Using different hydrological variables to assess the impacts of atmospheric forcing errors on optimization and uncertainty analysis of the CHASM surface model at a cold catchment

Youlong Xia[1]

Atmospheric and Oceanic Science Program, Princeton University, Princeton, New Jersey, USA

Zong-Liang Yang

Department of Geological Sciences, The John A. and Katherine G. Jackson School of Geosciences, University of Texas at Austin, Austin, Texas, USA

Paul L. Stoffa and Mrinal K. Sen

Institute for Geophysics, The John A. and Katherine G. Jackson School of Geosciences, University of Texas at Austin, Austin, Texas, USA

Received 15 June 2004; revised 19 October 2004; accepted 1 November 2004; published 4 January 2005.

[1] Estimation of parameters for land-surface models, along with their corresponding uncertainties, relies on the input data for the atmospheric forcing variables including atmospheric pressure, temperature, humidity, wind speed, precipitation, and incoming shortwave and longwave radiation. Most studies assume that forcing data are accurate and contain no random or systematic observational errors. In practice, there are indeed systematic errors in precipitation measurements, especially for snowfall, due to wind-caused undercatch. Incoming shortwave and longwave radiation fluxes are often not directly measured, but estimated from empirical formulations. Uncertainties in these forcing data may substantially affect optimization and uncertainty estimates of land surface models. In this study, we used 18-year forcing and calibration data as well as information about the uncertainties in the forcing variables at Valdai, Russia, to study the impacts of forcing errors on selection of optimal model parameters and their uncertainty estimates when three different hydrological variables were used for calibration. The results show that forcing errors have few effects on the selection of optimal model parameter sets when monthly evapotranspiration and runoff are calibrated. However, forcing errors do introduce significant effects on the selection of optimal model parameters when daily snow water equivalent is calibrated. Forcing errors also significantly affect uncertainty estimates of the land surface model parameters. In addition, constraints of forcing errors are different when different hydrological variables are calibrated. All three hydrological variables constrain the incoming longwave radiation error well, and the snow water equivalent and runoff constrain winter snowfall errors well. However, all three hydrological variables cannot constrain the incoming solar radiation error well. We highlight in this study that runoff is shown to be a good observable to use for calibration, the reason being that it integrates multiple hydrological processes; and the results support the theory that typical rain/snow gauges have 10–20% undercatch.

Citation: Xia, Y., Z.-L. Yang, P. L. Stoffa, and M. K. Sen (2005), Using different hydrological variables to assess the impacts of atmospheric forcing errors on optimization and uncertainty analysis of the CHASM surface model at a cold catchment, *J. Geophys. Res.*, *110*, D01101, doi:10.1029/2004JD005130.

1. Introduction

[2] The principal purpose of the Project for Intercomparison of Land-surface Parameterization Schemes (PILPS) is to understand land surface processes and to improve their parameterizations [*Henderson-Sellers et al.*, 1996]. To achieve this purpose, PILPS initiated a four-phase comparison effort. Phases 1 and 2 involve stand-alone and regional offline simulations of land surface models driven by synthetic and observed atmospheric forcing, respectively. Phases 3 and 4 investigate the performance of land surface models in fully coupled simulations with a host regional atmospheric model or general circulation model. The experiment in phase 1 was performed using synthetic forcing data [*Pitman et al.*, 1999]. The experiments in phase

[1]Also at Geophysical Fluid Dynamics Laboratory, NOAA, Princeton University, Princeton, New Jersey, USA.

0148-0227/05/2004JD005130$09.00

2 were performed using observed forcing data for different climate zones, soils and vegetation types from local comparisons, for example, phase 2a (a midlatitude grassland site at Cabauw, Netherlands [*Chen et al.*, 1997]), phase 2b (a soybean site at Camount, France [*Shao and Henderson-Sellers*, 1998]), and phase 2d (a boreal grassland site at Valdai, Russia [*Schlosser et al.*, 2000; *Slater et al.*, 2001]) to regional comparisons such as phase 2c (Red River basin, Arkansas, U.S. [*Wood et al.*, 1998]) and phase 2e (Torne River basin, Sweden [*Bowling et al.*, 2003]). The results from PILPS have revealed large uncertainties and significant differences among the land surface models [*Henderson-Sellers*, 1996; *Henderson-Sellers et al.*, 1996; *Sellers et al.*, 1997; *Desborough*, 1999]. In particular, PILPS phase 2d and 2e have identified and documented two key atmospheric variables, precipitation and downwelling longwave radiation, that are critical for accurately determining the snow mass balances, across different scales in cold regions.

[3] However, almost all these studies concerning the derivation of optimal land surface parameters assume that the forcing data are accurate and do not contain observational errors [*Sellers et al.*, 1989; *Franks and Beven*, 1997; *Gupta et al.*, 1999; *Leplastrier et al.*, 2002; *Xia et al.*, 2002; *Jackson et al.*, 2003; *Xia et al.*, 2004a, 2004b, 2004c]. In fact, significant systematic errors (biases) in precipitation measurement, obviously caused by wind, existed in all types of precipitation gauges, in particular for snowfalls [*Groisman et al.*, 1991; *Yang et al.*, 1995; *Groisman et al.*, 1996; *Yang et al.*, 1998]. These errors affect analysis of macro scale water fluxes [*Milly and Dunne*, 2002a, 2002b], simulations of regional land surface models [*Pan et al.*, 2003; *Lohmann et al.*, 2004], and simulations of global land water and energy fluxes [*Milly and Shmakin*, 2002a, 2002b]. Furthermore, *Milly and Shmakin* [2002b] believed 'insufficient characterization and control of errors in model forcing, especially precipitation, is currently the "bottleneck" or limiting factor for simulation of macro scale water fluxes.' Certainly, the precipitation errors also result from spatial and temporal sampling biases and topographic effects. However, for cold regions, bias of snowfall resulting from wind remains the major error. Besides precipitation errors, radiation errors also significantly affect simulations of water fluxes and energy fluxes both for a catchment simulation [*Yang et al.*, 1997; *Schlosser et al.*, 1997, 2000; *Slater et al.*, 2001] and for a global simulation [*Milly and Dunne*, 2002b], especially in cold regions.

[4] The errors in a variable computed by a land surface model may come from intrinsic model errors, forcing errors (errors in forcing data) and/or model parameter errors. These undoubtedly interact in nonlinear ways. Nonlinear interaction of errors and correlation between the errors can be studied using a correlation matrix derived by a Bayesian stochastic inversion [*Sen and Stoffa*, 1996]. In order to investigate how the errors in these forcing data impact the derivation of optimal parameters, and to study how the results depend on the calibration variables, we designed 2 experiments for each of the calibration variables (i.e. evapotranspiration, runoff, snow water equivalent). The first experiment is to use the fixed forcing data and varying model parameters, and the second experiment is to use both varying forcing data and varying model parameters. Therefore, we conducted a total of 6 experiments. For each experiment, Bayesian stochastic inversion selects 60,000–90,000 parameter sets. Therefore, we have almost 500,000 model runs. For all experiments, we used a one-year spin-up period to minimize the impact of the initial condition on simulations of monthly runoff, monthly evapotranspiration and daily snow water equivalent. This means that the model was integrated for one year prior to the start of each experiment. This one-year spin-up time is appropriate for the CHASM land surface model according to *Schlosser et al.* [2000].

[5] In order to explain our results more clearly in this study, we represent the forcing errors as overall errors from precipitation, solar radiation, and downward longwave radiation. The forcing error for a variable (say precipitation) represents the difference between the fixed forcing value multiplied by a ratio factor (see Table 1) and the fixed forcing value.

2. Data, Model, and Method

2.1. Forcing and Calibration Data Sets

[6] Observational data from Valdai (57.6°N, 33.1°E), Russia, have been used to test the representation of snow accumulation, snowmelt, and frozen soil processes in land surface models [e.g., *Robock et al.*, 1995; *Vinnikov et al.*, 1996; *Schlosser et al.*, 1997, 2000; *Luo et al.*, 2003]. *Fedorov* [1977], *Vinnikov et al.* [1996], and *Schlosser et al.* [1997] described the details of the continuous 18 years of atmospheric forcing and hydrologic data at Valdai. Here we give a brief overview, for completeness. The Valdai water-balance station is located at a catchment with an area of about 0.36 km^2 in a boreal forest region. The vegetation cover is mainly grassland meadow. The climate at Valdai is highly seasonal with an annual temperature range of 35°C and an annual average precipitation of 730 mm. The majority of precipitation falls in the summer and autumn months. Near surface air temperatures rise above 15°C in summer and fall below −10°C in winter. Continuous snow cover exists from November to April.

[7] The atmospheric forcing data include atmospheric pressure, air temperature, humidity, wind speed, and incoming shortwave and longwave radiation. Atmospheric pressure, air temperature, and humidity were recorded at a height of 2 m. Wind speed was recorded at a height of 10 m. Longwave radiation fluxes that were not directly measured, but estimated from air humidity and air temperature were used in this study, following *Schlosser et al.* [1997]. The original data recorded at 3-hour intervals were interpolated to 30-minute intervals.

[8] The calibration data include monthly evaporation, runoff and daily snow water equivalent. Monthly evaporation was measured using a lysimeter from May to October for the years 1966–1973 [*Fedorov*, 1977]. Evaporation data for the remaining months (November to April) were estimated using the algorithm of *Budyko* [1956]. *Schlosser et al.* [1997] compared the monthly evaporation calculated from the residual of the water balance from the top 1 m of soil with the lysimeter measurements and found that their seasonal cycles were in good agreement. A stream gauge at the catchment outflow site was used to measure monthly runoff. To assure a more consistent comparison of the observed catchment runoff to modeled runoff from the

Table 1. Descriptions and Ranges of 12 CHASM Parameters and 4 Forcing Error Factors

Parameter	Description	Minimum Value	Maximum Value
ALBG	bare ground albedo	0.15	0.25
ALBN	snow albedo	0.65	0.85
ALBV	vegetation albedo	0.15	0.25
LEFM	maximum leaf area index	3	5
LEFS	maximum LAI seasonality	0	3
VEGM	maximum fractional vegetation cover	0.70	0.95
VEGS	fractional vegetation cover seasonality	0.00	0.50
RCMIN	minimum canopy resistance, s/m	40.0	200
RHON	snow density, kg/m^3	50	450
WRMAX	available water holding capacity, mm	200	300
Z0G	ground roughness length, m	1.0×10^{-3}	0.01
Z0V	vegetation roughness length, m	0.00	0.20
SLR	ratio factor for solar radiation	0.9	1.1
LWR	ratio factor for incoming longwave radiation	0.9	1.1
RAS1	ratio factor for January and December precipitation	0.8	1.2
RAS2	ratio factor for February, March, and November precipitation	0.9	1.1

root-active zone, the observed runoff were modified by *Schlosser et al.* [1997] according to variations in the observed averaged water table depth. At forty-four sites within the catchment, snow measurements were made at least every month during the winter and more frequently (at intervals of days) during the spring snowmelt. The ranges of errors for both the evapotranspiration and runoff were estimated to be as large as ±0.25 mm/day [*Schlosser et al.*, 2000].

