

Global Modeling of Land Water and Energy Balances. Part III: Interannual Variability

A. B. SHMAKIN*

NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey

P. C. D. MILLY AND K. A. DUNNE

U.S. Geological Survey and NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey

(Manuscript received 18 July 2001, in final form 9 January 2002)

ABSTRACT

The Land Dynamics (LaD) model is tested by comparison with observations of interannual variations in discharge from 44 large river basins for which relatively accurate time series of monthly precipitation (a primary model input) have recently been computed. When results are pooled across all basins, the model explains 67% of the interannual variance of annual runoff ratio anomalies (i.e., anomalies of annual discharge volume, normalized by long-term mean precipitation volume). The new estimates of basin precipitation appear to offer an improvement over those from a state-of-the-art analysis of global precipitation (the Climate Prediction Center Merged Analysis of Precipitation, CMAP), judging from comparisons of parallel model runs and of analyses of precipitation–discharge correlations. When the new precipitation estimates are used, the performance of the LaD model is comparable to, but not significantly better than, that of a simple, semiempirical water-balance relation that uses only annual totals of surface net radiation and precipitation. This implies that the LaD simulations of interannual runoff variability do not benefit substantially from information on geographical variability of land parameters or seasonal structure of interannual variability of precipitation.

The aforementioned analyses necessitated the development of a method for downscaling of long-term monthly precipitation data to the relatively short timescales necessary for running the model. The method merges the long-term data with a reference dataset of 1-yr duration, having high temporal resolution. The success of the method, for the model and data considered here, was demonstrated in a series of model–model comparisons and in the comparisons of modeled and observed interannual variations of basin discharge.

1. Introduction

Global models of land water and energy balance can provide information of value in a variety of geophysical fields. Their use is commonplace to describe dynamic boundary conditions for the atmospheric general circulation models used for weather prediction and climate analysis (Chen et al. 1997). They are essential for the provision of important environmental information in an emerging class of global land vegetation models (e.g., Foley et al. 1996). They can generate estimates of changing water storage for analyses of time-varying global gravity fields (Wahr et al. 1998) and for evaluation of induced crustal deformations (van Dam et al. 2001).

* Current affiliation: Institute of Geography, Russian Academy of Sciences, Moscow, Russia.

Corresponding author address: Dr. P. C. D. Milly, Water Resources Division, U.S. Geological Survey, NOAA/Geophysical Fluid Dynamics Laboratory, P.O. Box 308, Princeton, NJ 08542.
E-mail: cmilly@usgs.gov

In most of these applications of land models, the capability of the model to reproduce variability at an annual timescale is an important requirement. This need is the motivation for the present investigation, in which we evaluate the capability of the Land Dynamics (LaD) model (Milly and Shmakin 2002a, henceforth Part I) to estimate interannual variations in river discharge. River discharge is arguably the best observable measure of water balance (and, indirectly, of energy balance) available for areas with horizontal length scales of hundreds of kilometers and greater. This investigation is distinct from, but complementary to, our evaluation of the capability of the LaD model to simulate land-characteristic influences on water and energy balances during a particular year (Milly and Shmakin 2002b, henceforth Part II).

A rigorous evaluation of model performance must rely on observational data. However, the considerable, often unknown, magnitude of observational errors creates a serious impediment to optimal use of observational data in model testing. The importance of carefully selecting river basins for analysis based on objective, a

priori measures of precipitation error was shown in Part I. Accordingly, we shall make use here of the error estimates provided by Milly and Dunne (2002a, manuscript submitted to *Water Resour. Res.*, hereafter MID-Ua), who evaluated systematic and random errors in their analyses of basin-mean precipitation.

Milly and Dunne (2002b, manuscript submitted to *Water Resour. Res.*, hereafter MIDUb) have investigated climatic controls on interannual variability of large-scale water and energy fluxes using a simple, semiempirical model. The simple model is a powerful tool for the elucidation of controls on basin-mean, annual-mean fluxes. However, it cannot provide the spatial and temporal resolution of a detailed, numerical land model, nor the internal physical (state) information. (Here we use the term “detailed” to describe models, such as the LaD model, that resolve processes in a spatially distributed manner and at a subdaily timescale; admittedly, a wide range in degree of detail is allowed within this definition.) On the other hand, we believe the simple model provides a fundamental point of reference in the performance assessment of more detailed models (Koster et al. 1999). In principle, a more detailed model should be able to perform better than a simple model, because it incorporates more information. The actual achievement of this hypothetical superiority could be taken as one measure of the maturity of a more detailed model. Accordingly, here we continue the investigation begun in Part II, asking whether the LaD model can perform as well as the semiempirical balance model. Whereas Part II addressed the case of geographic variability of water balance during a particular year, here we look at interannual variability of annual water balance.

Any attempt to apply a detailed land water- and energy-balance model globally at a multiyear timescale is immediately faced with the problem of the lack of suitable model forcing. Data are needed that resolve the diurnal cycle, with horizontal resolution on the order of a hundred kilometers, for various variables, some of which are not measured on an operational basis. The most suitable source of data for such an analysis is the International Satellite Land–Surface Climatology Project (ISLSCP) Initiative I dataset (Meeson et al. 1995). This dataset was created by merging of various land-based and satellite observations, using atmospheric models especially for interpolation. Because it is based on historical data, it reflects (if imperfectly) the true multivariate, space–time covariance structure of the suite of forcing variables.

The disadvantage of the ISLSCP Initiative I dataset for our purpose is its relatively short time span of only 2 yr. An effort (initiative II) has begun (IGPO 1999) to develop a similar data product spanning a much longer timescale, but no such dataset was available for this study. On the other hand, we do have various reasonable estimates of multiyear variability of precipitation at a monthly timescale. In this paper, we develop and test a

method of construction of high-frequency, long-term forcing datasets by merging of long-term monthly precipitation data with high-frequency, short-term forcing (i.e., ISLSCP initiative I). Adoption of such an approach ignores the interannual variability of the temporal structure of forcing at all timescales shorter than 1 month, as well as the interannual variability of all variables other than precipitation. It is important to evaluate the consequences of these approximations, and this issue is explored herein.

In summary, the objectives of this study are 1) to test the capability of the LaD model to reproduce observed interannual variations in river discharge; 2) to compare this capability with the corresponding capability of a simple, semiempirical model of annual water balance; and 3) to develop and test a simple methodology for construction of long-term, high-frequency forcing for use by land models in stand-alone mode.

