JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 109, D07101, doi:10.1029/2003JD004419, 2004

Impacts of data length on optimal parameter and uncertainty estimation of a land surface model

Youlong Xia,[1,2] Zong-Liang Yang,[3] Charles Jackson,[1]
Paul L. Stoffa,[1] and Mrinal K. Sen[1]

Received 4 December 2003; revised 9 February 2004; accepted 19 February 2004; published 3 April 2004.

[1] The optimal parameters and uncertainty estimation of land surface models require that appropriate length of forcing and calibration data be selected for computing error functions. Most of the previous studies used less than two years of data to optimize land surface models. In this study, 18-year hydrometeorological data at Valdai, Russia, were used to run the Chameleon Surface Model (CHASM). The optimal parameters were obtained by employing a global optimization technique called very fast simulated annealing. The uncertainties of model parameters were estimated by the Bayesian stochastic inversion technique. Forty-four experiments were conducted by using different lengths of data from the 18-year record, and a total of about 3 million parameter sets were produced. This study found that different calibration variables require different lengths of data to obtain optimal parameters and uncertainty estimates which are insensitive to the period selected. In the case of optimal parameters, monthly root-zone soil moisture, runoff, and evapotranspiration require 8, 3, and 1 years of data, respectively. In the case of uncertainty estimates, monthly root-zone soil moisture, runoff, and evapotranspiration require 8, 8, and 3 years of data, respectively. Spin-up has little impact on the selection of optimal parameters and uncertainty estimates when evapotranspiration and runoff were calibrated. However, spin-up affects the selection of optimal parameters when soil moisture was calibrated. *INDEX TERMS:* 1655 Global Change: Water cycles (1836); 1833 Hydrology: Hydroclimatology; 3307 Meteorology and Atmospheric Dynamics: Boundary layer processes; 3322 Meteorology and Atmospheric Dynamics: Land/atmosphere interactions; 3260 Mathematical Geophysics: Inverse theory; *KEYWORDS:* optimization, uncertainty estimate, land surface model

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

1. Introduction

[2] Over the past two decades, The Project for Intercomparison of Land-Surface Parameterization Schemes (PILPS) has shown poor agreement and large uncertainties among the schemes [*Henderson-Sellers*, 1996]. The PILPS experiments were performed for different climate zones, soils and vegetation types from stand-alone comparisons such as phase 2a [*Chen et al.*, 1997], phase 2b [*Shao and Henderson-Sellers*, 1998], and phase 2d [*Schlosser et al.*, 2000; *Slater et al.*, 2001] to regional comparisons such as phase 2c [*Wood et al.*, 1998] and phase 2e [*Bowling et al.*, 2003]. The results showed that these uncertainties came from different model development philosophies [*Henderson-Sellers*, 1996; *Sellers et al.*, 1997], different model structures [*Henderson-Sellers et al.*, 1996], and different definitions of effective parameters [*Desborough*, 1999]. Although intercomparison efforts have attempted to minimize these uncertainties by assigning a common set of parameter values for all schemes, no mechanism existed to ensure that the parameter values produce the same effect in terms of the land surface model simulation.

[3] One way to reduce parameter uncertainties is to use automated methods of parameter calibration. *Sellers et al.* [1989] used an iterative loop driven by a least squares reduction program and reliable micrometeorological measurements taken over the Amazonian tropical forests to estimate and to optimize physiological parameters in the simple biosphere model. Their results showed that specification of optimal parameters improved simulations of sensible and latent heat fluxes and reduced simulation uncertainties. Recently, *Gupta et al.* [1999] used a multicriteria parameter calibration technique to estimate optimal parameter values using prior ranges of model parameter. They showed that the biosphere-atmosphere transfer scheme improved the simulations of energy fluxes (i.e., sensible heat, latent heat) and state variables (i.e., soil temperature, soil moisture) when its parameters were optimized using a multicriteria method. *Xia et al.* [2002] investigated the

[1]Institute for Geophysics, John A. and Katherine G. Jackson School of Geosciences, University of Texas at Austin, Austin, Texas, USA.
[2]Geophysical Fluid Dynamics Laboratory, NOAA, Princeton, New Jersey, USA.
[3]Department of Geological Sciences, John A. and Katherine G. Jackson School of Geosciences, University of Texas at Austin, Austin, Texas, USA.

0148-0227/04/2003JD004419$09.00

relationship of model complexities to accuracies of modeled energy fluxes using the Chameleon Surface Model (CHASM) and this multicriteria method for one field site (i.e., Cabauw). Their results showed that a complex model had more accurate simulations of energy fluxes than a simple model when all models were optimized. More recently, *Jackson et al.* [2003] and Y. Xia et al. (Multi-data set study of optimal parameter and uncertainty estimation of a land surface model with Bayesian stochastic inversion and Multicriteria method, submitted to *Journal of Applied Meteorology*, 2003, hereinafter referred to as Xia et al., submitted manuscript, 2003a), compared the efficiency and ability of Bayesian stochastic inversion and multicriteria methods to find the optimal parameter values using the CHASM model at seven measurement sites. The results showed that two methods gave similar optimal parameter values which resulted in similar energy flux simulations.

[4] Previous efforts to make more uncertainty assessments of land surface models were undertaken by *Franks and Beven* [1997]. They used a Monte Carlo sampling of parameters within the soil-vegetation-atmosphere transfer scheme and generalized likelihood uncertainty estimation methodology to analyze uncertainties in land surface-atmosphere flux predictions for the FIFE site (First International Satellite Land Surface Climatology Project Field Experiment) and an Amazonian pasture site. The top 10% of 10,000 different parameter combinations were chosen to represent the uncertainty stemming from model parameters. The results showed that the range of model predictions on surface energy fluxes had typical widths of approximately a third of the maximum observed fluxes for both sites. *Franks and Beven* [1997] reported that the short-term field campaigns represented by the data sets (6–21 August 1987 and 5–16 October 1987 at FIFE site, 16 October to 2 November 1990 and 29 June to 10 September 1991 at Amazon site) may not be adequate to specify parameter values representing a site or area.

[5] Optimal parameter and uncertainty estimates of land surface models require the selection of: (1) a meteorological forcing data set (e.g., downwelling solar radiation, downwelling longwave radiation, precipitation, temperature, wind, humidity), (2) a calibration data set (e.g., soil moisture, evapotranspiration, and runoff), (3) an error function, (4) an automatic parameter search procedure (optimization algorithm), (5) a parameter region with feasible parameter ranges, and (6) a validation or evaluation procedure with independent data sets. The selection of an automatic parameter optimization algorithm has been studied extensively. Global optimization methods such as a multicriteria method [*Gupta et al.*, 1998, 1999] and Bayesian stochastic inversion method [*Sen and Stoffa*, 1996] have been widely applied to land surface models [*Bastidas et al.*, 1999; *Xia et al.*, 2002; *Jackson et al.*, 2003; *Xia et al.*, 2003; Xia et al., submitted manuscript, 2003a] to estimate optimal parameters and/or their uncertainties. Y. Xia et al. (Optimal parameter and uncertainty estimation of a land surface model: Sensitivity to parameter ranges and model complexities, submitted to *Agricultural and Forest Meteorology*, 2003, hereinafter referred to as Xia et al., submitted manuscript, 2003b) discuss impact of the selection of parameter ranges on optimal parameter and uncertainty estimates. Their results showed that local parameter ranges may be more reasonable for obtaining optimal parameters and their uncertainties than the global ranges as used in the work of *Xia et al.* [2002], L. A. Bastidas et al. (Comparative evaluation of land surface models using multi-criteria methods, submitted to *Journal of Geophysical Research*, 2003, hereinafter referred to as Bastidas et al., submitted manuscript, 2003), *Jackson et al.* [2003], and Xia et al. (submitted manuscript, 2003a). However, the sensitivity of optimal parameters and their uncertainties to the data length has not yet been investigated.