2.2. Chameleon Surface Model (CHASM)

[9] The CHASM [*Desborough*, 1999; *Pitman et al.*, 2003] land surface model has been used in offline intercomparison within PILPS phase 2d [*Schlosser et al.*, 2000; *Slater et al.*, 2001] and phase 2e [*Bowling et al.*, 2003], global climate simulations [*Desborough et al.*, 2001], and regional climate simulations [*Zhang et al.*, 2001]. CHASM was designed to explore the general aspects of the land-surface energy balance representation within a common modeling framework that can be run for a variety of surface energy balance modes ranging from the simplest energy balance formulation [*Manabe*, 1969] to a complex mosaic type structure [*Koster and Suarez*, 1992]. Here we use the two-tile mosaic-type representation, in which the land-atmosphere interface is divided into two tiles. The first tile is a combination of bare ground and exposed snow and the second tile consists of dense vegetation. The tiles may be of different sizes and the energy fluxes of each tile are area-weighted. Because a separate surface balance is calculated for each tile, temperature variations may exist across the land-atmosphere interface. A prognostic bulk temperature for the storage of energy and a diagnostic skin temperature for the computation of surface energy fluxes are calculated for each tile. Snow fraction cover for both ground and foliage surfaces are calculated as functions of the snow pack depth, density, and the vegetation roughness length. The vegetation fraction is further divided into wet and dry fractions if canopy interception is considered. This model has explicit parameterizations for canopy resistance, canopy interception, vegetation transpiration and bare ground evaporation, but has no explicit canopy-air space [*Pitman et al.*, 2003].

[10] CHASM uses the formulation of *Manabe* [1969] for the hydrologic component of the land surface in which the root zone is treated as a bucket with finite water holding capacity. Any water accumulation beyond this capacity is assumed to be runoff. In addition to storage as moisture in the root zone, water can be stored as snow on the ground or on the canopy. Soil temperature is calculated within four soil layers using a finite difference method and zero-flux boundary condition. Each tile has four evaporation sources including canopy evaporation, transpiration, bare ground evaporation, and snow sublimation.

[11] The CHASM model (previously known as the SLAM model in the PILPS 2d intercomparison) was used in the PILPS phase 2d experiment at Valdai, and showed reasonable performance [*Schlosser et al.*, 2000; *Slater et al.*, 2001; *Luo et al.*, 2003]. Furthermore, it was also used to investigate the impacts of data length on the optimal parameter and uncertainty estimates [*Xia et al.*, 2004c].

2.3. Bayesian Stochastic Inversion

[12] Bayesian stochastic inversion (BSI) [*Sen and Stoffa*, 1996] is based on Bayes theorem and, usually, a stochastic method to select sets of parameter values from a distribution of realistic choices for model parameters. Within the Bayesian nomenclature, the relative probability for each combination of parameter values is expressed as a "posterior" probability density function (PPD), which is given mathematically as

$$\sigma(\mathbf{m}/\mathbf{d}_{\mathrm{obs}}) = \frac{\exp[-sE(\mathbf{m})]p(\mathbf{m})}{\int \exp[-sE(\mathbf{m})]p(\mathbf{m})d\mathbf{m}}, \qquad (1)$$

where the domain of integration spans the entire model parameter space $\boldsymbol{m}$, $\sigma(\boldsymbol{m}/\boldsymbol{d}_{obs})$ is the PPD, vector $\boldsymbol{d}_{obs}$ is the observational data, $E(\mathbf{m})$ is the error function, $\exp[-sE(\mathbf{m})]$ is the likelihood function, and $p(\mathbf{m})$ is the "prior" probability density function for $\boldsymbol{m}$. The shaping factor, s, is derived from the estimated errors as described in *Jackson et al.* [2003]. Because only the range for each model parameter in $\boldsymbol{m}$ is known, a uniform distribution within the ranges is used as the "prior" probability density function. This selection is the least-biased as a uniform distribution corresponds to the maximum uncertainty range.

[13] Because the PPD is multidimensional, it is difficult to visualize. Therefore, a one-dimensional projection of the

PPD (i.e. the marginal PPD) is usually displayed. Parameter inter-dependencies may be described by the covariance matrix, which is defined by

$$\mathbf{I} = \int f(\mathbf{m})\sigma(\mathbf{m}/d_{\mathbf{obs}})d\mathbf{m} \tag{2}$$

where $f(\mathbf{m}) = (\mathbf{m} - \langle\mathbf{m}\rangle)(\mathbf{m} - \langle\mathbf{m}\rangle)^T$ and $\langle\mathbf{m}\rangle$ is the vector of parameter means.

[14] We use the Very Fast Simulated Annealing Algorithm (VFSA [*Ingber*, 1989]) to stochastically select parameter sets. VFSA is a form of importance sampling that reduces the computational burden of modeling the impact of every possible combination of model parameters. The VFSA algorithm will sample more frequently those regions of the PPD that are more probable [*Sen and Stoffa*, 1996].

2.4. Very Fast Simulated Annealing

[15] One may use the temperature construct within the Metropolis algorithm [*Metropolis et al.*, 1953] to locate the global minimum of an error function by very slowly lowering the temperature parameter within

$$P = \exp\left(\frac{-\Delta E}{T}\right) \tag{3}$$

where P is the probability of acceptance of a new parameter set with positive change of error function values, ΔE is the change in the error function between the new and previous parameter sets, and T is a control parameter analogous to temperature. If the change is negative, this new parameter set is accepted. If the change is positive, and if and only if P is less than a randomly generated number between 0 and 1, the new parameter set is rejected. This is analogous to the annealing process within a physical system where the lowest energy state between atoms or molecules is reached by the gradual cooling of the substance within a heat bath. Because of this physical analog, the algorithm is called simulated annealing. In order to enhance the ability of simulated annealing to converge to the global minimum of an error function, *Ingber* [1989] introduced a new procedure for selecting parameter sets according to a temperature dependent Cauchy distribution. This modified simulated annealing algorithm is called very fast simulated annealing. Ingber's algorithm can be described as follows:

[16] Let us assume that a model parameter m_i at the kth iteration (annealing step k) is represented by $m_i^{(k)}$ such that

$$m_i^{\min} \le m_i^{(k)} \le m_i^{\max} \tag{4}$$

where $m_i^{\min}$ and $m_i^{\max}$ are the minimum and maximum values of the model parameter m_i. This model parameter value is perturbed at iteration $(k + 1)$ using $m_i^{(k+1)} = m_i(k) + y_i(m_i^{\max} - m_i^{\min})$, $m_i^{\min} \le m_i^{(k+1)} \le m_i^{\max}$ and $y_i \in [-1, 1]$. The y_i is generated from the distribution $g_T(y) = \prod_{i=1}^{NM} \frac{1}{2(|y_i|+T_i)\ln\left(1+\frac{1}{T_i}\right)} = \prod_{i=1}^{NM} g_{T_i}(y_i)$ and has a cumulative probability $G_{T_i} = \frac{1}{2} + \frac{\mathrm{sgn}(y_i)}{2}\frac{\ln\left(1+\frac{|y_i|}{T_i}\right)}{\ln\left(1+\frac{1}{T_i}\right)}$. *Ingber* [1989] showed that, for such a distribution, a global minimum can be statistically obtained by using the following cooling schedule:

$$T_i(k) = T_{0i}\exp\left(-c_i k^{\frac{1}{NM}}\right) \tag{5}$$

where T_{0i} is the initial temperature for model parameter i and c_i is a parameter used to control the temperature. The acceptance rule of very fast simulated annealing algorithm is the same as that used in the Metropolis algorithm. However, very fast simulated annealing is more efficient when compared with conventional simulated annealing. More detailed descriptions and applications of this method can be found in *Sen and Stoffa* [1996] and *Jackson et al.* [2003].

2.5. Forcing Data Errors

[17] Atmospheric forcing data are usually used to drive land surface models to produce energy and water flux simulations, and calibration data are usually used to compare these simulated variables to evaluate the performance of land surface models. Both forcing data errors and calibration data errors may place constraints on theory and model development in the land surface model community. Forcing errors and model parameter errors may be major error factors affecting the simulations of energy and water fluxes when a land surface model is used. Precipitation error is one of the crucial factors to consider in the evaluation of simulations of the water budget variables such as runoff, evapotranspiration and snow water equivalent. This is especially true for cold region simulations, because wind results in undercatch of snowfalls. *Schlosser et al.* [1997] compared two observed precipitation datasets from 1974 to 1983. One is the original precipitation observation at Valdai, and the other is the new observation in which effects of wind biasing on the catchment of snowfall and rainfall at a gauge station were minimized by surrounding the rain gauge with natural shrub-type vegetation. The comparison results showed that the original precipitation estimate is low by about 20% for January and December and 10% for February, March and November due to the aerodynamic effects on wind-blowing snow. Therefore, *Schlosser et al.* [1997] used a monthly ratio of the two precipitation datasets to correct the original precipitation for each time step. This corrected precipitation was further used as the default forcing in the studies of PILPS phase 2d. However, *Slater et al.* [2001] argued that the corrected snowfall rate may have been overestimated. Either undercatch or overcatch would result in precipitation errors, affecting the simulations of monthly evapotranspiration, runoff and snow water equivalent.

[18] The other important forcing errors come from incoming solar radiation and longwave radiation, especially longwave radiation because the data obtained from the Valdai site did not include any incoming longwave measurements. Brutsaert's algorithm [*Brutsaert*, 1982] was used to derive incoming longwave radiation. The derived incoming longwave radiation was used as a default radiation forcing in PILPS phase 2d. In addition, other algorithms such as those of *Idso* [1981], *Satterlund* [1979], *Monteith* [1973], and *Brutsaert* [1975], were also used to derive the incoming longwave radiation. Among these schemes,

the *Idso* [1981] and *Brutsaert* [1975] schemes spanned the range of disparity for the estimated incoming longwave radiation. The range between the two schemes is about ± 10% during the winter and ±5% during the summer [*Schlosser et al.*, 2000]. A similar uncertainty range exists for the incoming solar radiation during the summer [*Schlosser et al.*, 1997], and a smaller uncertainty range exists in the winter. The sensitivity tests of PILPS phase 2d have shown that incoming longwave radiation has a significant effect on the hydrological simulations at the Valdai site [*Schlosser et al.*, 2000; *Slater et al.*, 2001].

