and Rosenburg 1993). Without comprehensive testing by each of the participating models to address these issues, it is difficult to quantify the significance and causes of the scatter.

### *b. Simulation of snow processes*

Model comparisons of the seasonal cycle of SWE indicate a stronger sensitivity of the simulated spring ablation period, as opposed to the accumulation of the winter snow pack, to the choice of model parameteri-

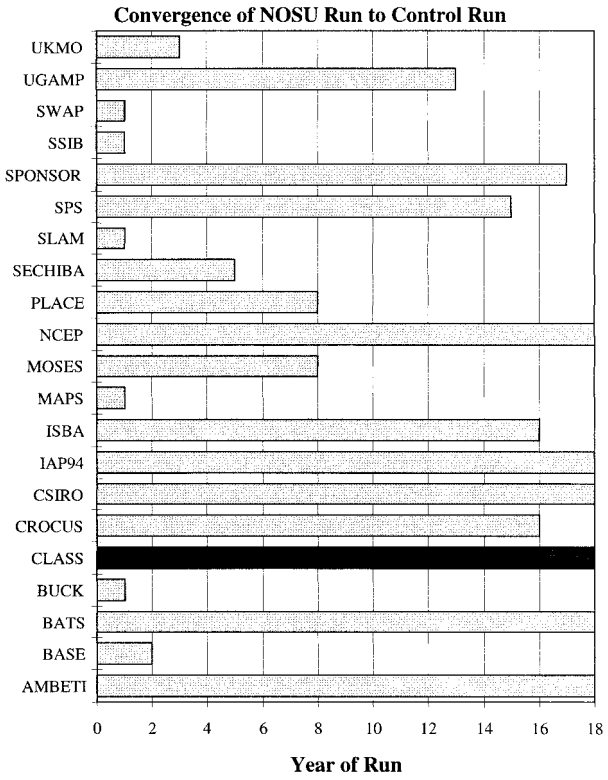


FIG. 13. Results of the NOSU test of convergence (as defined in the text) to the control run.

zation. The larger model scatter during the spring melt also occurs for the LWD sensitivity run (although not explicitly shown). The model scatter during the spring ablation period of the seasonal snow cover is twice the scatter over the accumulation period (excluding the model with the lowest SWE simulation in Fig. 8e, the difference then becomes a factor of 3). This result has also been found in comparisons of SWE simulations from GCMs (Foster et al. 1996) and suggests that these results have broader implications toward model development.

Given that the LSSs used identical atmospheric forcing for the PILPS 2(d) simulations, most notably precipitation, near-surface air temperature, and incoming radiation, the cause of the disparity of the model simulations are solely attributed to differences in snow ablation. Our initial analysis indicates that model disparity in snow ablation is a result of interactions among the various parameterizations of fractional snow cover, albedo, snow thermal properties, and snow model structure used by the LSSs. This was most evident for the aforementioned model with low simulated SWE (Fig. 8e). A combination of the prescribed snow parameters (section 3c) and its snow heterogeneity framework caused a lower albedo for snow-covered conditions, which lead to increased solar energy absorption; higher snowmelt rates; and decreased SWE and fractional snow coverage, which then further decreased albedo (i.e., a positive feedback). While this does not reflect an error in the model, it does serve as an example for the special care that must be taken when prescribing parameter values for coordinated, multimodel snow simulations.

### *c. Sensitivity results and their implications*

The results from the LNGW runs reveal some important considerations for the model simulations with regard to validation and development using the Valdai data. First, the shift in the partitioning of total runoff and evaporation (averaging about 0.2 mm day$^{-1}$ for all the models) in response to the different longwave forcing (Fig. 4) should be considered, in the context of model development, a source of model disagreement with the observations. Second, the difference in the two simulated LWD forcings is also quite comparable to the level of uncertainty in the instrumentation to measure LWD. Therefore, the results of the LNGW run can also be regarded as potential sensitivity issues to the random errors of instrumentation for the case of using observed LWD as forcing. Last, the simulations of the accumulation and ablation of the snowpack as well as meltwater partitioning between runoff and evapotranspiration show a marked sensitivity to longwave radiation. This result is consistent with the findings of previous work using data from other Russian sites (Yang et al. 1997; Slater et al. 1998a), but in those studies only one LSS was tested. From the sensitivities of 21 model simulations in this study, the results emphasize that any model development aimed to improve simulations of the snow processes using these data, or potentially other high-latitude datasets, should take into account their LWD sensitivity. In addition, LWD sensitivities to snow and meltwater processes should be carefully considered in the application of LSSs for river basin simulations (e.g., Wood et al. 1998), especially over regions where the winter snowpack is a large source for spring river discharge and water supply (e.g., the western United States).