[6] In previous studies, less than two years forcing and calibration data were used [e.g., *Gupta et al.*, 1999; *Xia et al.*, 2002; *Jackson et al.*, 2003; *Xia et al.*, 2003; Xia et al., submitted manuscripts, 2003a, 2003b; Bastidas et al., submitted manuscript, 2003] to derive optimal parameters, which were further used in climate simulations [*Sen et al.*, 2001]. *Franks and Beven* [1997] indicated that short-period data may be inadequate to estimate optimal parameters and their uncertainties for land surface models. The sensitivity of optimal parameters to data lengths for a conceptual rainfall-runoff model [*Yapo et al.*, 1996] showed that approximately 8 years of data were required to obtain optimal parameters that were relatively insensitive to the period selected. However, this type of study is rare for land surface models because of lack of long-period meteorological forcing and calibration data. Eighteen years of forcing data and long-period calibration data (e.g., 18-year runoff and soil moisture, and 8-year evapotranspiration) at Valdai provide us with an opportunity to investigate the sensitivity of optimal parameter sets to data lengths using land surface models. At the Valdai site, exist monthly runoff related closely to snowmelt, snow sublimation, evaporation, and soil moisture. This provides a useful setting to examine how the sensitivity varies when different calibration variables are used.

2. Data, Model, and Method

2.1. Forcing and Calibration Data Sets

[7] 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]. The continuous 18 years of atmospheric forcing and hydrologic data have been described in detail by *Fedorov* [1977], *Vinnikov et al.* [1996], and *Schlosser et al.* [1997]. Here we give an overview only for completeness. 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.

[8] The atmospheric forcing data include atmospheric pressure, air temperature, humidity, wind speed, and 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. Shortwave and longwave radiation fluxes were not directly measured, and their estimates are used in this study following

Schlosser et al. [1997]. Original data recorded at 3-hour intervals were interpolated to 30-min intervals.

[9] The calibration data include monthly evaporation, runoff and root zone soil moisture. Monthly evaporation was recorded from a lysimeter from May to October for the years 1966–1973 [*Federov*, 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. Monthly runoff was measured 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, the observed runoff were modified by *Schlosser et al.* [1997] according to variations in the observed averaged water table depth. Total soil moisture in the top 1 m of soil was taken from eleven point measurement sites at the end of every month and was calculated using the thermostat-weight (gravimetric) technique [*Robock et al.*, 1995]. The ranges of error for evapotranspiration, runoff and soil moisture were estimated to be 0.5 mm/day, 0.5 mm/day, and 10 mm, respectively [*Schlosser et al.*, 2000].

2.2. Chameleon Surface Model (CHASM)

[10] The CHASM [*Desborough*, 1999; *Pitman et al.*, 2003] land surface model has been used in offline intercomparison of the 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 land-surface energy balance representation within a common modeling framework that can be run in 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. Within the mosaic-type representation the land-atmosphere interface is divided into two tiles. The first tile is a combination of bare ground and exposed snow with the second tile consisting 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 snowpack 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].

[11] 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.

2.3. Bayesian Stochastic Inversion

[12] The Bayesian stochastic inversion (BSI) method 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) and is given mathematically as

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

where the domain of integration spans the entire model parameter space $\mathbf{m}$, $\sigma(\mathbf{m}|\mathbf{d_{obs}})$ is the PPD, vector $\mathbf{d_{obs}}$ is the observational data, $E(\mathbf{m})$ is the error function, exp[-$sE(\mathbf{m})$] is the likelihood function, $p(\mathbf{m})$ is the "prior" probability density function for $\mathbf{m}$. The shaping factor, s, was estimated by the estimated errors and the method is described in the work of *Jackson et al.* [2003]. Because only the range for each model parameter in $\mathbf{m}$ 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 indicates maximum uncertainty.

[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 estimated by the covariance matrix defined by

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

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) to stochastically select parameter sets. The VFSA is a form of importance sampling that reduces the computational burden of modeling the impact of every possible combination of model parameters. VFSA algorithm will sample more frequently those regions of the PPD that are more probable. As reported in the work of *Sen and Stoffa* [1995, 1996], the VFSA algorithm can be used with the BSI framework to approximate the multidimensional PPD, even when the relationship between parameters is nonlinear. Since VFSA converges to an optimal solution quickly, repeated runs of VFSA are used to sample the model space and all parameter evaluations from all VFSA runs are used to estimate the PPD.

2.4. Error Functions Used

[15] The square error for normalized variables (E^2) and the ratio of variance of the errors to the variance of observations were used in this study to investigate the sensitivity of optimal parameters and uncertainty estimates to different error functions. The normalization

Table 1. Descriptions and Ranges of 13 CHASM Parameters and 3 Initial Variables

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
Z0N	snow roughness length (m)	1.0×10^{-3}	4.0×10^{-3}
Z0V	vegetation roughness length (m)	0.00	0.20
TS	initial soil temperature (K)	260	265
SW	initial soil wetness	0.70	1.00
SWE	initial snow water equivalent (mm)	30.0	80.0

was used in computing E^2 for two reasons. First, the square error is sensitive to any differences between two data sets, and this sensitivity can be reduced by first normalizing these data [*Martinson et al.*, 1982]. Second, the square error calculated in this way is usually within a unity, which ensures the Bayesian stochastic inversion to be used more reasonably and effectively. Therefore E^2 is defined as

$$E^2 = \sum_{i=1}^{N} [o_n - s_n]^2, \tag{3}$$

where o_n and s_n are normalized observed and simulated data, respectively. N is the number of observed data. Here o_n and s_n are calculated by

$$o_n = \frac{obs_n}{\left[\sum\limits_{n=1}^{N} (obs_n)^2\right]^{\frac{1}{2}}} \tag{4}$$

$$s_n = \frac{sim_n}{\left[\sum\limits_{n=1}^{N} (sim_n)^2\right]^{\frac{1}{2}}}, \tag{5}$$

where obs_n is the observed data, and sim_n is the simulated data. Due to the use of normalized data, E^2 mainly measures the consistency of varying trends of observed and simulated data because E^2 is proportional to the coherence (C) between simulated and calibrated data sets, that is, E^2 = 2(1-C). The ratio of variance of the errors to the variance of observations is defined as $\frac{\sum_{n=1}^{N}(obs_n - sim_n)^2}{\sum_{n=1}^{N}\left(obs_n - \overline{obs}\right)^2}$. In this definition, $\overline{obs}$ is a mean value of obs_n. The ratio of variance of the errors to the variance of observations measures the fraction of the error variance of the observed data explained by the relative magnitude of the residual variance to the variance of the observed data. Its value is 0.0 when the simulated data match the observed data.

[16] Use of different error functions has little impact on the selection of optimal parameters and uncertainty estimates of model parameters for all CHASM parameters. This result is consistent with that derived by *Leplastrier et al.* [2002]. Therefore E^2 was used for all following analyses.