2.6. Error Functions

[19] The error function is defined as the ratio of the variance of the errors to the variance of the observations. It represents the mismatch between observations and model simulations for the BSI calibration. It is defined as

$$EF = \frac{\sum_{n=1}^{N} (obs_n - sim_n)^2}{\sum_{n=1}^{N} \left(obs_n - \overline{obs}\right)^2}, \quad (6)$$

where N is number of observational data points (i.e. number of months or number of days), obs_n is the observed data, sim_n is the simulated data, and $\overline{obs}$ is mean value of the observed data. In the following study, this definition is used for monthly evapotranspiration, monthly runoff and daily snow water equivalent, respectively.

3. Experiment Design

[20] Table 1 lists 12 CHASM model parameters and 4 forcing error factors, all with their assumed feasible ranges. The range of forcing factor (see the definitions of *Schlosser et al.* [1997]) is 0.9–1.1 for both incoming solar radiation and incoming longwave radiation. (We multiply the default forcing data by this factor for each time step to form a new set of forcing data.) For winter snowfalls, the range is 0.8–1.2 for January and December and 0.9–1.1 for February, March and November. It should be noted that precipitation data that were not modified by *Schlosser et al.* [1997] were used in the second experiment.

[21] In order to reduce the computing burden, we first used a traditional perturbation method (one factor at a time) as used by *Jackson et al.* [2003] to make an error profile analysis, to select sensitive parameters, and to remove insensitive parameters. The sensitivity analysis of 12 model parameters and 4 forcing error factors is shown in Figure 1. Different calibration variables show sensitivities to different parameters because different physical processes affect them. For example, evapotranspiration and runoff are very sensitive to minimum stomatal resistance (RCMIN), but snow water equivalent is not. Comparison of sensitivity tests shows that ALBN and LWR are important parameters for all three calibrated variables. The model results are also sensitive to other model parameters such as VEGS, WRMAX, and Z0V and forcing variables such as SLR, RAS1 and RAS2. None of the three calibrated variables were sensitive to ALBG, ALBV, LEFM, LEFS, VEGM, RHON and Z0G, so these were assigned their default values. Except for these parameters assigned to default values, the other parameters were allowed to vary within the specified ranges.

[22] Although complicated sensitivity analysis techniques such as variational (adjoint) methods [*Skaggs and Barry*, 1996; *Margulis and Entekhabi*, 2001], factorial methods [*Henderson-Sellers*, 1993], Fourier amplitude sensitivity tests [*Collins and Avissar*, 1994], multicriteria methods [*Bastidas et al.*, 1999], reduced form model [*Lynch et al.*, 2001], and response surface methods [*Niyogi et al.*, 2002], can be used to select model parameters and forcing error factors, we decide to use this simple error profile technique to select parameters because it has been used in previous studies [*Wilson et al.*, 1987; *Bonan et al.*, 1993; *Pitman*, 1994; *Alapaty et al.*, 1997; *Jackson et al.*, 2003; *Xia et al.*, 2004b]. The sensitivity analysis also shows that these sensitive parameters are consistent with previous results such as PILPS 2d [*Schlosser et al.*, 2000; *Slater et al.*, 2001] and a tundra region experiment in high latitude [*Lynch et al.*, 2001; *Beringer et al.*, 2002].

[23] Figure 2 is a simple schematic diagram that represents the impacts of the major sensitive model parameters and forcing error factors on simulations of monthly evapotranspiration, monthly runoff and daily snow water equivalent. Simulation of snow water equivalent is mainly determined by snowfall errors, incoming solar and longwave radiation errors, and snow albedo in the winter. In the summer, minimum stomatal resistance, incoming solar and longwave radiation errors mainly control the simulation of monthly evapotranspiration. The physical factors controlling runoff simulation may be different for the snow and snow-free seasons. For the snow season, snowfall errors, incoming solar radiation error, incoming longwave radiation error and snow albedo determine simulations of monthly runoff, whereas for the snow-free season, the minimum stomatal resistance, incoming solar radiation error, and incoming longwave radiation error may be major factors that control simulation of monthly runoff. Besides these factors, vegetation fractional cover seasonality, maximum water holding capacity and vegetation roughness length may also generate simulation errors for these three calibration variables.

[24] The performance of the CHASM model is assessed using the root mean square error, bias, PPDs of model parameters and forcing error factors. In addition, a comparison of the observed and simulated data and correlation matrices are also used to evaluate the performance of the CHASM model.

4. Results

4.1. Comparison of Simulations and Observations

[25] Figure 3 shows the ability of the CHASM model to reproduce the observed monthly evapotranspiration and runoff, and daily snow water equivalent. The simulations are generated using the optimal parameters derived by the Bayesian stochastic inversion and available calibration data (e.g., 17-year monthly runoff and daily snow water equivalent, 8-year monthly evapotranspiration) and the default parameter set provided in the PILPS 2d experiment. The results show that the default parameter set gives simulation results consistent with the observations. However, the optimal simulations agree better with the observations. The performances of the optimal parameter sets are superior

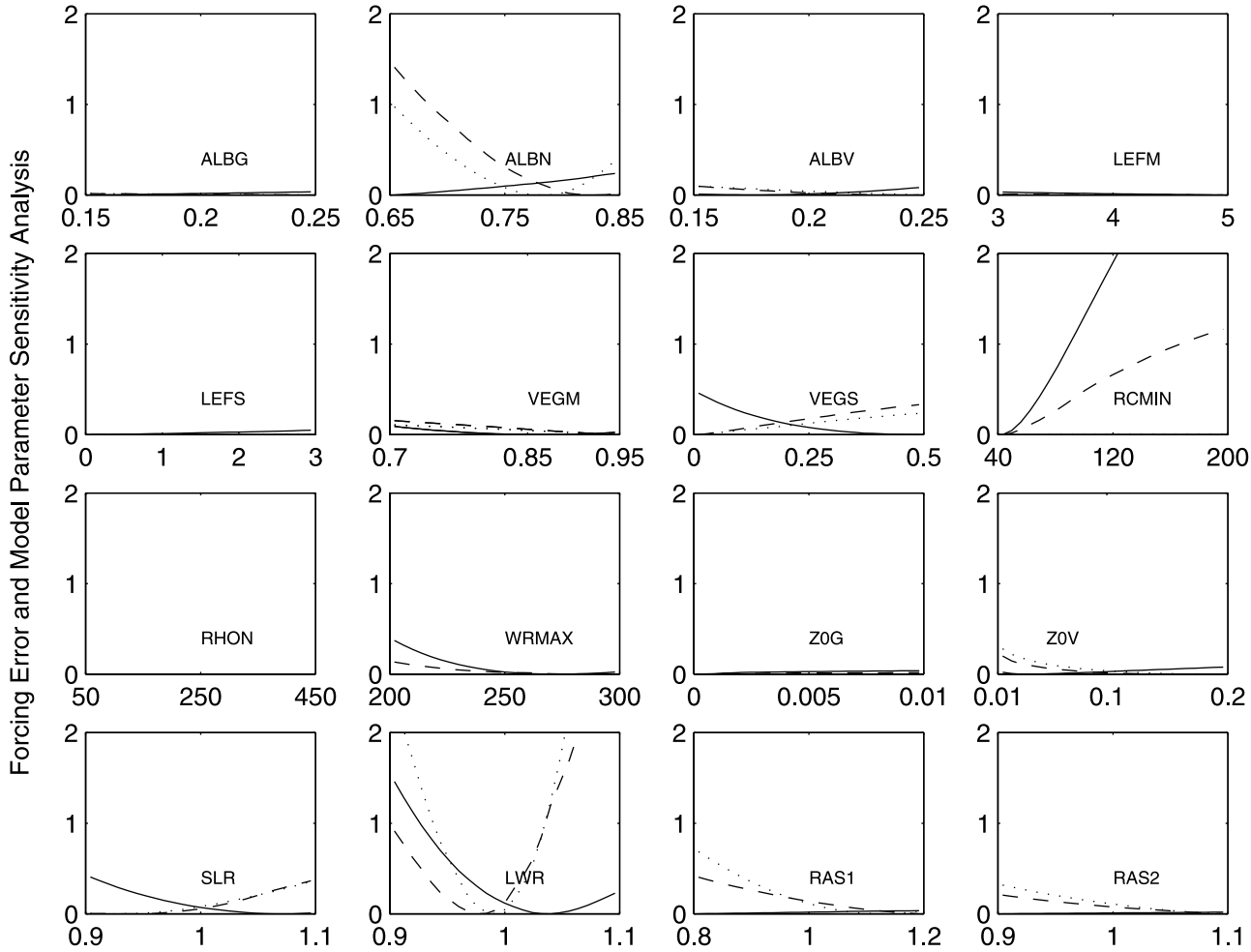

Figure 1. Sensitivity analysis of 12 CHASM parameters and 4 forcing error factors for monthly evaporation (solid), runoff (dashed), and daily snow water equivalent (dotted). Y-axis values were computed as a ratio of the difference between calculated error values and the minimum error value to the minimum error value. The minimum error value is the minimum value of all the calculated errors for each hydrological variable.

to the default parameter set, as indicated by their smaller RMSE and larger correlation coefficient, when the forcing errors were optimized (Figure 3e). Overall, the results show that the CHASM model is a reasonable tool to simulate the monthly evapotranspiration, monthly runoff and daily snow water equivalent.

