Global Modeling of Land Water and Energy Balances. Part II: Land-Characteristic Contributions to Spatial Variability

P. C. D. MILLY

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

A. B. SHMAKIN*

NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey

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

ABSTRACT

Land water and energy balances vary around the globe because of variations in amount and temporal distribution of water and energy supplies and because of variations in land characteristics. The former control (water and energy supplies) explains much more variance in water and energy balances than the latter (land characteristics). A largely untested hypothesis underlying most global models of land water and energy balance is the assumption that parameter values based on estimated geographic distributions of soil and vegetation characteristics improve the performance of the models relative to the use of globally constant land parameters. This hypothesis is tested here through an evaluation of the improvement in performance of one land model associated with the introduction of geographic information on land characteristics. The capability of the model to reproduce annual runoff ratios of large river basins, with and without information on the global distribution of albedo, rooting depth, and stomatal resistance, is assessed. To allow a fair comparison, the model is calibrated in both cases by adjusting globally constant scale factors for snow-free albedo, non-water-stressed bulk stomatal resistance, and critical root density (which is used to determine effective root-zone depth). The test is made in stand-alone mode, that is, using prescribed radiative and atmospheric forcing. Model performance is evaluated by comparing modeled runoff ratios with observed runoff ratios for a set of basins where precipitation biases have been shown to be minimal.

The withholding of information on global variations in these parameters leads to a significant degradation of the capability of the model to simulate the annual runoff ratio. An additional set of optimization experiments, in which the parameters are examined individually, reveals that the stomatal resistance is, by far, the parameter among these three whose spatial variations add the most predictive power to the model in stand-alone mode. Further single-parameter experiments with surface roughness length, available water capacity, thermal conductivity, and thermal diffusivity show very little sensitivity to estimated global variations in these parameters. Finally, it is found that even the constant-parameter model performance exceeds that of the Budyko and generalized Turc–Pike water-balance equations, suggesting that the model benefits also from information on the geographic variability of the temporal structure of forcing.

1. Introduction

Global models of land water and energy balances translate precipitation and radiative fluxes and near-surface atmospheric states to land water and energy storage and effluxes. This translation is achieved by use of functional relations that include parameters characterizing land processes. These relations and parameters differ from one model to the next; however, all land models need parameters to describe, in some way, shortwave reflectance, soil water retention characteristics and plant root density profile (which collectively control water storage capacity for evaporation), surface control of vapor release to the atmosphere, and aerodynamic roughness of the surface. Originally, land models used global constants for most land parameters (e.g., Manabe 1969). Dickinson et al. (1981) and Sellers et al. (1986) initiated the practice of assigning values to parameters of global land water- and energy balance models on the basis of estimated global distributions of vegetation and soil characteristics, and this method has since become standard practice among climate modelers. Underlying this approach is an implicit assumption that the characteristics and the functional dependences are sufficiently well known to make this an improvement over the use of global constants.

* 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

Various investigators (e.g., de Rosnay and Polcher 1998; Zeng et al. 1998; Kleidon and Heimann 1998) have reported improvements in model performance associated with changes in methods for specification of land parameters. However, such studies, including that described in Part I (Milly and Shmakin 2002) of this series of papers, generally have not distinguished between improvements due to information on geographical variations of parameters and those due to better global mean information.

An indication of the immature state of land modeling has been provided in a revealing analysis by Koster et al. (1999). They compared the annual runoff produced by nine models to observations for 106 river basins worldwide. They used the same set of observations to test the simple semiempirical equation of Budyko (1974), which uses only annual precipitation and net radiation as inputs. Overall, no model performed substantially better than Budyko's equation, and most models performed much worse. The superior performance of Budyko's equation was found despite the fact that most or all of the models had the advantage of using information on the global distributions of surface characteristics. Furthermore, all of the models used information on the temporal variations of forcing at a 6-h timescale. In view of these findings, it is reasonable to question whether land models gain any advantage from use of globally distributed information. Koster et al. remarked that runoff values derived from Budyko's equation

> *can be considered crude yardsticks against which to measure success in offline* [*land model*] *validation exercises.* [*A land model*] *can perhaps be said to contribute to the realism of the annual surface water balance across the globe only if the errors it produces in an equivalent validation experiment are, on average, significantly less than* [*values obtained from a Budyko analysis*].

Having introduced globally variable land parameters into a land model (Part I), we seek here to determine to what extent, if any, the use of globally varying parameters benefits model performance. Our chosen measure of model performance is the degree of agreement between observed and modeled annual runoff ratios of large river basins. The runoff ratio is a simple integral measure of the partitioning of precipitation into evaporation and runoff and, hence, of the partitioning of net radiation into latent and sensible heat fluxes. Furthermore, runoff and precipitation are arguably the best and second-best measured major water or energy fluxes at the basin scale. Our focus on the annual timescale helps to separate the problem of runoff generation, our main interest here, from that of runoff routing. On a monthly timescale, model performance would be affected by details of the formulation of storage lags, but on the annual timescale, such effects essentially will be removed for most basins.

We assume at the outset that the detection of a land-characteristic signal in runoff response may be difficult. We do not expect, for example, that differences in runoff between a forest and a grassland under similar climatic conditions will be much greater than the noise introduced into our model by errors in precipitation. To minimize the distortion of the analysis by measurement errors, we use only river basins for which errors in precipitation are known to be small. The analysis is facilitated by use of the error-characterized basin dataset of Milly and Dunne (2002a, manuscript submitted to *Water Resour. Res.,* hereafter MIDUa).