The results of the recursive year runs underscore the importance of running multiyear LSS simulations. Not only does it allow for an assessment of the LSSs' interannual variations, it also diminishes the considerable biases that may result for simulations with only 1 yr of forcing. In our analysis, all of the models show biases in the recursive runs (compared to the corresponding year of the control run) that can persist for the entire year. In addition, the features of these biases are found to be quite similar among the 3 yr tested and indicate that a climatologically normal year is not necessarily a good candidate for a recursive run. The biases in simulated SWE from January to the end of the spring snow melt (bottom right panel in Figs. 11 and 12) are a direct result of starting the recursive runs in midwinter 1 January. Therefore, these SWE biases could be significantly diminished if the forcing for the recursive year run started in a different (i.e., warm) season, and therefore would

contain a continuous winter time series of atmospheric forcing. While this would certainly be desirable for evaluating simulated snow/frozen soil processes, the biases during the warm season in which the recursive run begins may then be affected, and thus hinder the evaluation of liquid soil-water processes. Nevertheless, our results suggest that special attention must be drawn to validating an LSS simulation that is created by recursive techniques with the atmospheric forcing.

The results of the test simulations starting from prescribed initial conditions illustrate that a wide range of spinup timescales are inherent in the models, even when we account for the controls of these processes that have been cited in previous work (Yang et al. 1995). While a further analysis of these sensitivity tests is warranted, the large range of soil thermal discretization among the models that were allowed, coupled with the freezing conditions of the soil, likely contribute to the model scatter. A recent study focusing on the performance of one of the participating schemes at simulating soil freeze-thaw cycles (Slater et al. 1998b) suggests that seasonal changes in root-zone soil moisture are dependent on the proper spring thaw simulation. The results here suggest that the treatment of frozen soil processes in LSSs have the potential to affect longer timescales of simulated soil moisture variability. The results also draw special attention to LSS seasonal to interannual climate applications and the potential effects of improper LSS initialization at high latitudes.

## 6. Closing remarks

This paper is the first of a series that will focus on the simulations conducted by the participating models for PILPS 2(d). In the preceding sections, we presented the control run results, the models' performance against the observations, and the results of the suite of sensitivity tests. As this is the only PILPS model comparison to consider a site with seasonal snow cover and frozen soil, the implications of this particular experiment are limited in their application to the broader context of model development, but have merit to the LSS community.

From the results and discussion presented above, the continuing research with the PILPS 2(d) simulations aims to provide and publish a more thorough assessment of the model simulations of snow processes (Slater et al. 1999, manuscript submitted to *Climate Dyn.*) and runoff processes during the snow melt. Overall, the PILPS 2(d) experiment further emphasizes the utility and importance of comparing LSS simulations to each other and especially against observations. Studies such as these give modelers not only the opportunity to assess their models' performances, but also to investigate model sensitivities, detect sources of model deficiencies and improve them, expose modelers to existing sets of validation data, and provide an international forum for model development. However, this study is one among a list of four PILPS stand-alone experiments to date that have compared a collection of LSS simulations to observations. The remaining studies used data from Cabauw-Netherlands (Chen et al. 1997), HAPEX-MOBILHY (Shao and Henderson-Sellers 1996), and Red-River Arkansas (Wood et al. 1998). To assure the broadest possible evaluation of LSSs used in global climate models, efforts such as these must continue with as many validation datasets available that reflect a wide range of climatic, hydrologic, and geologic environments.

*Acknowledgments.* This work is supported by NOAA Climate and Global Change Program Grant NA56GPO212, NASA Grants NAGW-5227 and NAG55161, the Australian Research Council, and Russian Foundation for Basic Research Grants 98-05-64218 and 97-05-64467. We thank Linda Hopkins, Charmaine Franklin, and Paul Dirmeyer for their helpful reviews during the early version of the manuscript. We also wish to thank two anonymous reviewers who helped significantly clarify the paper.

REFERENCES

Auer, A. H., 1974: The rain versus snow threshold temperature. *Weatherwise,* **27,** 67.

Berlyand, M. E., 1956: *Forecasting and Control of the Heat Regime Surface Air Layer.* Gidrometeoizdat, 435 pp.

Berlyand, T. G., 1961: *The Distribution of Solar Radiation on Continents* (in Russian). Gidrometeoizdat, 227 pp.

Braden, H., 1995: The model AMBETI—A detailed description of a soil–plant–atmosphere model. Berichte des Deutschen Wetterdienstes, Offenbach am Main, Rep. 195, 117 pp.