3. Experiment Design

[17] Table 1 lists 13 CHASM model parameters and their feasible ranges. In order to reduce the computing burden, we 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. Individual parameter sensitivity analysis is shown in Figure 1. For different calibration variables, sensitive parameters are different because these variables are related to different physical processes and parameters. For example, WRMAX and Z0V are very sensitive parameters for soil moisture simulations but are less sensitive for evapotranspiration and runoff simulations when compared to other most sensitive parameters. ALBN is most sensitive for runoff simulations but not sensitive for soil moisture simulations. However, overall results showed that snow albedo (ALBN), vegetation albedo (ALBV), vegetation cover fraction (VEGM), vegetation cover seasonality (VEGS), minimum stomatal resistance (RCMIN), maximum water holding capacity (WRMAX), and vegetation roughness length (Z0V) are sensitive to the simulations of monthly evapotranspiration, runoff and soil moisture. In our experiment design, six insensitive parameters were assigned to default, fixed values, while the remaining seven sensitive parameters were allowed to vary according to the specified ranges.

[18] The impacts of spin-up on optimal parameter estimates and uncertainties were investigated for three calibration variables (i.e., evapotranspiration, runoff, soil moisture). For each of these calibration variables there was one pair of experiments, the first experiment having a one-year spin-up [*Schlosser et al.*, 2000] and the other no-spin-up but the initial state variables (soil moisture, ground temperature, and snow water equivalent) being used as optimization parameters. In the one-year spin-up experiment, CHASM was run continuously for 18 years (1966–1983), but the last 17 years (1967–1983) of calibration data were used to calculate error functions.

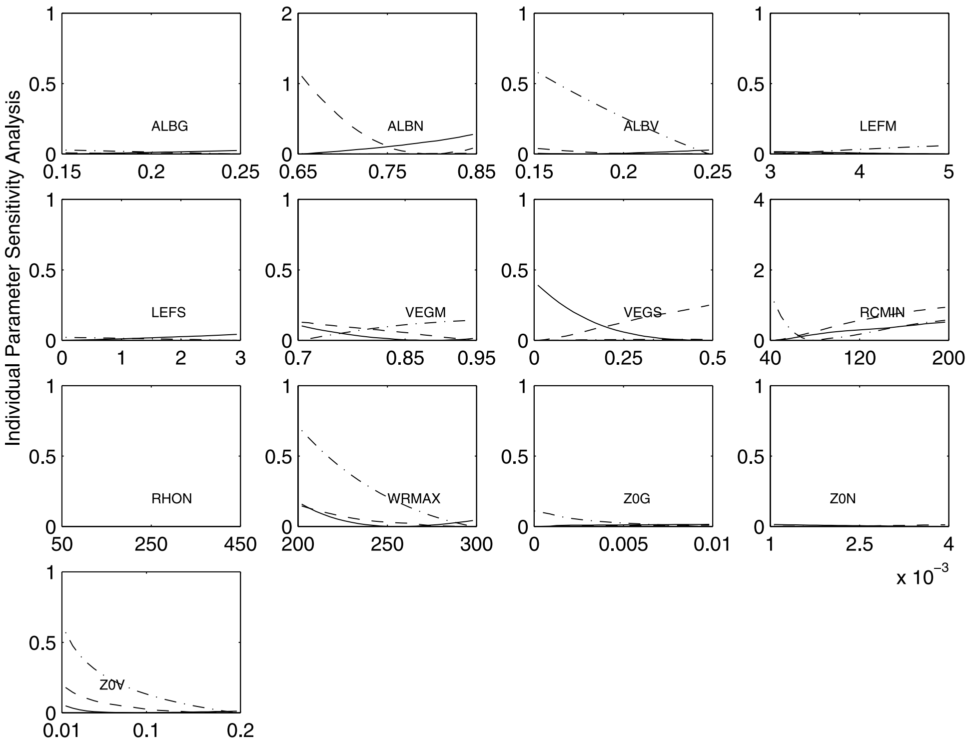

Figure 1. Sensitivity analysis of 13 CHASM parameters for monthly evaporation (solid line), runoff (dashed line), and root zone soil moisture (dashed-dotted line). *Y* axis values were computed as a ratio of difference between calculated and minimum error values to minimum error values.

In the no-spin-up experiment, CHASM was continuously run for 17 years starting from 1967, and these 17 years of calibration data were used to calculate error functions. Compared to the one-year spin-up experiment, the no-spin-up experiment has three additional parameters (initial soil moisture, initial ground temperature, initial snow water equivalent). Therefore no-spin-up experiment has a total of 10 parameters.

[19] In order to investigate the impacts of calibration data lengths on optimal parameter estimates and uncertainty of model parameters, we designed 44 experiments to run the CHASM model using 7 model parameters (see Table 2). In each experiment, a one-year spin-up period was used to minimize initialization errors. For the simulations of monthly runoff and soil moisture, model runs were conducted using three independent samples with lengths of 1, 2, 3, 4, and 6 consecutive years, two independent samples with lengths of 8 consecutive years, and one independent data set with a length of 17 years. To objectively assess these impacts, three samples were randomly selected from the 18-year data set and abnormal years (e.g., the driest year or the wettest year) were excluded in this study for 1-yr calibration. Because we have only 8 years of evapotranspiration data, we used three independent samples with lengths of 1 year and 2 years, two independent samples with lengths of 3 years, and one independent sample with length of 7 years.

[20] Performance of the CHASM model was assessed using two criteria: percent bias and Nash-Sutcliffe efficiency.

Table 2. Independent Data Samples Used in Sensitivity Study of Data Length for Three Calibration Variables[a]

Data Length	Sample 1	Sample 2	Sample 3
	Calibration Variable = Monthly Evapotranspiration		
1 year	1967	1970	1973
2 years	1967–1969	1970–1971	1972–1973
3 years	1967–1970	1971–1973	–
	Calibration Variable = Monthly Runoff or Soil Moisture		
1 year	1967	1977	1983
2 years	1970–1971	1978–1979	1982–1983
3 years	1967–1969	1975–1977	1980–1982
4 years	1967–1970	1972–1975	1978–1981
6 years	1967–1972	1973–1978	1978–1983
8 years	1967–1974	1976–1983	–

[a]Sample 1967–1973 for monthly evapotranspiration and 1967–1983 for monthly runoff and soil moisture, and total 44 samples were used in this study.

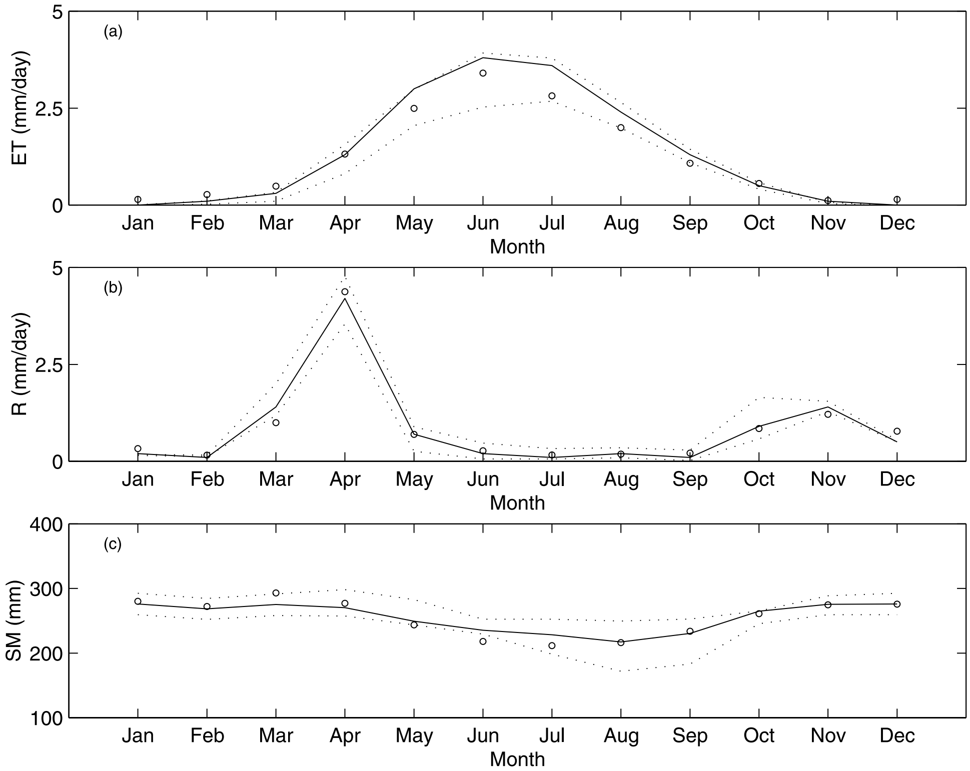

Figure 2. The 17-year average of observed and simulated monthly evapotranspiration, runoff, and root zone soil moisture and their uncertainty envelopes calculated with the BSI at 95% confidence level (observations, open circle; simulated, solid line; uncertainty envelope, dotted line).