2. Methodology

a. Land Dynamics (LaD) model

The LaD model has been described and tested in Parts I and II. Water storage is tracked in snow, glacier ice, root-zone, and groundwater stores. Heat is stored as latent heat of fusion of snow and glacier ice, and as sensible heat in the ground, the latter represented by a one-dimensional conduction equation. Runoff is generated as necessary to keep root-zone water content from exceeding a given capacity. All runoff passes through a groundwater reservoir of specified residence time, and then is summed over all grid cells in a river basin for calculation of river discharge. Evaporation is limited by a bulk stomatal resistance in series with the aerodynamic resistance and decreases below its maximum value as soil water decreases. Geographic variations of most land parameters are defined on the basis of their dependence on globally mapped soil and vegetation type. Seasonal and other temporal variations of land parameters are neglected. Part I evaluated the capability of the model to reproduce runoff ratios of a set of river basins for which precipitation is well known. Part II showed that the use of geographically varying information on land characteristics contributes to the capability of the model to reproduce observations.

b. Data

Model forcings for our investigations are constructed from four distinct information sources. The first of these is the ISLSCP Initiative I dataset, already mentioned in the introduction. ISLSCP data are available for land areas on a global 1° grid, with a 6-h temporal resolution. The dataset includes all forcing variables needed as input to the model: precipitation, downward shortwave and longwave radiation, surface pressure, and near-surface atmospheric temperature, humidity, and wind speed. The dataset spans 1987 and 1988.

The second source of data is the precipitation dataset of MIDUa. Monthly precipitation estimates are given as areal-mean values over 175 specific river basins having a median area of 51 000 km². The period of record differs across basins, but generally the records span three or more decades and terminate in the 1980s or 1990s.

The third source of data is the Climate Prediction Center Merged Analysis of Precipitation (CMAP) (Xie and Arkin 1997). CMAP contains monthly precipitation estimates on a global 2.5° grid for the period beginning with calendar year 1979. The “enhanced” CMAP precipitation estimates that we use were produced by merging information from gauges, multiple satellite observations, and model-based atmospheric reanalyses.

The final source of data is the 8-yr dataset of the Surface Radiation Budget (SRB) project of NASA Langley Research Center. The SRB dataset provides monthly global analyses, on an irregular grid, of various surface radiation components. The estimates are based on satellite observations and parameterized broadband radiative transfer model calculations (Darnell et al. 1988; Gupta et al. 1992). The SRB dataset spans the period July 1983 through June 1991. These were converted to a regular 1° grid and averaged over the eight years to obtain monthly grids of long-term mean downwelling shortwave and longwave radiation.

In addition to the forcing datasets cited above, we use monthly observations of river discharge from the dataset of MIDUa to evaluate model accuracy.

c. Basin selection

Our analyses are performed at the river-basin scale. Starting with the 82 basins used in Part I, we exclude those basins that are characterized by one or both of the following shortcomings:

- The characteristic annual precipitation error would induce a large runoff-ratio error. A quantity Δ^* was defined in Part I as the apparent error in annual runoff ratio that would be caused by a characteristic error in basin-mean, annual precipitation, if the model were perfect. Here we exclude basins for which Δ^* is greater than 0.1. The purpose of this constraint is to minimize distortion of our model evaluation by erroneous input data.
- The basin climate is characterized by strong annual-mean aridity, interrupted by an intense wet season. Part I showed that large model errors in a small number of basins appear to be associated with neglect of upward soil–water diffusion into the root zone during the dry season. To avoid distortion of our analysis by this recognized model error, we used only basins for which the index Ψ defined in Part I is less than 40 kg m⁻² y⁻¹.

These conditions, along with two additional constraints on land-cover type, were used in Part II; the conditions

applied here result in a set of 44 basins, of which the 22 considered in Part II are a subset. The land-cover constraints are not applied here, because our main focus is not on land characteristics.

d. LaD model experiments

All experiments are conducted using the “TUNED” version of the LaD model described in Part I. (The tuning of the model was the adjustment of one globally constant scale factor applied to the relatively uncertain global field of non-water-stressed bulk stomatal resistance, in such a way as to minimize errors in mean annual river discharge for 1988.) Each model grid cell spans 1 degree of latitude and longitude. Forcing is specified as 6-h means (in manners described below), and these values are interpolated to hourly values; integration is performed on a 1-h time step.

The experiments can be divided into two sets. The first set of experiments is used to evaluate, in the framework of the model, the importance of information on temporal variability of forcing at various timescales. Specifically, these experiments are designed to evaluate the proposed method of construction of long-term, high-temporal-resolution forcing datasets suitable for long-term model experiments. The second set of experiments uses this “forcing modulation” method to run the model for a multiyear period in an attempt to reproduce observed water fluxes.

The first set of experiments consists of a control experiment (CTRL), five approximations thereof, and an additional reference experiment (Table 1). In CTRL, we run the model for 4 yr using the forcing for 1987 three times in a row, followed by the forcing for 1988; the repetition of 1987 is intended to provide a spinup period for the model. All four years are run using the 6-hourly forcing provided by ISLSCP without modification. The other six experiments are run only for 1 yr (1988), taking their initial condition from the end of the third year of CTRL; thus, all seven experiments have the same initial condition for 1988.

The synthesis of the 1988 forcing for all experiments in the first set is summarized in Table 1. CTRL87 is similar to CTRL, but uses the full 1987 ISLSCP forcing to simulate the year of interest; we use differences between CTRL and CTRL87 as a measure of interannual variability. The five experiments intended to approximate CTRL use only monthly or annual-mean information on forcing during 1988. In the ANN (annual mean forcing) experiment, the 1988 forcing for any input variable at any gridpoint is taken to be the annual mean value of that variable for 1988 in the ISLSCP dataset; this value is identical for every time step of the year. In the MON (monthly mean forcing) experiment, the model is forced instead by monthly means of the 1988 ISLSCP data; thus, each forcing variable remains constant for each time step of a given month, but steps to a new value at the start of a new month. In the AMOD

TABLE 1. Prescription of atmospheric forcing for numerical experiments. Six-hour:month ratio is the ratio of 6-h-mean forcing to monthly mean forcing. Month:annual ratio is defined similarly. For radiation, ISLSCP values were all adjusted using the 8-yr SRB data, as described in the text.