4.2. Impact of Forcing Errors on Optimal Parameters and Simulations

[26] Table 2 shows the default values used in the PILPS 2d experiment for the CHASM model and the optimal parameter values selected using Bayesian stochastic inversion when monthly evapotranspiration, monthly runoff and daily snow water equivalent were calibrated. The results show that the optimal parameters are similar to the default parameters and that the forcing errors have only small additional effects on the estimation of the five sensitive model parameters when monthly evapotranspiration and runoff were used to minimize error function values. But, the optimal forcing data are different when the model is calibrated by the evapotranspiration data. Forcing errors have significant impacts on the estimation of model parameters (i.e., snow albedo) when the daily snow water equivalent is calibrated. Figure 4 shows a cross-validation test for the optimal model parameters. That is, we pair five optimal model parameters (see Table 2) derived with the consideration of forcing errors with the original forcing data (Figure 4c), and we pair five optimal model parameters derived using fixed forcing with optimal forcing data (Figure 4d) into two separate simulations. The validation results show that the simulated evapotranspiration and runoff are consistent with the observations because the comparison of Figures 3 and 4 show very similar results. This is consistent with our optimal parameter analysis because the two forcing data sets generate similar optimal model parameters when monthly evapotranspiration and runoff were used as the calibrated variables. However, when the optimal model parameters were derived using forcing errors and fixed forcing was used to simulate snow water equivalent, its bias is changed from -1.7 mm (Figure 3d) to -18.9 mm (Figure 4d), RMSE is increased from 21.9 mm (Figure 3d) to 35.1 mm (Figure 4d), and r is reduced from 0.89 (Figure 3d) to 0.79 (Figure 4d). When optimal model parameters derived from fixed forcing and optimal forcing

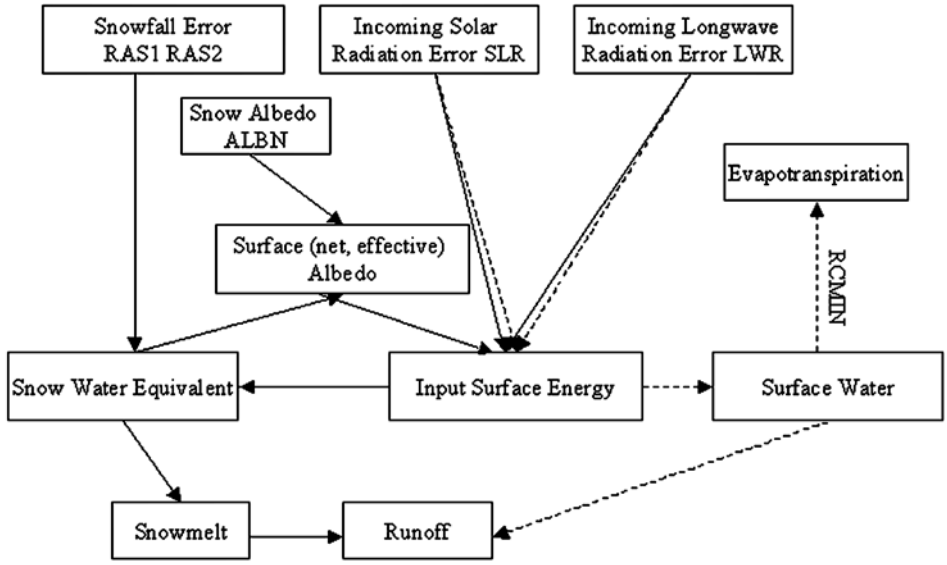

Figure 2. A simple schematic diagram representing impacts of sensitive model parameters and forcing errors (solid line represents snow seasons, and dashed line represents snow-free seasons).

were used to simulate snow water equivalent, its bias is changed from 6.2 mm (Figure 3e) to 20.6 mm (Figure 4c), RMSE is increased from 18.5 mm (Figure 3e) to 29.3 mm (Figure 4c), and r is reduced from 0.93 (Figure 3e) to 0.89 (Figure 4c). This significant change is a result of different optimal parameter sets, particularly snow albedo. Therefore, forcing errors have their most significant impact on the daily snow water equivalent simulations.

[27] The simulations for daily evapotranspiration, runoff, snowmelt and snow water equivalent are compared

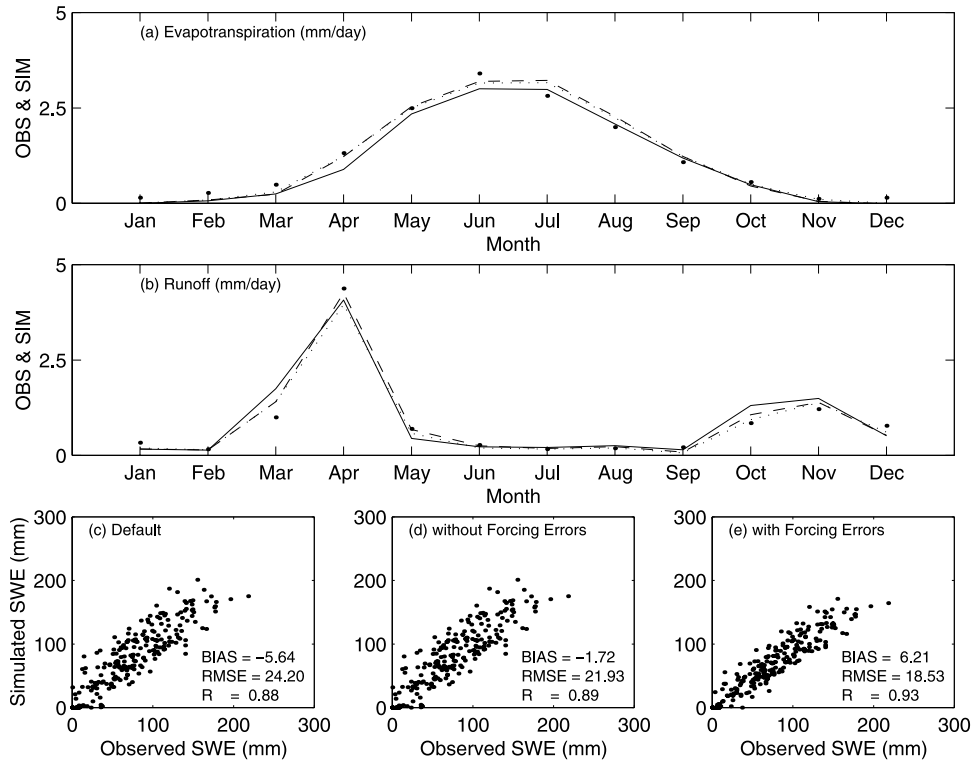

Figure 3. Observed and simulated (a) monthly evapotranspiration, (b) monthly runoff, (c) default daily snow water equivalent, (d) optimal simulations of daily snow water equivalent when fixed forcing was used, and (e) optimal simulations of daily snow water equivalent when forcing errors were used. (For Figures 3a and 3b, dot is observations, solid is default simulations, dashed is optimal simulations when fixed forcing was used, and dotted is optimal simulations when forcing factors were optimized.)

Table 2. Default Parameter Set Used in PILPS 2d Experiment and Optimal Parameter Sets Generated Using Bayesian Stochastic Inversion[a]

Calibration Variables		E		R		SWE	
Parameter	Default	Fixed Forcing	Optimized Forcing	Fixed Forcing	Optimized Forcing	Fixed Forcing	Optimized Forcing
ALBN	**0.75**	0.65	*0.66*	0.77	*0.85*	0.65	*0.85*
VEGS	**0.25**	0.50	*0.50*	0.00	*0.00*	0.00	*0.00*
RCMIN	**50**	57	*48*	40	*40*	197	*93*
WRMAX	**206**	215	*201*	206	*202*	*222*	*251*
Z0V	**0.10**	0.01	*0.01*	0.2	*0.18*	0.17	*0.18*
SLR	**1.00**	**1.00**	*0.90*	**1.00**	*0.96*	**1.00**	*1.08*
LWR	**1.00**	**1.00**	*1.02*	**1.00**	*1.03*	**1.00**	*1.05*
RAS1	**1.20**	**1.20**	*0.81*	**1.20**	*1.20*	**1.20**	*1.20*
RAS2	**1.10**	**1.10**	*0.90*	**1.10**	*1.09*	**1.10**	*1.10*

[a]The values represented in bold are not optimized; the optimal values with forcing error are represented in italics; E, evapotranspiration; R, runoff; SWE, snow water equivalent.

in Figures 5a–5c. Here optimal evapotranspiration (Figure 3a), optimal runoff (Figure 3b) and optimal snow water equivalent (Figure 3e) were used as comparisons (dotted). Optimal parameters estimated using monthly evapotranspiration give consistent simulations of daily evapotranspiration for both forcing data sets. However, two optimal parameter sets generate early snowmelt and runoff (Figure 5a) when compared to the optimal simulation. This different simulation is due to the use of a small snow albedo. Because the snow albedo is smaller, less incoming solar radiation is reflected into the air, and more solar energy is absorbed by the land surface, resulting in an increase in the land surface temperature. Increased surface temperature makes snow melt early, resulting in an early runoff peak. Therefore, monthly evapotranspiration is not a good calibration variable for the selection of the optimal snow albedo. This is consistent with our sensitivity analysis because snow albedo is not sensitive to evapotranspiration (Figure 1). Optimal parameters derived using daily snow water equivalent give good snowmelt and snow water equivalent simulations as well as winter runoff simulation. However, they give poor simulations for evapotranspiration and summer runoff. The reason for the poor simulations is that the optimal minimum stomatal resistance is too large

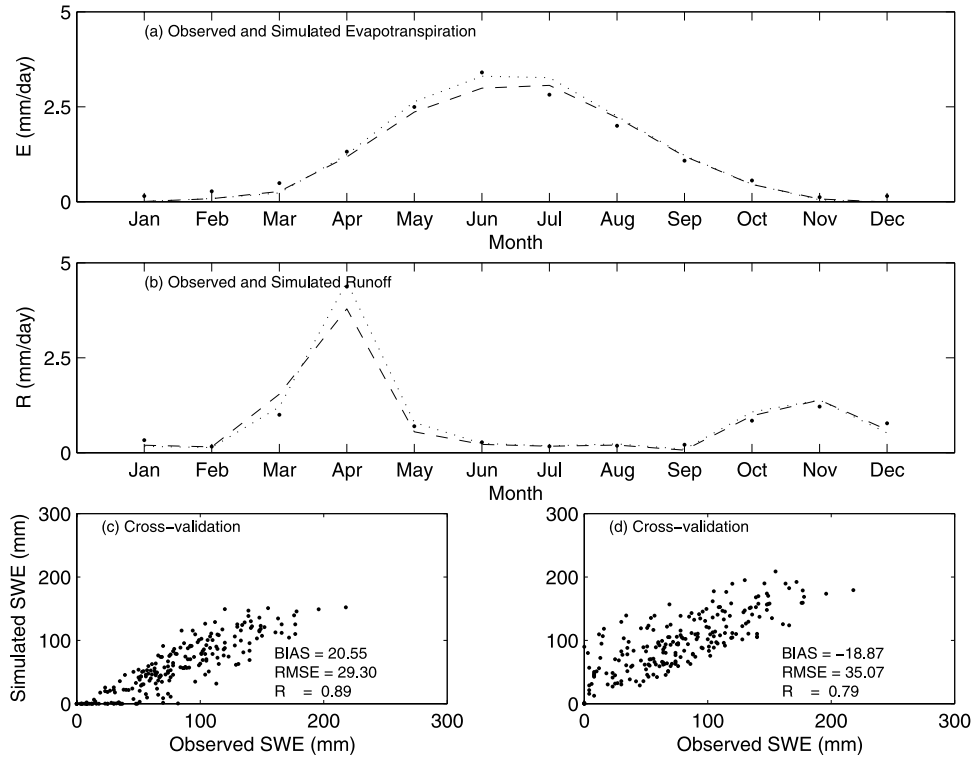

Figure 4. Observed and simulated (a) monthly evapotranspiration, (b) runoff, (c) cross-validation of daily snow water equivalent when optimal model parameters derived from fixed forcing and optimized forcing was used or when optimal model parameters derived from forcing error factors and fixed forcing was used, and (d) cross-validation of daily snow water equivalent when optimal model parameters derived from forcing errors and fixed forcing were used (dot is observations, dashed is simulations when optimal model parameters derived from fixed forcing and optimal forcing were used, and dotted is simulations when optimal model parameters derived from forcing errors and fixed forcing were used).