Brun, E., P. David, and M. Sudul, 1992: A numerical model to simulate snow-cover stratigraphy for operational avalanche forecasting. *J. Glaciol.,* **38,** 13–22.

Brutsaert, W., 1975: On a derivable formula for long-wave radiation from clear skies. *Water Resour. Res.,* **11,** 742–744.

Chen, F., and Coauthors, 1996: Modeling of land surface evaporation by four schemes and comparison with FIFE observations. *J. Geophys. Res.,* **101,** 1194–1215.

Chen, T. H., and Coauthors, 1997: Cabauw experimental results from the Project for Intercomparison of Land-Surface Parameterization Schemes (PILPS). *J. Climate,* **10,** 1194–1215.

Cox, P. M., R. A. Betts, C. B. Bunton, R. L. H. Essery, P. R. Rowntree, and J. Smith, 1999: The impact of new land surface physics on the GCM simulation of climate and climate sensitivity. *Climate Dyn.,* **15,** 183–203.

Dai, Y.-J., and Q.-C. Zeng, 1997: A land-surface model (IAP94) for climate studies. Part I: Formulation and validation in off-line experiments. *Adv. Atmos. Sci.,* **14,** 433–460.

de Rosnay, P., and J. Polcher, 1999: Modeling root water uptake in a complex land surface scheme coupled to a GCM. *Hydrol. Earth Sys. Sci.,* **2,** 239–255.

Desborough, C. E., 1998: Surface energy balance complexity in GCM land surface models. *Climate Dyn.,* **15,** 389–403.

——, and A. J. Pitman, 1998: The BASE land surface model. *Global Planet. Change,* **19,** 3–18.

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

Fedorov, S. F., 1977: *A Study of the Components of the Water Balance in Forest Zone of the European Part of the USSR* (in Russian). Gidrometeoizdat, 264 pp.

Foster, J., and Coauthors, 1996: Snow cover and snow mass intercomparisons of general circulation models and remotely sensed datasets. *J. Climate,* **9,** 409–426.

Gedney, N., 1995: Development of a land surface scheme and its application to the Sahel. Ph.D. dissertation, University of Reading, 200 pp. [Available from Dept. of Meteorology, University of Reading, Reading, RG6 6BB United Kingdom.]

Gusev, Y. M., and O. N. Nasonova, 1998: The land surface parameterization scheme SWAP: Description and partial validation. *Global Planet. Change,* **19,** 63–86.

Henderson-Sellers, A., A. Pitman, P. Love, P. Irannejad, and T. Chen, 1995: The Project for Intercomparison of Land Surface Parameterization Schemes (PILPS): Phases 2 and 3. *Bull. Amer. Meteor. Soc.,* **76,** 489–503.

Idso, S. B., 1981: A set of equations for full spectrum and 8–14m and 10.5–12.5m thermal radiation from cloudless skies. *Water Resour. Res,* **17,** 295–304.

Kim, J., and M. Ek, 1995: A simulation of the surface energy budget and soil water content over the HAPEX-MOBILHY forest site. *J. Geophys. Res.,* **100,** 20 845–20 854.

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

McKenney, M. S., and N. J. Rosenburg, 1993: Sensitivity of some potential evapotranspiration estimation methods to climate change. *Agric. For. Meteor.,* **64,** 81–110.

Miskolczi, F., and R. Guzzi, 1993: Effect of the non-uniform spectral dome transmittance on the accuracy of the IR radiation measurements using shielded pyradiometers and pyrgeometers. *Appl. Opt.,* **31,** 608–612.

Monteith, J. L., 1973: *Principles of Environmental Physics.* Edward Arnold, 241 pp.

Noilhan, J., and J.-F. Mahfouf, 1996: The ISBA land surface parameterization scheme. *Global Planet. Change,* **13,** 145–159.

Pitman, A. J., and Coauthors, 1999: Key results and implications for phase 1(c) of the Project for Intercomparison of Land-Surface Parameterization Schemes. *Climate Dyn.,* **15,** 673–684.

Polcher, J., K. Laval, L. Dumenil, J. Lean, and P. R. Rowntree, 1996: Comparing three land surface schemes used in general circulation models. *J. Hydrol.,* **180,** 373–394.

Qu, W., and Coauthors, 1998: Sensitivity of latent heat flux from PILPS land-surface schemes to perturbations of surface air temperature. *J. Atmos. Sci.,* **55,** 1909–1927.