These two criteria used by *Yapo et al.* [1996] are defined

$$100\% \times \sum_{n=1}^{N}(obs_{n-sim_n}) / \sum_{n=1}^{N} obs_n, \text{ and } 1.0 - \frac{\sum_{n=1}^{N}(obs_n - sim_n)^2}{\sum_{n=1}^{N}\left(obs_n - \overline{obs}\right)^2},$$

respectively. In these two definitions, obs_n is the observed data, sim_n is the simulated data, $\overline{obs}$ is a mean value of obs_n, and N is the number of observed data. Percent bias measures the tendency of the simulated data to be larger or smaller than their observed counterparts. Its optimal value is 0.0, positive values indicating underestimation and negative values indicating overestimation. Nash-Sutcliffe efficiency measures the fraction of the variance of the observed data explained by the model in terms of the relative magnitude of the residual variance ("noise") to the variance of the observation ("signal").

4. Results

4.1. Comparison Against Observations

[21] Figure 2 shows the ability of the CHASM model in reproducing the observed evapotranspiration, runoff and soil moisture. The simulations are generated using the optimal parameters derived by the BSI method and available calibration data (e.g., 17-year for runoff and soil moisture, 8-year for evapotranspiration). Uncertainty estimates of simulated evapotranspiration, runoff, and soil moisture at the 95% confidence level were derived by the selected best parameter sets. These best parameter sets were selected from over 50,000 parameter sets using the BSI method and estimated *s* in section 2.3. It should be noted that one-year spin-up has been used for the above results when the CHASM model was run. The results show that the optimal simulations agree well with the observations for the three calibration variables. The optimal simulations were enveloped by the uncertainty ranges for the three variables with the exception of the observed evapotranspiration and runoff in January, March and December and observed soil moisture in June. These poor uncertainty estimates of winter evaporation and runoff and early spring runoff may be mainly due to crudeness of the snow model. The snow scheme uses only one layer to represent the vertical structure, and excludes snow melting and refreezing processes.

4.2. Impacts of Spin-Up

[22] Figure 3 shows the impacts of spin-up on normalized optimal parameters and the simulations of runoff, evapotranspiration, and soil moisture. These optimal parameters are normalized such that for each parameter and each calibration experiment, the difference between the optimized and the minimum parameter values is divided by the difference between the maximum and minimum parameter values (see Table 1). Three conclusions are in order. First, the optimal parameter values derived by the observed evapotranspiration are very similar for one-year spin-up and no-spin-up runs (Figure 3a), and therefore generate the similar simulations (Figure 3b). Second, all of the optimal parameter values derived by the observed runoff are similar for one-year spin-up and no- spin-up runs except for WRMAX

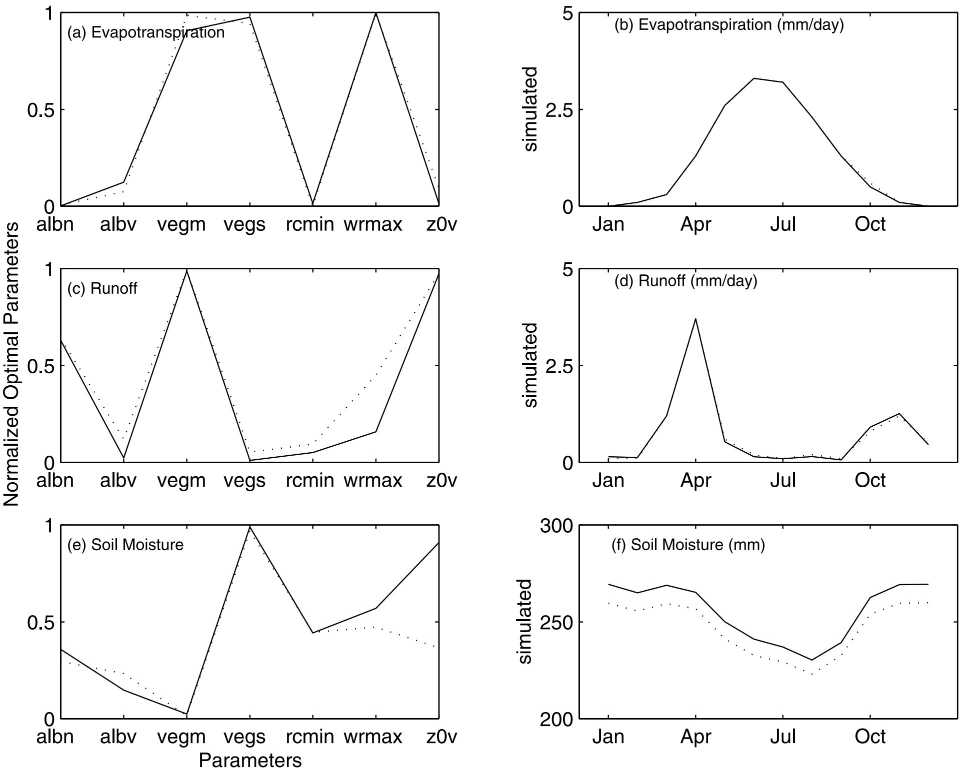

Figure 3. Normalized optimal parameters and 17-year (1967–1983) averages of simulated evapotranspiration, runoff, and soil moisture using the optimal values (one-year spin-up run, solid; no-spin-up run, dashed; albn, ALBN; albv, ALBV; vegm, VEGM; rcmin, RCMIN; wrmax, WRMAX; z0v, Z0V).

(Figure 3c). These similar optimal parameter values lead to similar runoff simulations (Figure 3d) because WRMAX is less sensitive to runoff simulation. Third, most of the optimal parameter values derived by the observed soil moisture are similar for one-year spin-up and no-spin-up runs except for WRMAX and Z0V (Figure 3e). However, because WRMAX and Z0V are two of most sensitive parameters for soil moisture simulations, differences between spin-up and no-spin-up runs for these two parameters lead to different soil moisture simulations (Figure 3f). The results in Figure 3 show that spin-up has little impact on the selection of optimal parameters when monthly evapotranspiration and runoff were calibrated. However, spin-up significantly impacts the selection of optimal parameters when root zone soil moisture was calibrated. This may be because initial conditions (e.g., initial soil moisture) may impact simulations of root zone soil for longer than one year. An examination of the multiyear impact of spin-up on root zone soil moisture can be found in the work of *Cosgrove et al.* [2003]. Marginal PPDs of seven CHASM parameters show that spin-up and no-spin-up runs have almost the same PPD distributions for each CHASM parameter and calibration variable, meaning that spin-up and no-spin-up runs result in similar parameter uncertainties. Most of the previous works used 1 year of data for calibration, and therefore a similar analysis was performed using 1 year of (i.e., 1967) data. The results support the above conclusion except for the runoff case where initial snow water equivalent had significant effects on optimal parameter and uncertainty estimates of model parameters.