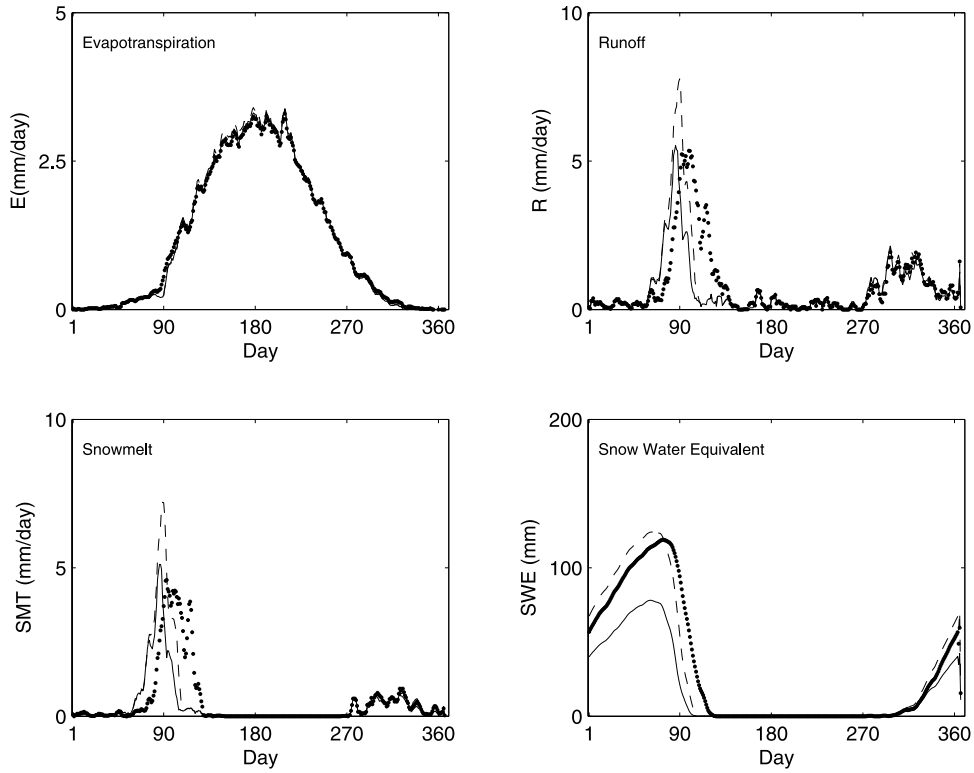

Figure 5a. The 17-year average (1967–1983) of optimal simulations for daily evapotranspiration (E), runoff (R), snowmelt (SMT), and snow water equivalent (SWE) when monthly E was used as a calibration variable (dotted is optimal simulations for calibration variables, solid is simulations using the optimal parameters derived by forcing with errors, and dashed is simulations using the optimal parameters derived by fixed forcing).

and leads to a small evapotranspiration simulation and large runoff simulation in the summer (Figure 5c). Minimum stomatal resistance is not searched well using VFSA when the daily snow water equivalent is used as a calibration variable because it is not sensitive to daily snow water equivalent. In addition, the simulations of evapotranspiration and runoff are different for the fixed and optimized forcing runs because these runs produce different values of minimum stomatal resistance (Figure 5c). Comparisons of the snowmelt for Figures 5a and 5c show that small snow albedo causes early snowmelt for the evapotranspiration case compared to the snow water equivalent case. This is due to the use of different VEGS and Z0V in two optimal parameter sets. Optimal parameters derived using monthly runoff give relatively good simulations for evapotranspiration, runoff, snowmelt and snow water equivalent (Figure 5b), although differences exist for the spring snowmelt period because of different ALBN, VEGS, and Z0V (Table 1). In addition, the forcing errors affect the snow water equivalent simulations. However, from this study we know runoff is a more appropriate calibration variable for a cold catchment.

4.3. Impacts of Forcing Errors on PPDS of Model Parameters

[28] Figure 6 shows the marginal posterior probability density distributions for the three most sensitive parameters: snow albedo (ALBN), fractional vegetation cover seasonality (VEGS) and minimum stomatal resistance (RCMIN). For evapotranspiration, the forcing errors (solid line) result in wider PPDs, which means larger uncertainty when compared to the fixed forcing case (dashed line). For the runoff case, the forcing errors not only influence the uncertainty range of ALBN and RCMIN but also influence the shape of the PPD's distribution for ALBN. The same conclusion can be drawn for ALBN for the snow water equivalent case (note that VEGS and RCMIN are moderately sensitive for the snow water equivalent case (see Figure 1)). Forcing errors result in quite different PPDs, that is, the snow albedo favors small values for the fixed forcing case, and it favors large values for the case of forcing with error.

[29] Therefore, forcing errors indeed influence the width and shape of the PPD's for the most important model parameters. This is a result of the model parameters and forcing errors interdependency (correlation of model parameters and forcing errors). A schematic diagram derived from the correlation matrices calculated using Bayesian stochastic inversion is shown in Figure 7. Values in the figure represent the correlation coefficients between the model parameters and forcing errors. Large correlations can be seen in Figures 7b and 7c, for example, 0.75 for LWR and RAS2, 0.65 for LWR and ALBN, 0.61 for LWR and RAS1 when the daily snow water equivalent was calibrated. 0.65 is for LWR and ALBN when the monthly runoff was calibrated. This large correlation indicates the existence of dependencies between forcing errors (e.g., LWR) and model parameters (e.g., ALBN) as well as dependencies between forcing errors. The dependencies between model parameters

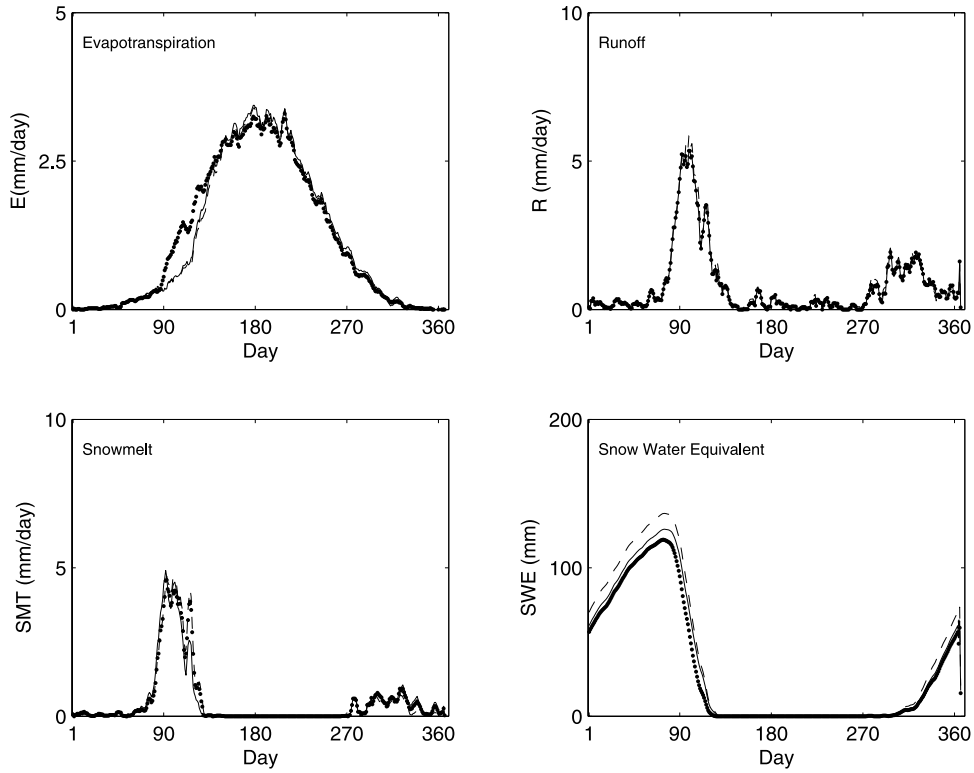

Figure 5b. Same as Figure 5a when monthly R was used as a calibration variable.

also exist in Figure 6. For the evapotranspiration case, LWR first affects VEGS. Interaction among VEGS, RCMIN and ALBN results in wider PPDs for these three model parameters (see Figure 6). For the snow water equivalent case, all forcing errors except for SLR affect the snow albedo so that this effect results in totally different snow albedo PPD. As the incoming longwave radiation increases (this would cause more snowmelt because the land surface gets more

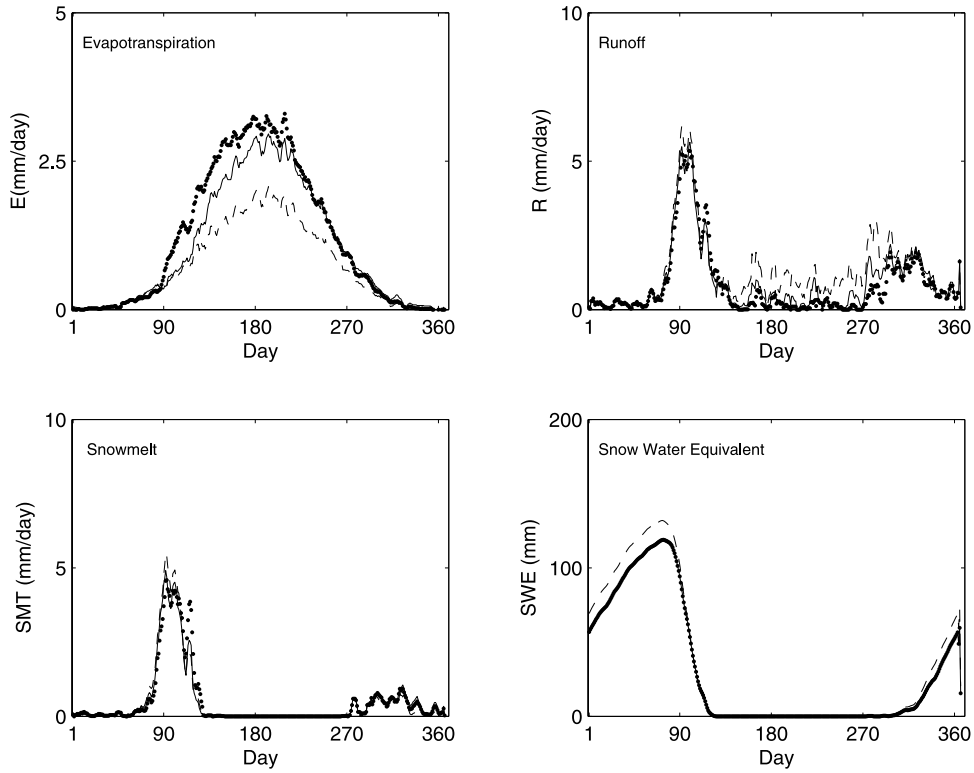

Figure 5c. Same as Figure 5a when daily SWE was used as a calibration variable.