Experiment	Forcing variables	6-h:month ratio	Month:annual ratio	Annual mean
CTRL	All	ISLSCP 1988	ISLSCP 1988	ISLSCP 1988
CTRL87	All	ISLSCP 1987	ISLSCP 1987	ISLSCP 1987
ANN	All	=1	=1	ISLSCP 1988
MON	All	=1	ISLSCP 1988	ISLSCP 1988
AMOD	All	ISLSCP 1987	ISLSCP 1987	ISLSCP 1988
MMOD	All	ISLSCP 1987	ISLSCP 1988	ISLSCP 1988
MMODP	Precipitation	ISLSCP 1987	ISLSCP 1988	ISLSCP 1988
	All others	ISLSCP 1987	ISLSCP 1987	ISLSCP 1987
MD	Precipitation	ISLSCP 1987	MIDUa 1979–96	MIDUa 1979–96
	All others	ISLSCP 1987	ISLSCP 1987	ISLSCP 1987
CMAP	Precipitation	ISLSCP 1987	CMAP 1979–96	CMAP 1979–96
	All others	ISLSCP 1987	ISLSCP 1987	ISLSCP 1987

(annual modulation) and MMOD (monthly modulation) experiments, forcing variables change value every time step, but the variability at the time step timescale is defined using the 1987 ISLSCP data. In AMOD, the 6-hourly 1987 forcing is scaled by the ratio of 1988 to 1987 annual-mean values to obtain forcing for 1988. In MMOD, a similar scaling is performed, but the scaling ratio is defined on a monthly basis. MMODP is similar to MMOD, but the modulation is applied only to precipitation; other variables are set equal to their 1987 time step values.

The second set of experiments is a pair of 20-yr model runs (Table 1). These experiments are similar in design to MMODP. Forcing for all variables except precipitation is taken directly from the ISLSCP dataset for 1987. Precipitation forcing is formed by multiplying the 1987 ISLSCP values by the ratios of estimated monthly precipitation to the monthly totals of the 1987 ISLSCP values. Two sources are used for the estimated monthly precipitation. The first source is MIDUa, which is designated MD. For the experiment with this source, a scale factor is computed, for each month of each year, in a given basin, as the ratio of basin-mean amounts; that scale factor is applied at every grid point in the basin. Basin mean values are used because MIDUa analyzed only basin means. The second 20-yr experiment uses the CMAP precipitation dataset. Because these data are available on monthly grids, we formed distinct monthly scale factors at each grid point in the CMAP-forced run of the LaD model, which here is termed the CMAP run.

Interannual variability of radiation is ignored in the pair of 20-yr experiments. The radiation fields are based on the 1988 ISLSCP initiative I data, scaled on a monthly basis at each grid point to make monthly means consistent with the Surface Radiation Budget 8-yr means, as described in Part I. This adjustment was based on the assumption that the more recently produced 8-yr dataset provides a more representative estimate of long-term means than the 1988 ISLSCP fields.

e. Semiempirical water-balance calculations

Independently of the LaD experiments, interannual variations in river discharge are estimated by application of the semiempirical water-balance equation of Budyko (1974). Budyko's relation can be written

$$\bar{q}/\bar{p} = 1 - \phi(\bar{r}/\bar{p}), \quad (1)$$

where q is runoff, p is precipitation, r is net radiation expressed as equivalent evaporative flux, overbars denote long-term averages, and

$$\phi(x) = [x(\tanh x^{-1})(1 - \cosh x + \sinh x)]. \quad (2)$$

It is assumed that (1) applies not only to the long-term mean, but also for any single water year, and that interannual variability of radiative energy supply is negligible. For sufficiently small anomalies of precipitation δp_n in water year n , it follows that the resulting runoff anomaly δq_n is (MIDUa)

$$\delta q_n = [1 - \phi + (\bar{r}/\bar{p})\phi']\delta p_n, \quad (3)$$

where ϕ' is the derivative of ϕ . The coefficient of δp_n in (3) is the runoff sensitivity and can be evaluated as a function of the index of dryness, \bar{r}/\bar{p} , upon which ϕ depends.

In general, even with the use of a water year, runoff may not appear as discharge during its year of production. To account for this effect, we route δq_n through a simple linear reservoir to model, in a lumped fashion, all storage delays after runoff production by the root zone and before river discharge (Milly and Wetherald 2002, manuscript submitted to *Water Resour. Res.*, hereafter MIWE). This delay parameterization is identical to that used in the LaD model, and identical, basin-dependent residence times are used in the LaD model and in the processing of the Budyko-based estimates. We do not attempt to apply the Budyko model at a timescale shorter than 1 yr. Thus, it is assumed that the runoff anomaly is constant through the water year, with step changes across water years. However, its conversion to discharge, through the delay model, is performed

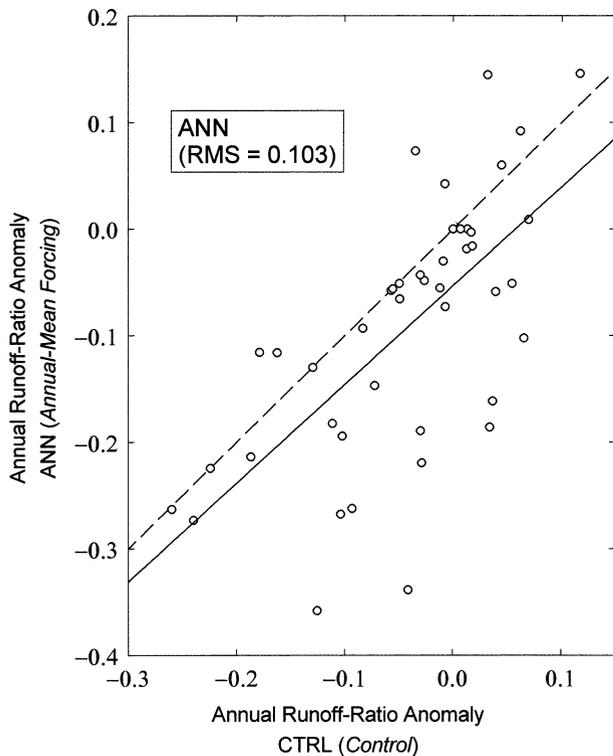


FIG. 1. Scatterplot comparing annual runoff ratio anomalies in the ANN run (annual-mean forcing) with those from the CTRL run. Dashed line is 1:1 line. Solid line is least squares fit. Rms is root-mean-square difference between values from the two runs.