4.3. Impacts of Data Length

[23] The issue explored in this section is how many years of data are required to obtain a consistent optimal parameter set and uncertainty estimate of model parameters. The impacts of data length were evaluated by analyzing normalized optimal parameter values, empirical cumulative distribution functions, and marginal PPD distributions. This evaluation was performed for monthly runoff, soil moisture and evapotranspiration.

4.3.1. Impacts on Optimal Parameters

[24] The normalized optimal parameter values for 1-year, 2-year, 3-year, 4-year, 6-year, 8-year and 17-year calibrations are shown in Figure 4 when monthly runoff was used as a calibration variable. In the case of 1-year and 2-year calibrations, the optimal parameter values vary widely among the three samples. Sample 3 (the year 1983) in the 1-year calibration run produces the lowest value of snow albedo among the samples. All the conditions being equal, the lowest snow albedo leads to the most absorbed solar radiation, and hence the earliest snowmelt and runoff peak [*Slater et al.*, 2001] because runoff at Valdai is mainly determined by snowmelt process [*Schlosser et al.*, 2000]. Indeed, the observed snow ablation and runoff in sample 3 peaked in March, the earliest among the three samples (Figure 5).

[25] As the calibration lengths increase from three to eight years, the optimal parameter values of the three independent

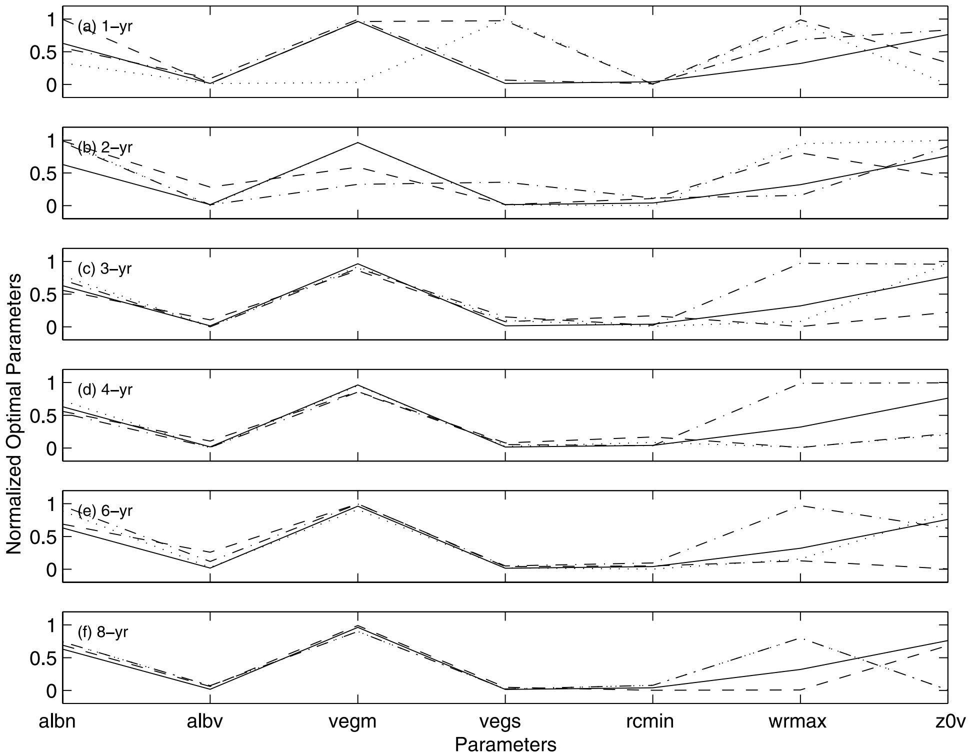

Figure 4. Normalized optimal parameters for three independent samples for one-year, two-year, three-year, four-year, six-year, and eight-year calibrations when observed monthly runoff was used (sample 1, dashed-dotted; sample 2, dashed; sample 3, dotted; solid, 17-year calibration; albn, ALBN; albv, ALBV; vegm, VEGM; rcmin, RCMIN; wrmax, WRMAX; z0v, Z0V).

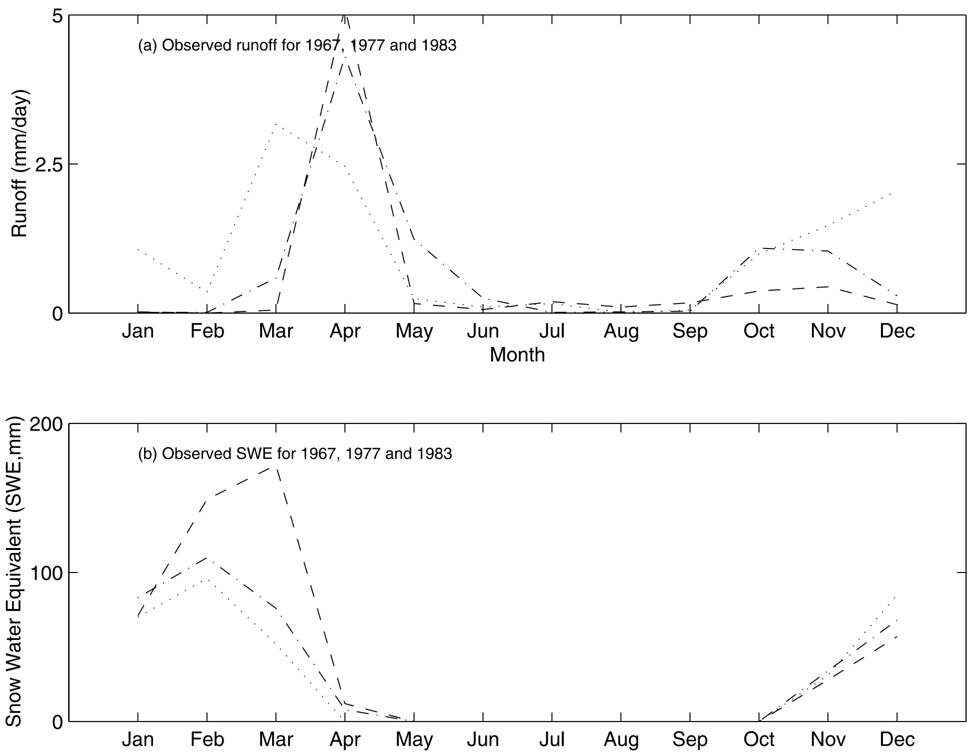

Figure 5. Observed runoff and snow water equivalent for the years 1967, 1977, and 1983 (1967, dashed-dotted line; 1977, dashed line; 1983, dotted line).