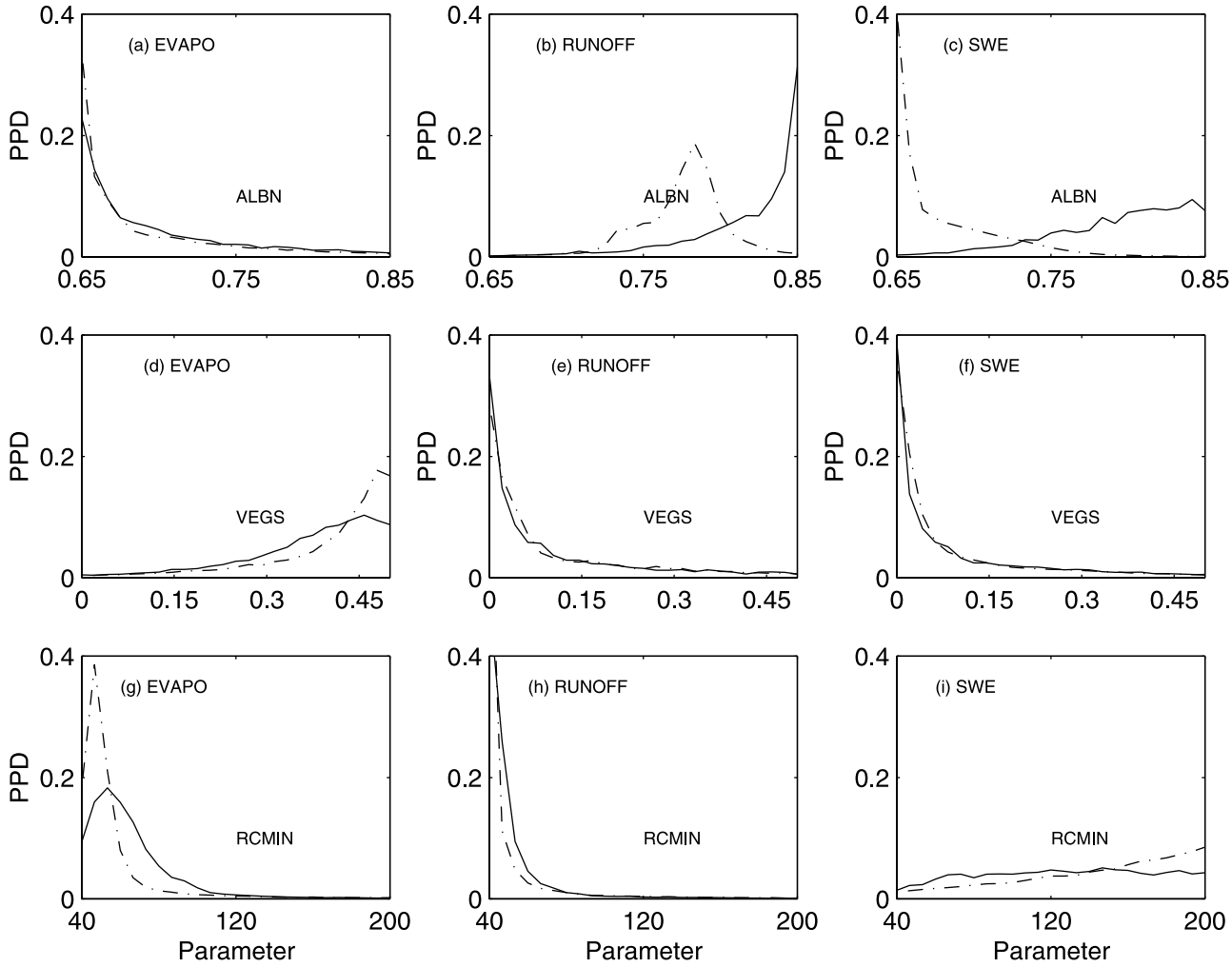

Figure 6. Calculated PPDs for the three most sensitive model parameters when observed monthly evapotranspiration, monthly runoff, and daily snow water equivalent were used as calibration variables (solid is fixed forcing, dash-dotted is forcing with errors, EVAPO is evapotranspiration, and SWE is snow water equivalent).

incoming energy), the snow albedo should increase to reduce the incoming surface energy and to control snowmelt rate. When we compare Table 2 and Figure 6 for the fixed forcing and forcing with error, we find that larger snow albedo exists for the latter because of the larger incoming longwave radiation. This is consistent with our analysis here. The combination of the evapotranspiration and snow water equivalent case forms the runoff case (see Figure 7c). This result echoes the sensitivity analysis of Figures 1 and 2 because the runoff simulations are determined by snowmelt and evapotranspiration. Therefore, the forcing errors affect snow albedo, and further affect snowmelt for the snow season. For snow-free seasons they affect VEGS and RCMIN, and further affect evapotranspiration. As a result, changes in either the snowmelt or evapotranspiration will result in changes the runoff simulations.

[30] Comparisons of Figures 7a and 7c as well as Figures 7b and 7c show that correlations between the forcing errors and model parameters (those between model parameters and those between forcing errors) are similar except that the correlation coefficient between VEGS and RCMIN is changed from a negative to positive value. This means that the relationships between the forcing errors and the model parameters are relatively stable. If we compare Figure 7 with Figures 1 and 2, we believe this schematic diagram may be reasonable. For example, larger incoming longwave radiation implies larger snow albedo for the conservation of surface energy (see Figure 6b). However, it should be noted that some model parameters and forcing errors may be correlated although we cannot give a physical explanation.

4.4. Constraint of Hydrological Variables on Forcing Errors

[31] Errors in forcing data can be optimized and their uncertainties can be estimated using Bayesian stochastic inversion. Figure 8 shows the marginal posterior probability densities (PPD) for SLR, LWR, RAS1, and RAS2 when different hydrological variables (i.e., evapotranspiration, runoff, snow water equivalent) were used to constrain the errors between observations and simulations. The results show that all hydrological variables give good constraints for LWR although the peaks of the PPDs are different. Most

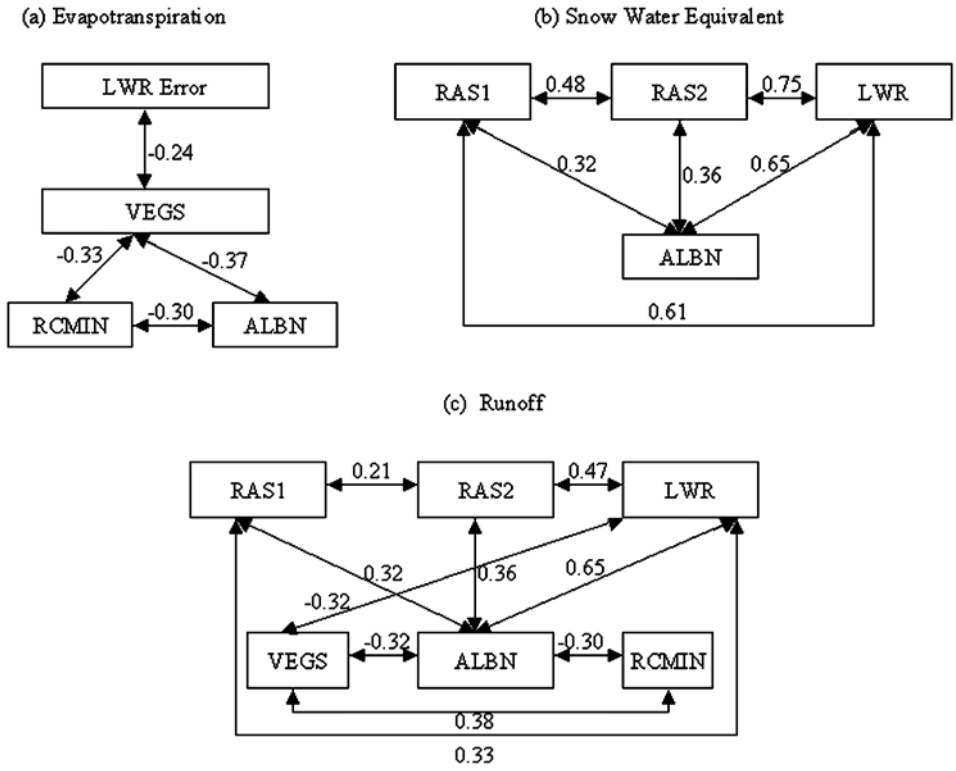

Figure 7. A schematic diagram derived from correlation matrices calculated using Bayesian stochastic inversion and very fast simulated annealing when (a) monthly evapotranspiration, (b) daily snow water equivalent, and (c) monthly runoff were calibrated (values in this figure represent correlation coefficients; total parameters set is over 50,000 for each experiment).

likely values exist within 0.98 to 1.05, which is consistent with the value of 1.0 used in the PILPS 2d experiment. In addition, optimal values are very close to each other (i.e., 1.02, 1.03, 1.05) for the three calibrations. Snow water equivalent and runoff give strong constraints on winter snowfalls although the constraints are much stronger on RAS1 than RAS2. Evapotranspiration gives only a weak constraint on winter snowfalls. All hydrological variables give weak constraints on SLR because their PPDs are almost uniform, indicating large uncertainties. Comparisons of Figures 8 and 1 show that the most sensitive parameters such as LWR, RAS1, ALBN, and RCMIN can be well constrained by using the observed hydrological variables.