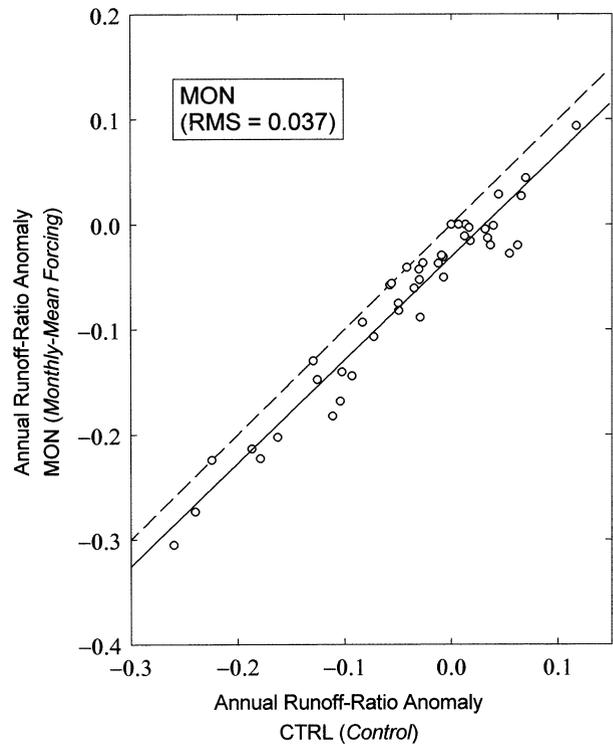


FIG. 2. Scatterplot comparing annual runoff ratio anomalies in the MON run (monthly mean forcing) with those from the CTRL run. Dashed line is 1:1 line. Solid line is least squares fit. Rms is root-mean-square difference between values from the two runs.

on a daily time step for consistency with the LaD treatment.

f. Defining annual runoff ratio anomalies

The term runoff ratio refers to the ratio of runoff to precipitation. For the first set of experiments, we define an annual runoff ratio anomaly for each basin as the difference in discharge between a given experiment and the CTRL87 experiment, normalized by the 1988 ISLSCP precipitation. Our evaluation of various forcing approximations is based on comparison of the annual runoff ratio anomalies for each experiment with those from the CTRL experiment.

For the second set of experiments, the anomaly is defined similarly. However, the base value for the modeled (or observed) anomalies is taken as the mean of the model output (or of the observations) over the period of available discharge observations for a given basin, instead of the CTRL87 output. For both MD and CMAP, the difference is normalized by the mean precipitation in MD, also computed over the period of available discharge observations; the choice of this normalization is arbitrary and does not bias the results against CMAP.

3. Results

a. Sensitivity of LaD-modeled interannual variations to resolved scale of temporal variability

Basin-mean annual runoff computed in the ANN run is compared with that computed in the CTRL run in Fig. 1. Any discrepancy is indicative of error induced by ignoring temporal variations in model input (forcing) at timescales less than 1 yr. Although the ANN run captures some of the variability of runoff anomalies, the scatter is large. Additionally, a significant negative bias is present, with most points below the 1:1 line.

Consideration of seasonal variations in forcing considerably improves model computations of runoff (Fig. 2). The MON run retains information on monthly mean variations in forcing but ignores the intramonthly variations present in the CTRL experiment. Both the bias and the scatter, though significant, are reduced considerably from the comparison in Fig. 1.

Results for the AMOD run, wherein 1988 is modeled using full, 6-hourly inputs from 1987, scaled simply so that their annual means are appropriate for 1988, are shown in Fig. 3. Use of this technique leads to a smaller bias than that present in the ANN run (Fig. 1). However, because the temporal structure of the 1988 precipitation anomaly is not supplied, significant random errors in runoff are present.

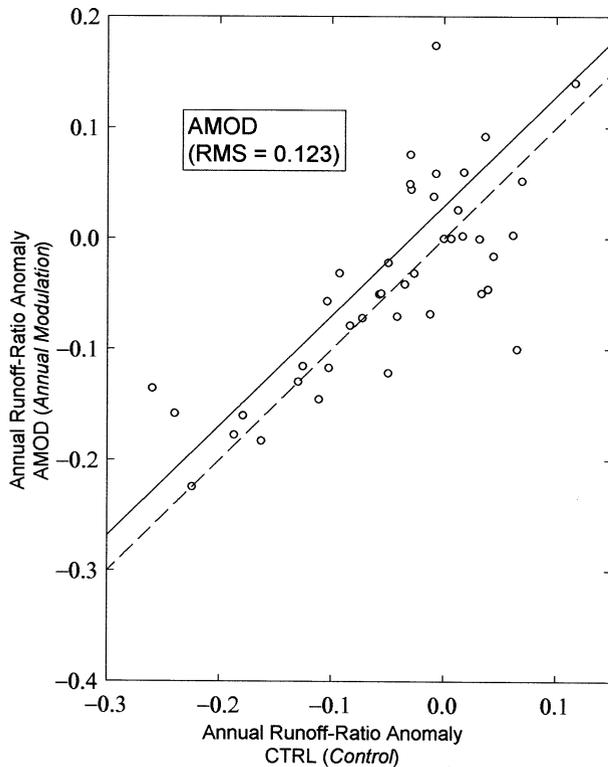


FIG. 3. Scatterplot comparing annual runoff ratio anomalies in the AMOD run (annual modulation) with those from the CTRL run. Dashed line is 1:1 line. Solid line is least squares fit. Rms is root-mean-square difference between values from the two runs. Two data points do not appear in this plot because their values were greater than 0.2 in the AMOD run.

When 6-hourly forcing having statistically realistic variability is used in conjunction with information on monthly variations in forcing (MMOD run), considerably better results are achieved than in the other cases (Fig. 4). Comparison of Fig. 4 with Fig. 2 shows the value of using 6-hourly forcing having realistic variability, even if that variability is unrelated to the true historical time series (i.e., from 1987 rather than 1988). The most obvious difference is the virtual removal of bias in the case of Fig. 4. Comparison of Fig. 4 with Fig. 3 shows the value of knowing the monthly distribution of annual forcing anomalies. The main effect is a reduction in scatter.

The runoff results for the MMODP run are shown in Fig. 5. In MMODP, precipitation is treated as in MMOD, and all other input variables are treated as in CTRL87. Thus, high-frequency variability is statistically realistic for all variables, but only the precipitation input contains (monthly) information specific to 1988. Results are similar to those for MMOD; the root-mean-square error of 0.027 in MMODP compares favorably with the value of 0.026 in MMOD. Similarity of MMOD and MMODP implies that precipitation is the dominant driver of interannual variability in the LaD model driven by ISLSCP forcing.