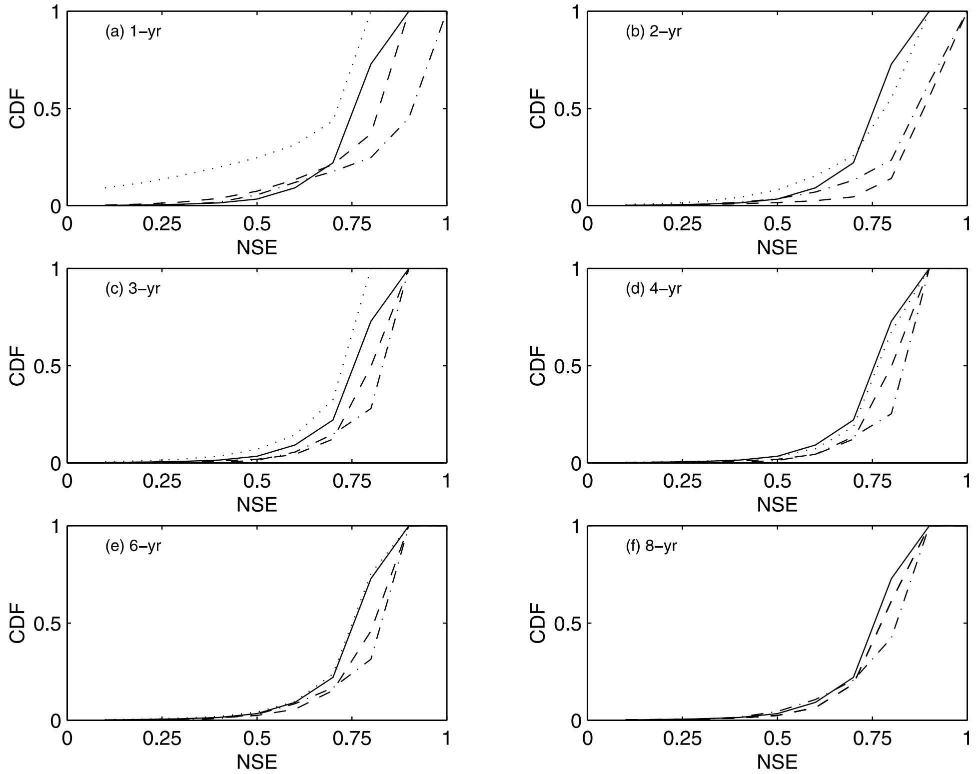

Figure 6. Empirical cumulative distribution functions (CDFs) of the Nash-Sutcliffe efficiency (NSE) for different calibration data lengths when observed monthly runoff was used as a calibration variable (sample 1, dashed-dotted; sample 2, dashed; sample 3, dotted; solid, 17-year calibration).

samples merge toward those of 17-year calibration for all parameters except for WRMAX and Z0V. Differences in optimal values of WRMAX and Z0V exist for the three samples even though 8-year data were used. Despite these differences, the variations in the values of the two parameters do not significantly affect the model output (Figure 1). Another special case is the 6-year calibration where ALBN and ALBV show some differences for three samples. However, these small differences have little impact on the runoff simulations. This result is confirmed by comparing the runoff simulations from a series of 17-year runs. In each of these runs, the optimal parameters derived from an *n*-year runoff calibration run were used, where *n* is 1, 2, 3, 4, 6, 8, or 17. It is found that all the runoff simulations are similar when more than three years of calibration data were used. Therefore we conclude that, in the case of runoff, three years of calibration data are required to obtain the optimal parameters that are insensitive to the period selected.

[26] In the case of evapotranspiration, 1-year calibration data are required to obtain the optimal parameters which are insensitive to the period selected. Indeed, all the optimal parameter values except WRMAX and Z0V are similar for the three samples after 1-year data were used. The variations in WRMAX and Z0V values do not affect the simulations of evapotranspiration. This conclusion was also confirmed by a series of 17-year runs. In the case of root zone soil moisture, 8-year calibration data are required to obtain the optimal parameters which are insensitive to the period selected.

[27] In conclusion, the lengths of calibration data required to generate consistent optimal parameter values and model outputs may depend on the calibration variables because these variables are closely related to model physical processes which have different timescales. Furthermore, the calibration variables may depend on site-specific characteristics concerning vegetation, site and climate. Therefore the minimal lengths of calibration data may be different at different sites. More discussion on this point can be seen in section 4.4. It is clear that longer calibration data may give more consistent optimal parameter values and result in more consistent simulations. However, due to huge cost for measurement and collection of calibration and forcing data, 1-year long calibration data sets have often been used in previous land surface modeling studies. Our results here indicate that 1-year data sets may be adequate for deriving the optimal parameter values if evapotranspiration is calibrated at midlatitude grassland such as Valdai, but not so if runoff or root zone soil moisture is calibrated.

4.3.2. Impacts on Uncertainty Estimates

[28] The empirical Cumulative Distribution Functions (CDFs) were constructed for different statistic measures (e.g., Nash-Sutcliffe efficiency, percent bias) and for different data lengths (Figures 6 and 7). Each CDF indicates an uncertainty estimate of results simulated by all parameter sets (over 50,000). If a specific percentage of the best parameter sets was given, say, 10% [*Franks and Beven*, 1997], we are able to use the selected parameter sets to estimate uncertainties of simulated runoff. Figures 6a–6f show that the CDFs of the Nash-Sutcliffe efficiency merge as the data length increases from 1 to 8 years. The closer CDFs mean more consistent uncertainty estimates of the simulated runoff. When 8-year calibration data are used,

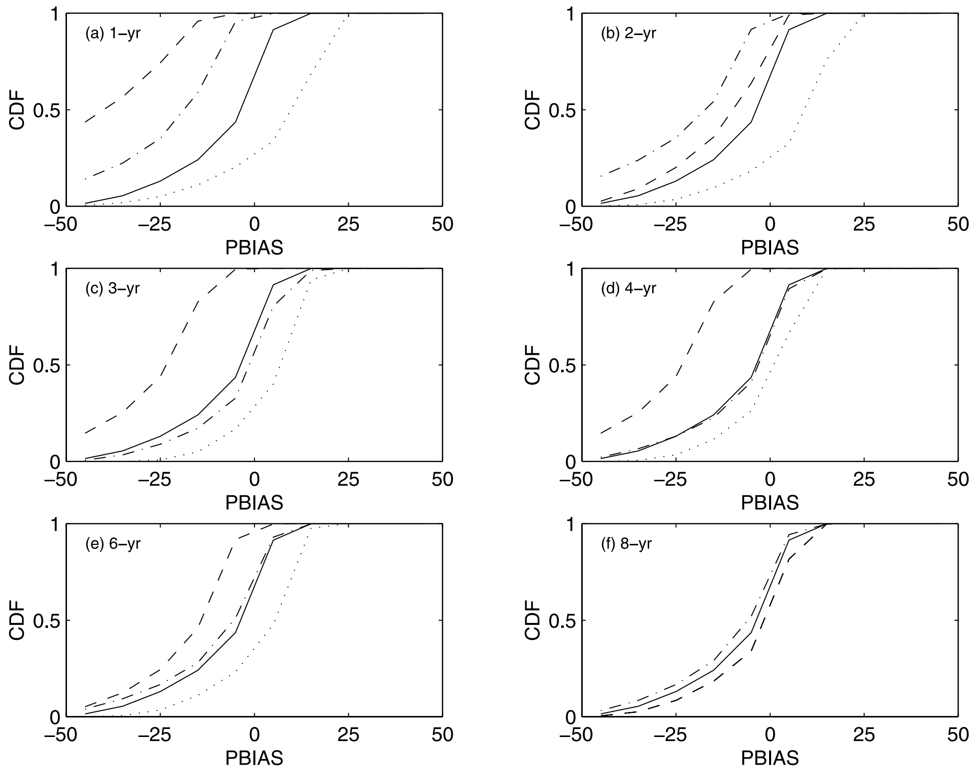

Figure 7. Same as Figure 6 but for percent bias (PBIAS).

CDFs are closest for all cases. This result is more obvious for the percent bias analysis (Figure 7). As the length of calibration data increases from 1 to 8 years, CDFs for three independent samples merge toward those for the 17-year data. Significant differences exist for CDFs when 1-, 2-, 3-, 4- or 6-year calibration data are used. However, the 8- and 17-year CDFs are very similar. Therefore 8-year calibration data may be required to estimate the uncertainties that are insensitive to the period selected.