Robinson, D. A., K. F. Dewey, and R. R. Heim, 1993: Global snow cover monitoring: An update. *Bull. Amer. Meteor. Soc.,* **74,** 1689–1696.

Robock, A., K. Ya. Vinnikov, C. A. Schlosser, N. A. Speranskaya, and Y. Xue, 1995: Use of midlatitude soil moisture and meteorological observations to validate soil moisture simulations with biosphere and bucket models. *J. Climate,* **8,** 15–35.

——, C. A. Schlosser, K. Ya. Vinnikov, N. A. Speranskaya, J. K. Entin, and S. Qiu, 1998: Evaluation of the AMIP soil moisture simulations. *Global Planet. Change,* **19,** 181–208.

Satterlund, D. R., 1979: An improved equation for estimating longwave radiation from the atmosphere. *Water Resour. Res.,* **15,** 1649–1650.

Schlosser, C. A., 1995: Land-surface hydrology: Validation and intercomparison of multi-year off-line simulations using midlatitude data. Ph.D. dissertation, University of Maryland, 135 pp. [Available from Office of Graduate School, University of Maryland, College Park, MD 10742.]

——, A. Robock, K. Ya. Vinnikov, N. A. Speranskaya, and Y. Xue, 1997: 18-year land surface hydrology model simulations for a midlatitude grassland catchment in Valdai, Russia. *Mon. Wea. Rev.,* **125,** 3279–3296.

Sellers, P. J., and Coauthors, 1997: BOREAS in 1997: Experiment overview, scientific results, and future directions. *J. Geophys. Res.,* **102,** 28 731–28 769.

Shao, Y., and A. Henderson-Sellers, 1996: Validation of soil moisture simulations in landsurface parameterization schemes with HAPEX data. *Global Planet. Change,* **13,** 11–46.

Shmakin, A. B., 1998: The updated version of SPONSOR land surface scheme—PILPS influenced improvements. *Global Planet. Change,* **19,** 49–62.

Slater, A. G., A. J. Pitman, and C. E. Desborough, 1998a: The validation of a snow parameterization designed for use in General Circulation Models. *Int. J. Climatol.,* **18,** 595–617.

——, ——, and ——, 1998b: The simulation of freeze–thaw cycles in a GCM land surface scheme. *J. Geophys. Res.,* **103,** 11 303–11 312.

Smirnova, T. G., J. M. Brown, and S. G. Benjamin, 1997: Performance of different soil model configurations in simulating ground surface temperature and surface fluxes. *Mon. Wea. Rev.,* **125,** 1870–1884.

Verseghy, D., N. McFarlane, and M. Lazare, 1991: CLASS—A Canadian Land Surface Scheme for GCMs. Part I: Soil model. *Int. J. Climatol.,* **11,** 111–133.

Vinnikov, K. Ya., A. Robock, N. A. Speranskaya, and C. A. Schlosser, 1996: Scales of temporal and spatial variability of midlatitude soil moisture. *J. Geophys. Res.,* **101,** 7163–7174.

Warrilow, D. A., and E. Buckley, 1989: The impact of land surface processes on the moisture budget of a climate model. *Ann. Geophys.,* **7,** 439–450.

Wetzel, P., and A. Boone, 1995: A parameterization for land–atmosphere–cloud exchange (PLACE): Documentation and testing of a detailed process model of the partly cloudy boundary layer over heterogeneous land. *J. Climate,* **8,** 1810–1837.

Williams, P. J., and M. W. Smith, 1989: *The Frozen Earth.* Cambridge University Press, 306 pp.

Wood, E. F., and Coauthors, 1998: The Project for the Intercomparison of Land-Surface Parameterization Schemes (PILPS) Phase 2(c) Red-Arkansas River Basin Experiment. Part 1: Experiment design and summary intercomparisons. *Global Planet. Change,* **19,** 115–135.

Xue, Y., F. J. Zeng, and C. A. Schlosser, 1996: SSiB and its sensitivity to soil properties—A case study using HAPEX-MOBILHY data. *Global Planet. Change,* **13,** 183–194.

Yang, Z.-L., R. E. Dickinson, A. Henderson-Sellers, and A. J. Pitman, 1995: Preliminary study of spin-up processes in land surface models with the first stage of Project for Intercomparison of Land Surface Parameterization Schemes Phase 1(a). *J. Geophys. Res.,* **100,** 16 553–16 578.

——, ——, A. Robock, and K. Ya. Vinnikov, 1997: Validation of the snow submodel of the Biosphere–Atmosphere Transfer Scheme with Russian snow cover and meteorological observational data. *J. Climate,* **10,** 353–373.