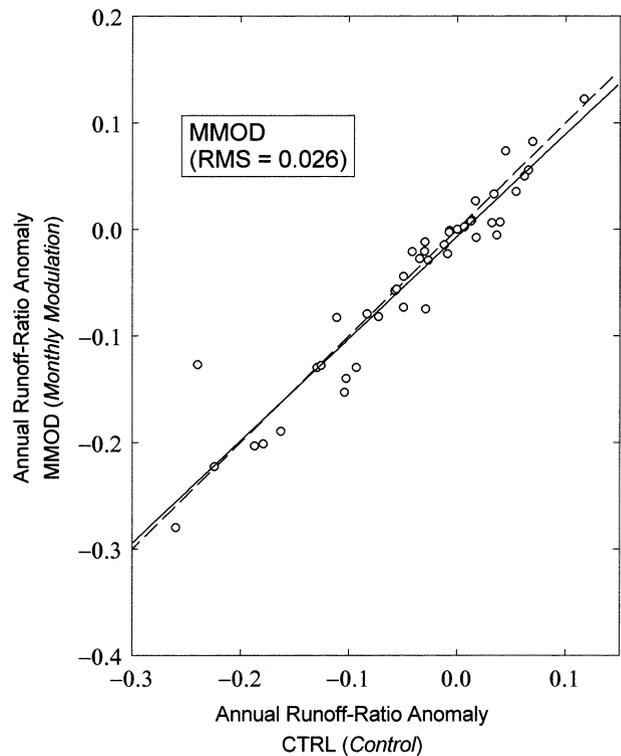


FIG. 4. Scatterplot comparing annual runoff ratio anomalies in the MMOD run (monthly modulation) with those from the CTRL run. Dashed line is 1:1 line. Solid line is least squares fit. Rms is root-mean-square difference between values from the two runs.

b. Observation-based evaluation of LaD-modeled interannual variations

Results in the preceding section are based only on model-model comparisons. Assuming that the LaD model and the ISLSCP forcing are sufficiently realistic, those results indicate an efficient strategy for modeling interannual variability: we may use the ISLSCP initiative I dataset to specify high-frequency variability of all forcing variables and supplement it with information only on monthly variations in precipitation. Here we test this approach against historical river discharge observations.

The annual runoff ratio anomalies computed in the MD run are compared to the observations in Fig. 6. The 0.054 rms error in runoff ratio means that the typical error in departure of annual discharge from its long-term mean in any given year is equal to about 5.4% of the long-term annual mean precipitation. Overall, the MD run explains about two-thirds of the variance in the observed runoff ratio anomaly. As seen in Fig. 7, however, the fraction of variance explained by the model is not distributed symmetrically across basins; r^2 values exceed 0.6 in 33 of the 44 basins.

Interquartile values of r^2 for the MD run are about 0.5, 0.75, and 0.85. We use these values to select representative basins for display of modeled and observed

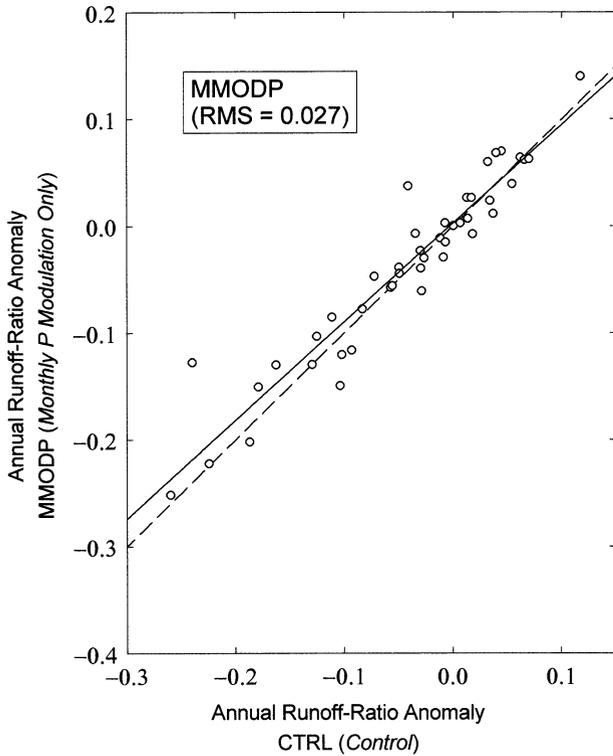


FIG. 5. Scatterplot comparing annual runoff ratio anomalies in the MMODP run (monthly modulation of precipitation only) with those from the CTRL run. Dashed line is 1:1 line. Solid line is least squares fit. Rms is root-mean-square difference between values from the two runs.

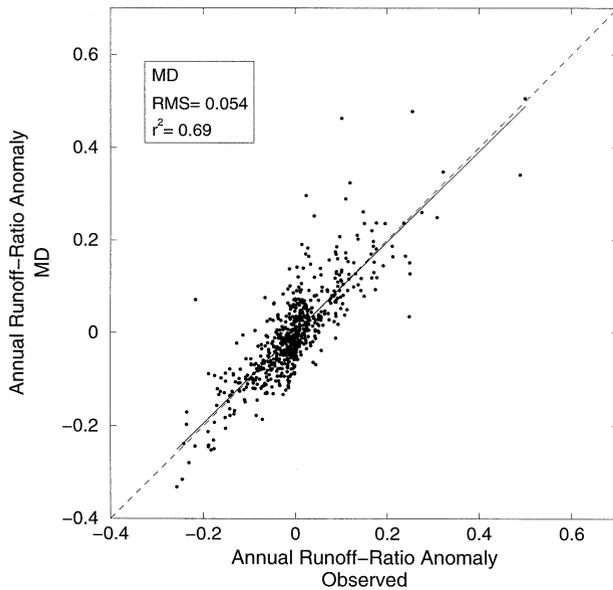


FIG. 6. Scatterplot comparing annual runoff ratio anomalies in the MD run with corresponding observations. Each symbol represents 1 yr (having discharge observations) in one basin. Dashed line is 1:1 line. Solid line is least squares fit. Rms is root-mean-square difference between model and observations.