[29] Similar conclusions may be drawn in the analysis of a marginal posterior probability density function because it can be used to quantify uncertainties in the derived parameter sets. Figures 8 and 9 show that marginal PPDs for the seven CHASM model parameters evaluated by three independent 1-year calibration data and two independent 8-year data. The marginal PPDs calculated by 17-year data are also shown in Figures 8 and 9 for comparison. For 1-year analysis (Figure 8), the width of the PPD distribution varies broadly for three samples for ALBN, ALBV, VEGS, WRMAX and Z0V if a specific percentage of best parameter sets (e.g., 10%) is given. Furthermore, the marginal PPDs are also significantly different for 1-year and 17-year results. Therefore different marginal PPDs calculated from three independent samples would have resulted in different estimates of parameter uncertainties. However, for 8-year analysis (Figure 9), the width and shape of the PPD distribution are similar for three samples (two 8-year samples and one 17-year sample) for seven CHASM parameters except for WRMAX and Z0V. The possible reasons for the peculiar behavior of WRMAX and Z0V include (a) parameter interdependency (parameters interact with each other), (b) criterion inadequacy (the objective function does not properly extract the information contained in the data), and (c) insensitivity (variations in the values of the parameters do not significantly affect the model output). Our correlation matrices for parameters showed that none of correlations was large enough to support parameter interaction as the dominant reason for this inconsistency although the matrices only represent linear relationships between parameters. In addition, a comparison of two error functions also showed that PPDs are not sensitive to the selection of error functions as discussed in section 2.4. Therefore the insensitivity of WRMAX and Z0V to runoff simulation is a dominant reason. Therefore use of 8-year calibration data may be required to obtain consistent PPD estimates which are not sensitive to the period selected.

[30] Again, the conclusion above is dependent on the calibration variables. The CDFs of the percent bias for monthly evapotranspiration show that they merge as the data length increases from 1 to 3 years (Figure 10). Analysis of marginal PPDs for different data lengths also shows similar PPDs for 3-year calibration data. Therefore 3-year of data may be sufficient to estimate consistent uncertainties for monthly evapotranspiration. The similar analysis for monthly soil moisture shows that 8-year calibration data are needed for obtaining a consistent uncertainty estimate.

4.4. Discussion

[31] It should be noted that 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]. Therefore the use of a multicriteria method

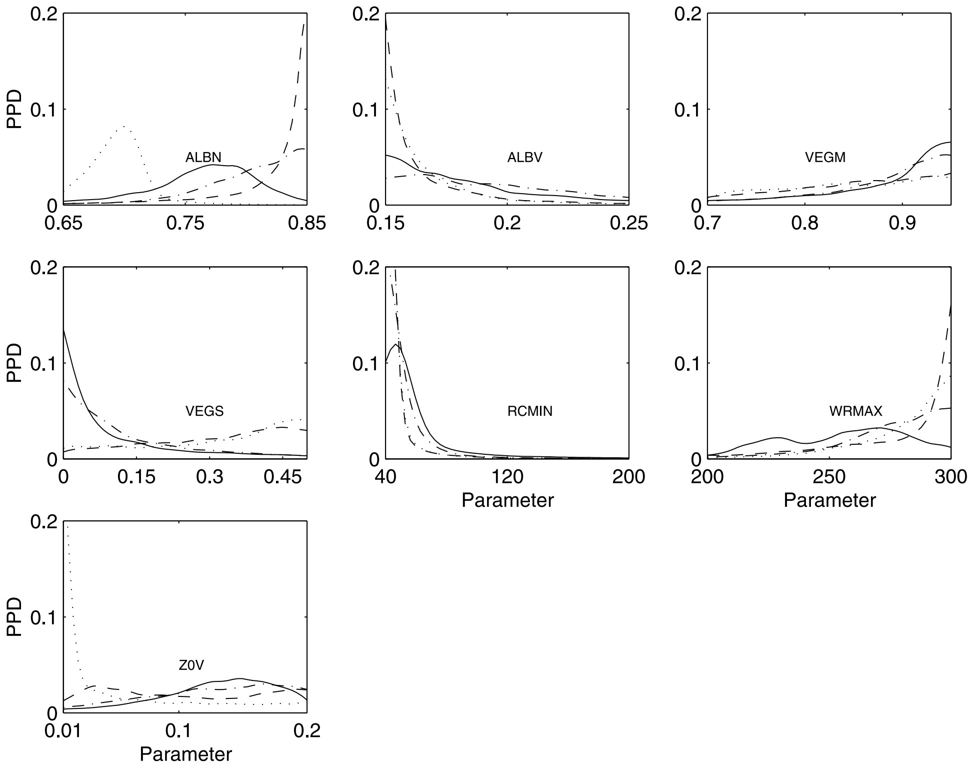

Figure 8. Calculated PPDs for three 1-year independent calibration samples and 7 CHASM model parameters when observed monthly runoff was used as calibration variables (sample 1, dashed-dotted line; 2, dashed line; sample 3, dotted line; solid line, 17-year calibration).

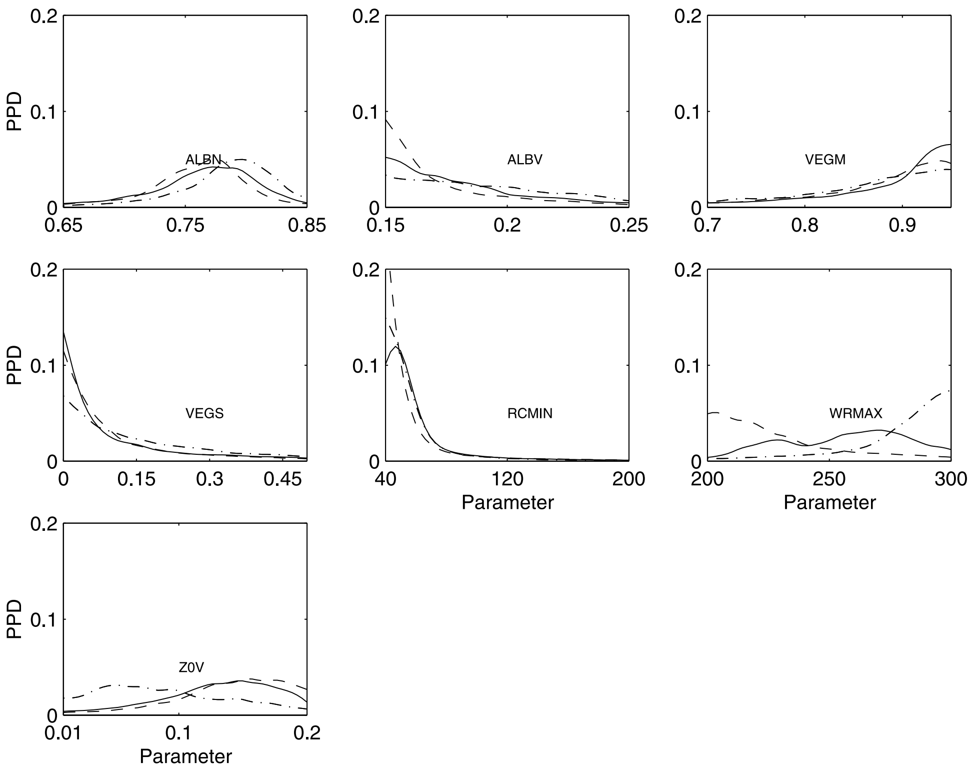

Figure 9. Same as Figure 8 but for two 8-year independent calibration samples.