[32] Different hydrological variables show different sensitivities to forcing errors. Evapotranspiration, runoff and snow water equivalent are sensitive to incoming longwave radiation, so these hydrological variables may be used to constrain the incoming longwave radiation. However, they may not constrain the incoming shortwave radiation. Snow water equivalent and runoff are sensitive to winter snowfall, so both variables may be used to constrain the winter snowfall forcing. Evapotranspiration is small in the winter, so it cannot be used to constrain the winter snowfall forcing. The optimal rate of winter snowfall obtained in this study appears to be consistent with the estimates in *Schlosser et al.* [1997], who suggested that the gauged snowfall at Valdai is subject to a 20% undercatch for December and January snowfalls, and a 10% undercatch for November, February and March snowfalls. Our results do not support *Slater et al.* [2001], who asserted the overcatch of winter snowfalls at Valdai.

4.5. Discussion

[33] This study was performed using a single criterion method (e.g., one calibration variable at a time) instead of a multicriteria method [*Gupta et al.*, 1998, 1999]. Usually, the optimal parameter set derived using a single criterion method (e.g., a variable) is difficult to use to simulate another variable because this may result in poor simulation for another variable [see *Leplastrier et al.*, 2002]. However, our cross-validation shows that the optimal model parameter set derived using monthly runoff can simulate evapotranspiration and snow water equivalent well. However, the optimal model set derived using monthly evapotranspiration cannot simulate runoff and snow water equivalent well, and the optimal parameter set derived using the daily snow water equivalent is also unable to simulate evaporation and runoff well. The reason is that the runoff process comprises snowmelt (snow water equivalent) and evapotranspiration processes, and therefore the runoff calibration indirectly contains the snow water equivalent calibration and evapotranspiration calibration. However, there is no direct relationship between snowmelt and evapotranspiration processes, and the important parameters that control the two processes are different (see Figure 1). As a result, optimal parameter values such as optimal RCMIN and VEGS that derived by evapotranspiration are quite different from optimal parameter values derived by snow water

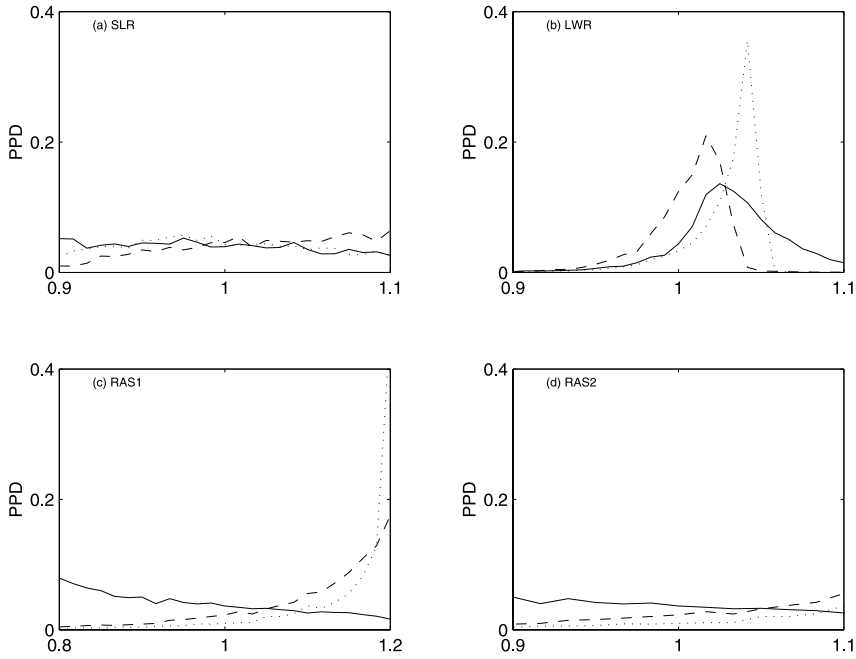

Figure 8. Calculated PPDs for four forcing error factors (a) SLR, (b) LWR, (c) RAS1, and (d) RAS2 when observed monthly evapotranspiration, monthly runoff, and daily snow water equivalent were used as the calibration variables (solid is evapotranspiration, dashed is runoff, and dotted is snow water equivalent).

equivalent. The cross-validation of these different optimal parameter values result in different simulations.

[34] Based on our experiments, runoff appears to be an appropriate variable for model optimization at a cold catchment when water flux is analyzed. When an appropriate variable is difficult to choose, a multicriteria method [*Gupta et al.*, 1999] may be a more appropriate choice if at least two hydrological variables are measured. For the study of regional water fluxes [*Lohmann et al.*, 2004], macroscale water fluxes [*Milly and Dunne*, 2002a] and global modeling of land water and energy balances, precipitation error is a serious limit for evaluating runoff simulation because precipitation contains several errors including gauge errors, spatial sampling errors, spatial interpolation errors, and topographic spatial sampling errors as well as their interactions. As indicated by *Milly and Dunne* [2002a], a 10% to 20% bias in precipitation is typical among the basins they used, and in some cases, the bias in precipitation is 50% or larger. These errors may produce a 100% error in runoff simulation [*Milly and Dunne*, 2002a]. In the case of inaccurate LWR, runoff can be used to constrain the incoming longwave radiation and winter snowfalls by Bayesian stochastic inversion.

[35] Both single and multiple criteria methods have their benefits and limits for calibrating land surface models. The benefit of the multiple criteria method is that it can select more reasonably optimal parameter sets if calibrated variables such as sensible heat fluxes, latent heat fluxes, evapotranspiration, soil moisture, runoff, and soil temperature are available. (It should also be noted that a selection of different combinations of the calibrated variables is needed because some combinations may result in poor simulations [*Gupta et al.*, 1999].) As indicated by *Leplastrier et al.* [2002], the use of a single hydrological variable or energy flux may result in the reduction of errors in the calibrated variable and an increase in the errors of the other variables. Through multiple criteria calibration, the physical mechanism or relationships between model parameters found from calibrating land surface models may be more reliable than single criterion calibration because more observed data are used to constrain land surface models. These benefits are also the limits of the single criterion method. It should be noted that these benefits are dependent on the available measurement data. At the current time, these data are difficult to measure for GCM grid cells, small to large basins, on a national scope or a global scope except for at some sites [see *Xia et al.*, 2004a], and therefore, lack of available measurement data limit the use of the multiple criteria method. Therefore, the single criterion method had to be used to calibrate land surface models. (For example, streamflow from basins are widely measured in the world, particularly, the U.S. Geological Survey has archived high-quality streamflow data for both the U.S. and on a global scope.) This study is a basis for calibrating land surface model for regional (e.g., U.S.) or global simulation, given that reliable observations are not widely available.

[36] Finally, it should be noted that we did not use soil moisture as a calibration variable although monthly soil moisture at the root zone are available at Valdai. The reason is that we have previously conducted a similar study for soil moisture [*Xia et al.*, 2004c] although forcing data errors were not included in that study.

5. Conclusions

[37] The primary goal of this study is to discuss the effects of forcing errors on the estimates of optimal parameters and uncertainty of the CHASM model parameters. The paper has

shown that the optimal parameters give relatively accurate simulations of evapotranspiration, runoff and snow water equivalent. The BSI method also gives a relatively reasonable uncertainty estimate by considering the marginal posterior probability densities. We have shown that forcing errors have little effect on the estimation of the optimal model parameters when monthly evapotranspiration and runoff were calibrated. However, forcing errors do have significant effects on the estimation of the optimal model parameters when the daily snow water equivalent was calibrated. Forcing errors also affect the uncertainty estimates of the CHASM model parameters. Forcing errors result in large uncertainties of model parameters and totally different PPDs in some cases. These uncertainties are a result of interactions between the forcing errors and model parameters, interactions between model parameters, and interactions between forcing error factors. Therefore, all these interactions contribute to uncertainties in the simulations of evapotranspiration, runoff and snow water equivalent.

[38] All calibrated hydrological variables are sensitive to incoming longwave radiation errors. Snow water equivalent and runoff are sensitive to winter snowfalls errors. Therefore, they are well constrained by Bayesian stochastic inversion. However, all the hydrological variables cannot constrain incoming solar radiation error well. Evapotranspiration cannot constrain winter snowfall errors because of small evaporation in winter. Optimal winter snowfalls are relatively consistent with the results of *Schlosser et al.* [1997], and our results suggest a low possibility for overcatch of winter snowfalls at Valdai as suggested by *Slater et al.* [2001].

[39] **Acknowledgments.** The authors wish to thank A. J. Pitman for providing us the CHASM land surface model. Y.X. wishes to acknowledge the financial support from NOAA grant NA17RJ2612. Z.L.Y. was supported under NASA grant NAG5-12577 and NOAA grant NA03OAR4310076. Y.X. appreciates the helpful reviews provided from P. C. D. Milly, S. Fan, and two anonymous referees.

References

Alapaty, K., S. Raman, and D. S. Niyogi (1997), Uncertainty in specification of surface characteristics: A study of prediction errors in the boundary layer, *Boundary Layer Meteorol.*, *82*, 473–500.

Bastidas, L. A., H. V. Gupta, S. Sorooshian, W. J. Shuttleworth, and Z. L. Yang (1999), Sensitivity analysis of a land surface scheme using multicriteria methods, *J. Geophys. Res.*, *104*(D16), 19,481–19,490.

Beringer, J., S. McIlwaine, A. H. Lynch, F. S. Chapin III, and G. B. Bonan (2002), The use of a reduced form model to assess the sensitivity of a land surface model to biotic surface parameters, *Clim. Dyn.*, *19*, 455–466.

Bonan, G. B., D. Pollard, and S. L. Thompson (1993), Influence of subgrid-scale heterogeneity in leaf area index, stomatal resistance, and soil moisture on grid-scale land-atmosphere interactions, *J. Clim.*, *6*, 1882–1897.

Bowling, L. C., et al. (2003), Simulation of high latitude hydrological processes in the Torne-Kalix basin: PILPS Phase 2 (e), 1, Experiment description and summary intercomparisons, *Global Planet. Change*, *38*, 1–30.

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

Brutsaert, W. H. (1982), *Evaporation Into Atmosphere*, 299 pp., Springer, New York.

Budyko, M. I. (1956), Heat balance of the Earth's surface (in Russian), *Gidrometeoizlat*, 255 pp.

Chen, T. H., et al. (1997), Cabauw experimental results from the Project for Intercomparison of Land-surface Parameterization Schemes, *J. Clim.*, *10*, 1194–1215.

Collins, D. C., and R. Avissar (1994), An evaluation with the Fourier amplitude sensitivity test (FAST) of which land-surface parameters are of greatest importance in atmospheric modeling, *J. Clim.*, *7*, 681–703.

Desborough, C. E. (1999), Surface energy balance complexity in GCM land surface models, *Clim. Dyn.*, *15*, 389–403.