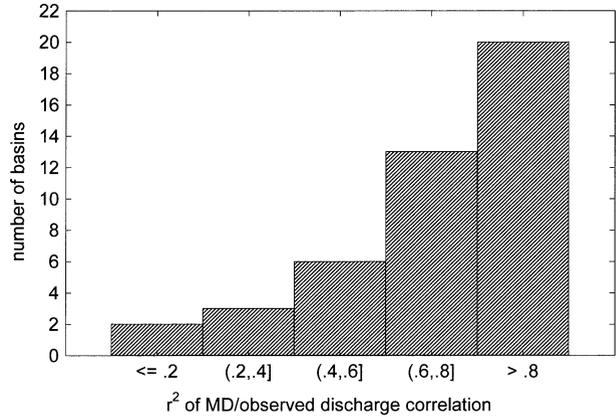


FIG. 7. Histogram of the square of the correlation between MD-modeled and observed annual discharge.

time series. We selected one arid basin and one humid basin with r^2 approximately equal to each of these three values. The results are shown in Figs. 8, 9, and 10. In the Amazon (Fig. 8), systematic errors are more obvious than the deficiencies in correlation, the mean runoff is underestimated significantly, and the interannual variability is overestimated. Interannual variability is overestimated also for the Powder River, with the model yielding near-zero flow in the drier years. The interannual variability is reproduced well for both the Nelson and Warta Rivers (Fig. 9). The mean runoff is reproduced well for the Nelson, but the model shows a positive bias for the Warta. Results for the Potomac and Humboldt Rivers are good, with the main deficiency being the underestimation of mean runoff from the Humboldt. As in the case of the Powder River, there is a tendency for modeled runoff to be zero during the drier years of the period of record.

c. Comparison of precipitation datasets

The performance of the LaD model forced by CMAP precipitation is illustrated in Fig. 11. Comparison with

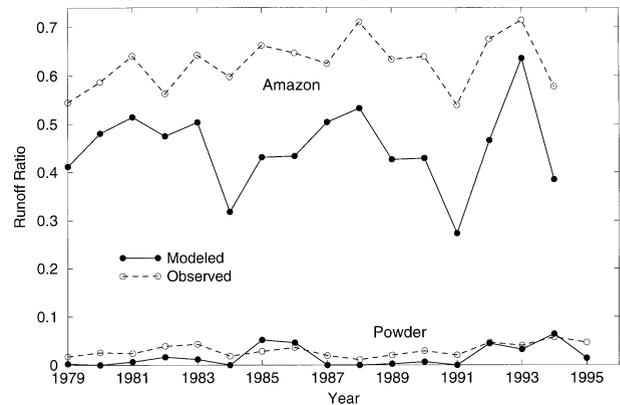


FIG. 8. MD-modeled and observed time series of annual discharge for the Amazon River at Manacapuru, Brazil ($r^2 = 0.50$), and the Powder River near Locate, Montana ($r^2 = 0.51$).

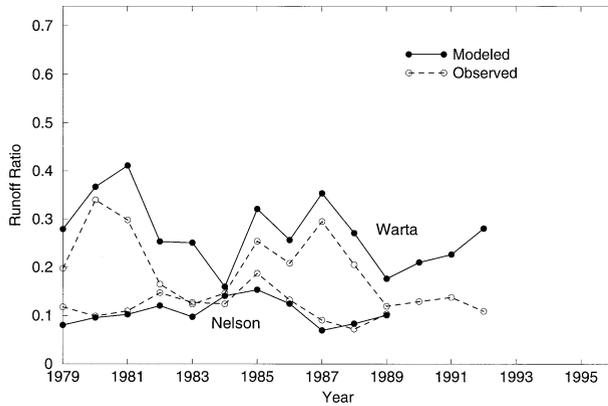


FIG. 9. MD-modeled and observed time series of annual discharge for the Warta River at Gorzów, Poland ($r^2 = 0.72$), and the Nelson River above Bladder Rapids, Manitoba ($r^2 = 0.73$).

Fig. 6 shows that the LaD model produces better results with the MD precipitation than with the CMAP precipitation. Correlation drops from 0.82 in MD to 0.68 in CMAP, and the rms error increases from 0.054 to 0.086.

It could be inferred from the comparison of MD and CMAP results that the MD precipitation data are more accurate than the CMAP data in the basins analyzed. On the other hand, it must be considered that the LaD model has been calibrated (Part I), to some degree, using the data of MIDUa, which form the basis for the MD runs here. To make a separate, simple assessment, independent of the LaD model, we computed, for each basin, the correlation between estimated precipitation anomalies and observed discharge anomalies. The precipitation estimates used in MD are consistently better correlated with observed discharge than are precipitation estimates used in CMAP (Fig. 12).

d. Comparison with performance of semiempirical relation

Finally, we compare the performance of the MD run (Fig. 6) with equivalent results from the much simpler

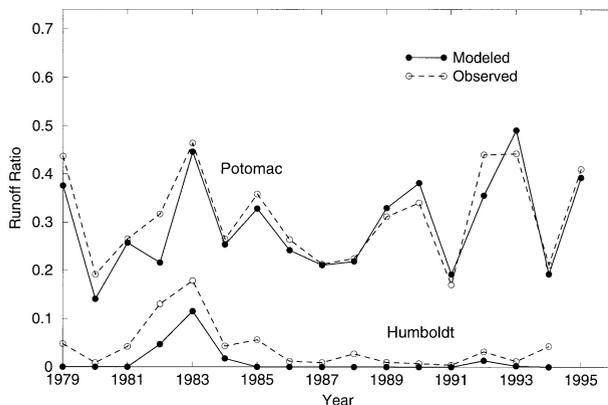


FIG. 10. MD-modeled and observed time series of annual discharge for the Potomac River at Point of Rocks, Maryland ($r^2 = 0.85$), and the Humboldt River at Comus, Nevada ($r^2 = 0.85$).