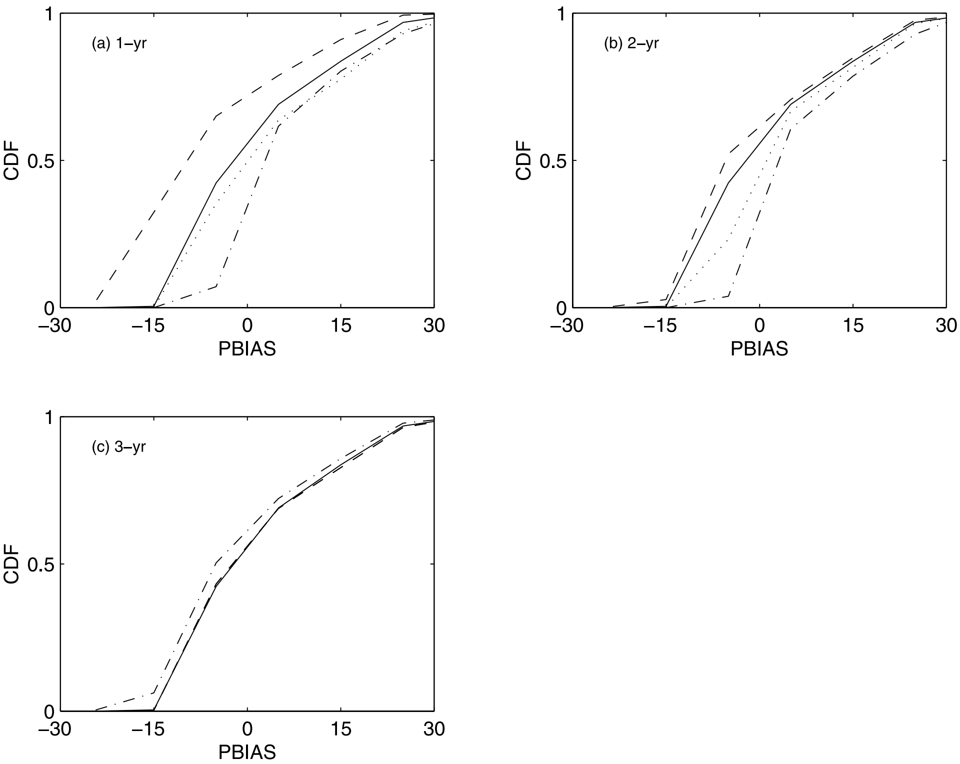

Figure 10. Empirical cumulative distribution functions (CDFs) of the percent bias (PBIAS) for different calibration data lengths when observed monthly evapotranspiration was used as a calibration variable (sample 1, dashed-dotted line; 2, dashed line; sample 3, dotted line; solid line, 17-year calibration).

in uncertainty analysis of land surface models will be addressed in the future.

[32] In addition, this study focused on one site and one model. It remains to be addressed how the optimal data length depends on site characteristics concerning vegetation, soil and climate as well as on land surface models. However, the calibration results of monthly evapotranspiration at the Valdai site show that only one year of calibration data is required, suggesting that the interannual variability of evaporation at this site may be very low. Further insight will be gained from the ongoing PILPS San Pedro (see www.sahra.arizona.edu/pilpssanpedro) which involves the use of different sites, long-term data records, and different land surface models.

5. Conclusions

[33] The primary goal of this study is to demonstrate the importance of calibration data (e.g., data length, different calibration variables, and initial conditions) to estimates of optimal parameters and uncertainty of the CHASM model. The paper has shown that the optimal parameters give relatively accurate simulations of evapotranspiration, runoff and soil moisture. The BSI method also gives a relatively reasonable uncertainty envelope for evapotranspiration, runoff and soil moisture at the 95% confidence level. We have shown that one year of data for evapotranspiration, three years of data for runoff and eight years of data for root zone soil moisture are required to obtain consistent optimal parameters which are insensitive to the period selected. Approximately three-year data for evapotranspiration, eight-year data for runoff and root zone soil moisture may be needed to obtain a consistent uncertainty estimate which is insensitive to the period selected. The sensitivity of optimization process and uncertainty estimates to calibration data length is dependent on the calibration variables. This conclusion is physically reasonable because these variables are closely related to different physical processes which have different timescales.

[34] Both spin-up and no-spin-up runs have similar optimal parameter values for monthly evapotranspiration and runoff. For monthly soil moisture, the two runs result in different optimal values for Z0V and hence different simulations. However, they lead to similar PPD distributions for all seven model parameters regardless of calibration variables. Analysis of one-year (e.g., 1967) and 17-year calibrations give similar conclusions except for the simulation of runoff where initial snow water equivalent has a significant effect on the selection of optimal parameters and marginal PPD distributions of model parameters.

[35] **Acknowledgments.** The authors wish to thank A. J. Pitman for providing us the CHASM land surface model. YX also wishes to acknowledge the technical support provided by M. Wiederspahn at the Institute for Geophysics, University of Texas at Austin (UTIG). YX and CJ were supported by the Unger Vetlessen Foundation. ZLY was supported under NASA grant NAG5-12577 and NOAA grant NA03OAR4310076.

References

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.

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.

Budyko, M. I. (1956), *Heat Balance of the Earth's Surface* (in Russian), 255 pp., Gidrometeoizlat, St. Petersburg.

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

Cosgrove, B. A., et al. (2003), Land surface model spin-up behavior in the North Land Data Assimilation System (NLDAS), *J. Geophys. Res.*, *108*(D22), 8845, doi:10.1029/2002JD003316.

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), 264 pp., Gidrometeoizlat, St. Petersburg.

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.

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, 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. (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.

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.

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.

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.

Martinson, D. G., W. Menke, and P. Stoffa (1982), An inverse approach to signal correlation, *J. Geophys. Res.*, *87*(B6), 4807–4818.

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.

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 (1995), *Global Optimization Methods in Geophysical Inversion*, 281 pp., Elsevier Sci., New York.

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

Sen, O. L., L. A. Bastidas, W. J. Shuttleworth, Z.-L. Yang, H. V. Gupta, and S. Sorooshian (2001), Impact of field-calibrated vegetation parameters on GCM climate simulations, *Q. J. R. Meteorol. Soc. London B*, *127*(574), 1199–1223.

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

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.

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., P. L. Stoffa, C. Jackson, and M. K. Sen (2003), Effect of forcing data errors on calibration and uncertainty estimates of CHASM model: A multi-dataset study, in *Observations, Theory, and Modeling of Atmospheric and Oceanic Variability*, *World Sci. Ser. on Meteorol. of East Asia*, vol. 3, World Sci., Tokyo, in press.

Yapo, P. O., H. V. Gupta, and S. Sorooshian (1996), Automatic calibration of conceptual rainfall-runoff models: Sensitivity to calibration data, *J. Hydrol.*, *181*, 23–48.

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.

C. Jackson, M. K. Sen, and P. L. Stoffa, Institute for Geophysics, John A. and Katherine G. Jackson School of Geosciences, University of Texas at Austin, 4412 Spicewood Springs Rd., Building 600, Austin, TX 78759-8500, USA. (charles@utig.ig.utexas.edu; mrinal@utig.ig.utexas.edu; pauls@ utig.ig.utexas.edu)

Y. Xia, Geophysical Fluid Dynamics Laboratory, NOAA, Forrestal Campus, U.S. Route 1, Princeton, NJ 08542, USA. (youlong.xia@noaa.gov)

Z.-L. Yang, Department of Geological Sciences, John A. and Katherine G. Jackson School of Geosciences, University of Texas at Austin, 1 University Station #C1100, Austin, TX 78712-0254, USA. (liang@mail.utexas.edu)