Desborough, C. E., A. J. Pitman, and B. McAvaney (2001), Surface energy balance complexity in GCM land surface models, part II: Coupled simulations, *Clim. Dyn.*, *17*, 615–626.

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.

Franks, S. W., and K. J. Beven (1997), Bayesian estimation of uncertainty in land surface-atmosphere flux predictions, *J. Geophys. Res.*, *102*, 23,991–23,999.

Groisman, P. Y., V. V. Koknaeva, T. A. Belokryloya, and T. R. Karl (1991), Overcoming biases of precipitation measurement: A history of USSR experience, *Bull. Am. Meteorol. Soc.*, *11*, 1725–1732.

Groisman, P. Y., D. R. Easterling, R. G. Quayle, and V. S. Golubev (1996), Reducing biases in estimates of precipitation over the United States: Phases 3 adjustments, *J. Geophys. Res.*, *101*, 7185–7195.

Gupta, H. V., S. Sorooshian, and P. O. Yapo (1998), Towards improved calibration of hydrologic models: multiple and noncommensurable measures of information, *Water Resour. Res.*, *34*, 751–763.

Gupta, H. V., L. A. Bastidas, S. Sorooshian, S. W. J. Shuttleworth, and Z.-L. Yang (1999), Parameter estimation of a land surface scheme using multicriteria methods, *J. Geophys. Res.*, *104*, 19,491–19,503.

Henderson-Sellers, A. (1993), A factorial assessment of the sensitivity of the BATS land-surface parameterization scheme, *J. Clim.*, *6*, 227–247.

Henderson-Sellers, A. (1996), Soil moisture simulation: Achievements of the RICE and PILPS intercomparison workshop and future directions, *Global Planet. Change*, *13*, 99–115.

Henderson-Sellers, A., K. McGuffie, and A. Pitman (1996), The Project for Intercomparison of Land-surface Parameterization Schemes (PILPS): 1992–1995, *Clim. Dyn.*, *12*, 849–859.

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

Ingber, L. (1989), Very fast simulated reannealing, *Math. Comput. Model.*, *12*, 967–993.

Jackson, C., Y. Xia, M. K. Sen, and P. Stoffa (2003), Optimal parameter estimation and uncertainty analysis of a land surface model: A case study from Cabauw, Netherlands, *J. Geophys. Res.*, *108*(D18), 4583, doi:10.1029/2002JD002991.

Koster, R. D., and M. J. Suarez (1992), Modeling the land surface boundary in climate models as a composite of independent vegetation stands, *J. Geophys. Res.*, *97*, 2697–2715.

Leplastrier, M., A. J. Pitman, H. Gupta, and Y. Xia (2002), Exploring the relationship between complexity and performance in a land surface model using the multicriteria method, *J. Geophys. Res.*, *107*(D20), 4443, doi:10.1029/2001JD000931.

Lohmann, D., et al. (2004), Streamflow and water balance intercomparisons of four land-surface models in the North American Land Data Assimilation System project, *J. Geophys. Res.*, *109*, D07S91, doi:10.1029/2003JD003517.

Luo, L., et al. (2003), Effects of frozen soil on soil temperature, spring infiltration and runoff: Results from the PILPS 2 (d) experiment at Valdai, Russia, *J. Hydrometeorol.*, *4*, 334–351.

Lynch, A. H., S. McIlwaine, J. Beringer, and G. Bonan (2001), An investigation of the sensitivity of a land surface model to climate change using a reduced form model, *Clim. Dyn.*, *17*, 643–652.

Manabe, S. (1969), Climate and the ocean circulation: 1, The atmospheric circulation and the hydrology of the earth's surface, *Mon. Weather Rev.*, *97*, 739–805.

Margulis, S. A., and D. Entekhabi (2001), Feedback between the landsurface energy balance and atmospheric boundary layer diagnosed through a model and its adjoint, *J. Hydrometeorol.*, *2*, 599–619.

Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller (1953), Equation of state calculations by fast computing machines, *J. Chem. Phys.*, *21*, 1087–1092.

Milly, P. C. D., and K. A. Dunne (2002a), Macroscale water fluxes: 1. Quantifying errors in the estimation of basin mean precipitation, *Water Resour. Res.*, *38*(10), 1205, doi:10.1029/2001WR000759.

Milly, P. C. D., and K. A. Dunne (2002b), Macroscale water fluxes: 2. Water and energy supply control of their interannual variability, *Water Resour. Res.*, *38*(10), 1206, doi:10.1029/2001WR000760.

Milly, P. C. D., and A. B. Shmakin (2002a), Global modeling of land water and energy balances. part I: The land dynamics (LaD) model, *J. Hydrometeorol.*, *3*, 283–299.

Milly, P. C. D., and A. B. Shmakin (2002b), Global modeling of land water and energy balances. part II: Land-characteristic contributions to spatial variability, *J. Hydrometeorol.*, *3*, 301–310.

Monteith, J. L. (1973), *Principles of Environmental Physics*, 241 pp., Edward Arnold, London.

Niyogi, D. S., Y. K. Xue, and S. Raman (2002), Hydrological land surface response in a tropical regime and a midlatitudinal regime, *J. Hydrometeorol.*, *3*, 39–56.

Pan, M., et al. (2003), Snow process modeling in the North American Land Data Assimilation System (NLDAS): 2. Evaluation of model simulated snow water equivalent, *J. Geophys. Res.*, *108*(D22), 8850, doi:10.1029/2003JD003994.

Pitman, A. J. (1994), Assessing the sensitivity of a land-surface scheme to the parameter values using a single column model, *J. Clim.*, *7*, 1856–1869.

Pitman, A. J., et al. (1999), Key results and implications for phase 1(c) of the Project for Intercomparison of Land-surface Parameterization Schemes, *Clim. Dyn.*, *15*, 673–684.

Pitman, A. J., Y. Xia, M. Leplastrier, and A. Henderson-Sellers (2003), The CHAmeleon Surface Model (CHASM): Description and use with the PILPS Phase 2e forcing data, *Global Planet. Change*, *38*, 121–135.

Robock, A., K. Y. 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. Clim.*, *8*, 15–35.

Satterlund, D. R. (1979), An improved equation for estimating long-wave radiation from the atmosphere, *Water Resour. Res.*, *15*, 1649–1650.

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

Schlosser, C. A., et al. (2000), Simulation of a boreal grassland hydrology at Valdai, Russia: PILPS phase 2(d), *Mon. Weather Rev.*, *128*, 301–321.

Sellers, P. J., et al. (1989), Calibrating the simple biosphere model for Amazonian tropical forest using field and remote-sensing data. 1. Average calibration with field data, *J. Appl. Meteorol.*, *28*, 727–759.

Sellers, P. J., et al. (1997), Modeling the exchanges of energy, water and carbon between continents and the atmosphere, *Science*, *275*, 502–509.

Sen, M. K., and P. L. Stoffa (1996), Bayesian inference, Gibbs' sampler and uncertainty estimation in geophysical inversion, *Geophys. Prospect.*, *44*, 313–350.

Shao, Y. P., and A. Henderson-Sellers (1998), Validation of soil moisture simulation in land surface parameterization schemes with HAPEX data, *Global Planet. Change*, *13*, 11–46.

Skaggs, T. H., and D. A. Barry (1996), Sensitivity methods for time-continuous, spatially discrete groundwater contaminant transport models, *Water Resour. Res.*, *32*, 2409–2420.

Slater, A. G., et al. (2001), The representation of snow in land-surface scheme: Results from PILPS 2(d), *J. Hydrometeorol.*, *2*, 7–25.

Vinnikov, K. Y., 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.

Wilson, M. F., A. Henderson-Sellers, R. E. Dickinson, and P. J. Kennedy (1987), Sensitivity of the Biosphere-Atmosphere Transfer Scheme (BATS) to the inclusion of variable soil characteristics, *J. Clim. Appl. Meteorol.*, *26*, 341–362.

Wood, E. F., et al. (1998), The project for intercomparison of land-surface parameterization schemes (PILPS) phase 2(c) Red–Arkansas River basin experiment: 1. Experiment description and summary intercomparisons, *Global Planet. Change*, *19*, 115–135.

Xia, Y., A. J. Pitman, H. V. Gupta, M. Leplastrier, A. Henderson-Sellers, and L. A. Bastidas (2002), Calibrating a land surface model of varying complexity using multi-criteria methods and the Cabauw data set, *J. Hydrometeorol.*, *3*, 181–194.

Xia, Y., M. K. Sen, C. Jackson, and P. L. Stoffa (2004a), Multi-dataset study of optimal parameter and uncertainty estimation of a land surface model with Bayesian stochastic inversion and multicriteria method, *J. Appl. Meteorol.*, 1477–1497.

Xia, Y., Z.-L. Yang, P. L. Stoffa, and M. K. Sen (2004b), Optimal parameter and uncertainty estimation of a land surface model: sensitivity to parameter ranges and model complexities, *Adv. Atmos. Sci.*, in press.

Xia, Y., Z.-L. Yang, C. Jackson, P. L. Stoffa, and M. K. Sen (2004c), Impacts of data length on optimal parameter and uncertainty estimation of a land surface model, *J. Geophys. Res.*, *109*, D07101, doi:10.1029/2003JD004419.

Yang, D., et al. (1995), Accuracy of Tretyakov precipitation gauge: Results of WMO intercomparison, *Hydrol. Processes*, *9*, 877–895.

Yang, D., B. E. Goodison, C. S. Benson, and S. Ishida (1998), Adjustment of daily precipitation at 10 climate stations in Alaska: Application of WMO intercomparison results, *Water Resour. Res.*, *34*, 241–256.

Yang, Z. L., R. E. Dickinson, A. Robock, and K. Y. Vinnikov (1997), On validation of the snow sub-model of the Biosphere-Atmosphere Transfer Scheme with Russian snow cover and meteorological observational data, *J. Clim.*, *10*, 353–373.

Zhang, H., A. Henderson-Sellers, A. J. Pitman, J. L. McGregor, C. E. Desborough, and J. J. Katzfey (2001), Limited-area model sensitivity to the complexity of representation of the land surface energy balance, *J. Clim.*, *14*, 3965–3986.

M. K. Sen and P. L. Stoffa, Institute for Geophysics, The John A. and Katherine G. Jackson School of Geosciences, University of Texas at Austin, Austin, TX 78759, USA.

Y. Xia, Geophysical Fluid Dynamics Laboratory, NOAA, Princeton University, Princeton, NJ 08542, USA. (youlong.xia@noaa.gov)

Z.-L. Yang, Department of Geological Sciences, The John A. and Katherine G. Jackson School of Geosciences, University of Texas at Austin, Austin, TX 78712, USA.