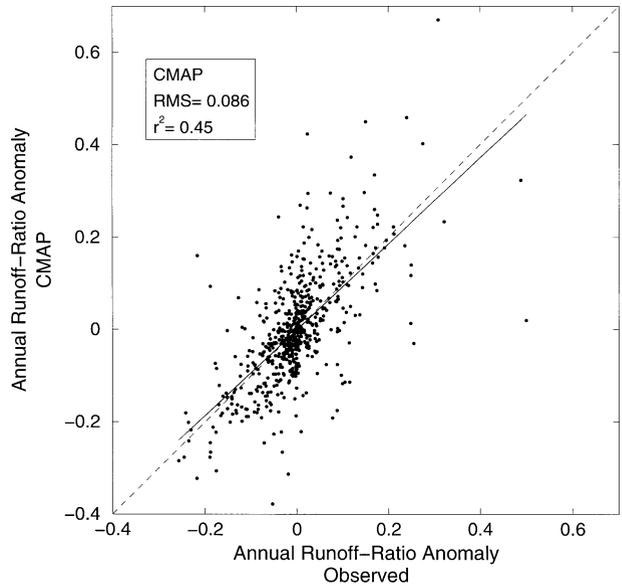


FIG. 11. Scatterplot comparing annual runoff ratio anomalies in the CMAP run with corresponding observations. Each symbol represents 1 yr (having discharge observations) in one basin. Dashed line is 1:1 line. Solid line is least squares fit. Rms is root-mean-square difference between model and observations.

semiempirical relation of Budyko (1974) (Fig. 13). The MD-forced LaD model produces a slightly higher rms error, a higher correlation, and a slope (model versus observed anomaly) closer to unity than does the Budyko relation. Overall, difference in performance is small.

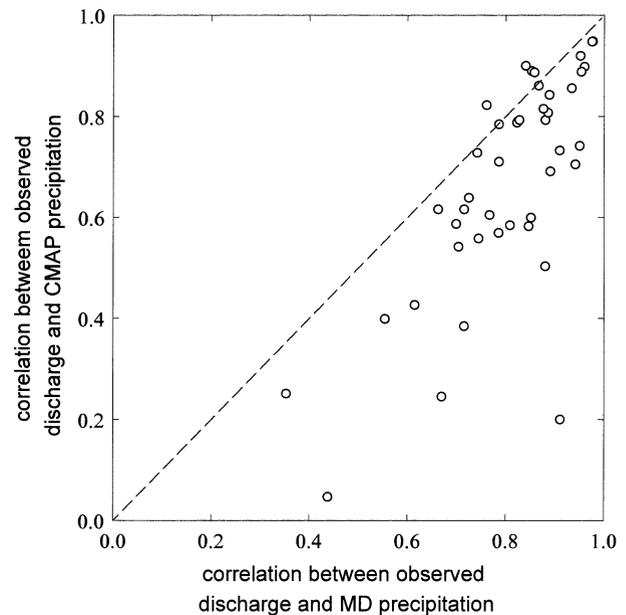


FIG. 12. Scatterplot comparing correlations of interannual precipitation variations (CMAP and MD) with same-year observations of discharge.

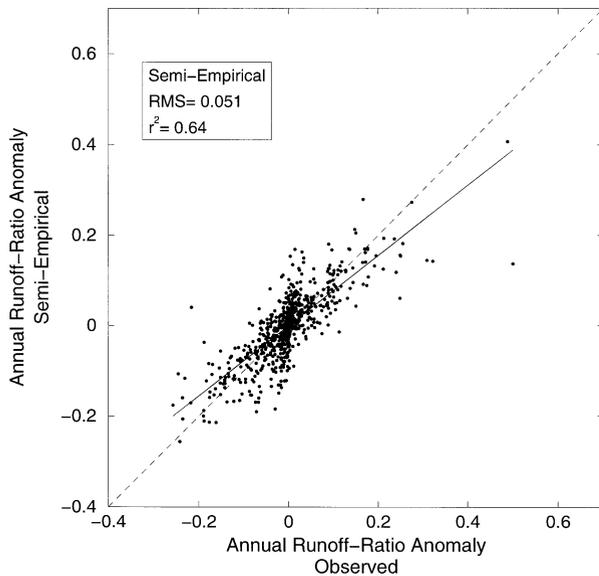


FIG. 13. Scatterplot comparing annual runoff-ratio anomalies from the semiempirical relation of Budyko (1974) with corresponding observations. Each symbol represents 1 yr (having discharge observations) in one basin. Dashed line is 1:1 line. Solid line is least squares fit. Rms is root-mean-square difference between model and observations.

4. Summary and discussion

a. Summary

We have tested the capability of the Land Dynamics (LaD) model to reproduce interannual variations in runoff. The model was tested by comparison with observational data for 44 large river basins. River basins were included in the analysis only if it appeared that estimates of precipitation were sufficiently accurate. A few basins were excluded from the analysis because they experience a climate that is believed to highlight soil water diffusion processes absent from the model. Overall, the model explained 67% of the variance of annual runoff ratio anomalies. In half of the basins, the model explained more than 75% of the variance.

The performance of the LaD model was compared to that of a simple relation based on Budyko's (1974) semiempirical water balance equation. This approach used only information on long-term radiation balance and annual precipitation amounts. In contrast, the LaD model used information on the seasonal distribution and typical 6-hourly variability of precipitation and several other atmospheric forcing variables. Furthermore, the LaD model used information on geographical variability of land characteristics. Overall, performance of the LaD model was similar to that of the semiempirical relation.

In order to carry out the LaD experiments, we developed a method for downscaling of long-term monthly precipitation data to the relatively short timescales necessary for running the model. The method merges the long-term data with a reference dataset of 1-yr duration,

having high temporal resolution. The success of the method was demonstrated in a model-model comparison and in the comparisons of modeled and observed interannual variations of runoff.

b. Lessons from the model-model comparisons

Comparisons among the ANN, MON, and CTRL outputs show that temporal variability at the monthly timescale and at shorter timescales is important for simulation of annual mean water and energy balances. Ignoring temporal variability tended to create a negative runoff bias; the bias was greater when annual mean forcing was used than when monthly mean forcing was used. This set of results is consistent with theoretical analysis of Milly (1994a), who pointed out that temporal variability tends to generate runoff by creating imbalances between the water and energy supplies that support evaporation. Use of average forcing removes such imbalances and enhances evaporation at the expense of runoff. Thus, the great reduction in bias from the ANN run to the MON run can be explained by the incorporation of seasonal variability and its effect on runoff in seasonal climates (Milly 1994b). The systematic bias still present in the MON run can be explained in terms of neglect of the random nature of storm arrivals (Milly 1993), which also contribute to the production of runoff.

Results from the AMOD and MMOD runs show that most of the effect of submonthly temporal variability on water balance can be captured without knowledge of the actual forcing time series. Presumably, as long as the temporal variability in the assumed forcing has the proper statistical characteristics, it will produce results similar to the actual forcing. Similarity between MMOD and MMODP runs suggests that even interannual variability of all variables other than precipitation is a very minor control on water balances. Monthly precipitation was the overwhelming control on interannual variability of water balance in the ISLSCP-forced LaD model. This finding supports the use of our monthly modulation method for modeling multiyear water and energy balances.

It must be kept in mind that the results of the model-model comparisons could be different if similar experiments were conducted using different models and/or different forcing. Unlike the LaD model, many other models have water stores of small capacity, such as canopy interception stores. Runoff generation in the LaD model is essentially a soil-store-excess mechanism, with no limitation on infiltration capacity. The ISLSCP forcing contains no variability at scales shorter than 6 h. Although we think that such approximations collectively do not distort the analysis greatly, it is not difficult to imagine certain combinations of models and forcings that might give qualitatively different results. For example, a model that generates its runoff mainly through the infiltration-excess mechanism might show much greater bias in the MON run and much more scatter in

the MMOD run, because runoff would be much more sensitive to high-frequency extremes of precipitation rates. In such a situation, our monthly modulation approach might not be as applicable as it was in this study. Thus, similar model–model tests should be conducted in other models before the technique is adopted for similar multiyear experiments.

c. Performance of MD-forced run

The model–model tests can only indicate the possibility of effective modeling approaches under the assumption that the model and forcing are useful approximations of reality. Therefore, testing against observations is a crucial complement to the model–model tests. We found that the apparent errors in comparisons with observations (e.g., rms MD–OBS difference in runoff ratio of 0.054) were greater than similarly defined MMODP–CTRL differences (rms difference of 0.027). This result was to be expected, because the MMODP–CTRL differences are theoretically a lower limit on MD–OBS errors. The MD–OBS differences, by design, should include errors similar to those of MMODP–CTRL, and also errors in model response, parameter values, and forcing.

d. Interannual variability of radiation

One interesting implication of the small difference between the MMOD and MMODP runs and the success of the MD run is the implied minimal effect of interannual variations of energy supply (surface net radiation) on water balances. This implication is consistent with a simpler data analysis by MIDUb, which did not detect an independent effect of radiation variability (although it was suggestive of a radiation influence correlative with precipitation). From a purely theoretical standpoint, we know there must be a contribution of radiation variability. However, it appears that this effect is so small that it is currently being hidden among various errors, including, possibly, significant errors in the SRB radiation forcing data.

e. Performance of CMAP-forced run

Our analyses suggest that the precipitation estimates of MIDUa may do a better job of capturing interannual variations than the CMAP estimates. However, the comparison has been made only in basins having good networks of precipitation gauges. A relative strength of the CMAP dataset is its blending of gauge information with other sources of information to produce globally complete estimates of precipitation. It is arguable that CMAP accuracy may surpass that of MD in regions of few gauge measurements. Still, the comparison made here is suggestive of potential for improvement of the CMAP precipitation estimation algorithms.

f. Performance compared to semiempirical relation

Koster et al. (1999) noted the apparent failure of current land water- and energy-balance models to perform better than (or, in many cases, as well as) a simple Budyko-type equation in their predictions of annual-mean quantities. We believe that the implicit challenge laid out by Koster et al. (1999) provides a meaningful performance measure for adoption by land modelers. Part II presented results indicating that the information on both geographic variations in land characteristics and high-frequency variability of forcing enabled the LaD model to exceed the performance of a simple semiempirical relation in predicting geographic variability of annual runoff. As noted in the summary above, the LaD model performed as well as, but not better than, the semiempirical model in the prediction of interannual anomalies of runoff. This result provides an interesting contrast to the success of Part II and leaves a worthy challenge to land modelers for the future.

Acknowledgments. Support for A.B.S. was provided by NASA's Water Cycle Processes Program through the University Corporation for Atmospheric Research. Helpful reviews were provided by Hiram Levy II and Michael Spelman.

REFERENCES

- Budyko, M. I., 1974: *Climate and Life*. Academic, 508 pp.
- Chen, T. H., and Coauthors, 1997: Cabauw experimental results from the Project for Intercomparison of Land-Surface Parameterization Schemes. *J. Climate*, **10**, 1194–1215.
- Darnell, W. L., W. F. Staylor, S. K. Gupta, and F. M. Denn, 1988: Estimation of surface insolation using sun-synchronous satellite data. *J. Climate*, **1**, 820–835.
- Foley, J. A., I. C. Prentice, N. Ramankutty, S. Levis, D. Pollard, S. Sitch, and A. Haxeltine, 1996: An integrated biosphere model of land surface processes, terrestrial carbon balance, and vegetation dynamics. *Global Biogeochem. Cycles*, **10**, 603–628.
- Gupta, S. K., W. L. Darnell, and A. C. Wilber, 1992: A parameterization of longwave surface radiation from satellite data: Recent improvements. *J. Appl. Meteor.*, **31**, 1361–1367.
- IGPO, 1999: ISLSCP Initiative II receives funding. *GEWEX News*, Vol. 9, No. 3, International GEWEX Project Office, 8.
- Koster, R. D., T. Oki, and M. J. Suarez, 1999: The offline validation of land surface models: Assessing success at the annual time-scale. *J. Meteor. Soc. Japan*, **77**, 257–263.
- Meeson, B. W., F. E. Coprew, J. M. P. McManus, D. M. Myers, J. W. Closs, K.-J. Sun, D. J. Sunday, and P. J. Sellers, 1995: *ISLSCP Initiative I—Global Data Sets for Land–Atmosphere Models, 1987–1988*, Vols. 1–5, NASA, CD-ROM, USA. NASA_GDAAC-ISLSCP.001/002/003/004/005.
- Milly, P. C. D., 1993: An analytic solution of the stochastic storage problem applicable to soil water. *Water Resour. Res.*, **29**, 3755–3758.
- , 1994a: Climate, soil water storage, and the average annual water balance. *Water Resour. Res.*, **30**, 2143–2156.
- , 1994b: Climate, interseasonal storage of soil water, and the annual water balance. *Adv. Water Resour.*, **17**, 19–24.
- , 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.

- , and —, 2002b: Global modeling of land water and energy balances. Part II: Land-characteristic contributions to spatial variability. *J. Hydrometeor.*, **3**, 301–310.
- van Dam, T., J. Wahr, P. C. D. Milly, A. B. Shmakin, G. Blewitt, D. Lavallée, and K. M. Larson, 2001: Crustal displacements due to continental water loading. *Geophys. Res. Lett.*, **28**, 651–654.
- Wahr, J., M. Molenaar, and F. Bryan, 1998: Time variability of the earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE. *J. Geophys. Res.*, **103**, 205–230.
- Xie, P., and P. A. Arkin, 1997: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc.*, **78**, 2539–2558.