2. Methodology

a. Model

The Land Dynamics (LaD) model has been described and tested in Part I. 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. Most land parameters are defined as globally varying fields, as described below.

The LaD model has nine parameters (of which the first two listed appear only as a product): effective depth of the root zone (Z_R), available water capacity (AWC), bulk heat capacity of the ground (C), thermal conductivity of the ground (λ), surface roughness length (z_o), non-water-stressed bulk stomatal resistance (r_s), groundwater residence time (τ), snowfree surface albedo (A_n), and snow-masking depth (W_S^*). Each land parameter is prescribed as a temporally constant function of soil type (S) (=1, 2, . . . , 9) and/or vegetation type (V) (=1, 2, . . . , 10) indices, which, in turn, are determined from global datasets. In general, for any parameter ω,

$$\omega_{ij} = F_\omega(S_{ij}) \quad \text{or} \quad \omega_{ij} = F_\omega(V_{ij}), \qquad (1)$$

where indices i and j denote dependence on longitude and latitude, respectively, and F_ω is a table of parameter values indexed by soil type or vegetation type. The geographic distribution of nine soil types (seven texture classes or combinations for mineral soils plus ice and organic soil) are specified following Zobler (1986). The distribution of 10 vegetation types (5 forest types, grassland, desert, tundra, agriculture, and ice) is obtained from Matthews (1983), whose 32 types were aggregated in Part I to form the LaD model types. Both fields are defined at 1° resolution.

b. Strategy

How can we assess the predictive value of global distributions of model parameters? First, we need to introduce some objective measure of agreement between the model and some set of observations. We then could compare the model performance using its global distribution of parameters to that with globally constant parameter values. The result of such a comparison, however, would undoubtedly depend upon the values chosen for the globally constant parameters. An objective and meaningful choice for the global constants would be those values that optimize the value of the chosen performance measure. Such an optimization would, however, give unfair advantage to the globally constant parameter set. To make a meaningful comparison, therefore, we should optimize both the globally constant and the globally varying parameter distributions, using an equal number of degrees of freedom in both optimizations.

To describe the strategy more precisely, let us consider first the case of any single parameter ω. In order to quantify the predictive power of the model, we define a measure of performance Φ (e.g., the root-mean-square difference between modeled and observed runoff). The globally constant ω field can be denoted by $f_\omega^C\omega_o$, where ω_o is some arbitrarily chosen global constant value of ω and f_ω^C is the tuning factor for ω in the globally constant parameter case. The globally variable ω field can be denoted by $f_\omega^V\omega_{ij}$, where f_ω^V is the analogous tuning factor for the globally variable parameter case. We define $f_\omega^C{*}$ as that value of f_ω^C where $\Phi(f_\omega^C\omega_o)$ is optimized, and $f_\omega^V{*}$ as that value of f_ω^V where $\Phi(f_\omega^V\omega_{ij})$ is optimized. The difference between the corresponding optima, $\Phi(f_\omega^V{*}\omega_{ij})$ and $\Phi(f_\omega^C{*}\omega_o)$, is our measure of the predictive value of globally distributed parameter estimates. For the case of multiple parameters, we still perform only two optimizations, one with all parameters globally constant, and the other with all globally variable. However, each optimization must be carried out in a multidimensional space of tuning factors, with one factor for each parameter.

The general approach outlined above is applied here. Because the computational magnitude of the problem grows rapidly with the number of parameters optimized, it is critical that we minimize that number within reason. In preliminary sensitivity studies, we have found that model results for annual runoff (and all other water and energy fluxes) are, under many conditions, highly sensitive to values of Z_R and r_s within the ranges over which they are believed to vary geographically. In general, annual runoff from the model is relatively insensitive to differences in the values of C, λ, z_o, τ, and W_S^*. Sensitivities to A_n and AWC are intermediate in magnitude. For our main analysis, we shall consider only the three parameters Z_R, r_s, and A_n. The other parameters will be allowed to follow their usual global distributions, described in Part I. Thus, we initially are investigating only the predictive power that (imperfect) knowledge of spatial distributions of Z_R, r_s, and A_n brings to the model. (Analyses involving the other parameters will be described below.)

For the optimization, our treatment of r_s and A_n is as described in the outline of the approach above. In contrast, Z_R is handled differently. The effective root-zone depth is given by (Part I)

$$Z_R = \zeta \ln(R_o/R_c), \tag{2}$$

where $R_o(V)$ and $\zeta(V)$ are parameters in the relation (Jackson et al. 1996)

$$R(z) = R_o e^{-z/\zeta} \tag{3}$$

that describes root biomass density as a function of depth z below the ground surface, and our parameter R_c is the critical root-biomass density at the bottom of the effective root zone, which is assigned a globally constant value of 0.5 kg m^{-3} (Part I). For optimization of both globally constant and globally varying parameter cases, we apply a globally constant scale factor to R_c. For the case of globally varying parameters we use $R_o(V)$ and $\zeta(V)$, and for the globally constant case we instead use constants for both of these fields. As for the parameters r_s and A_n, the choice of values for these constants is immaterial, as the tuning factor is allowed to vary freely to find the optimum. We chose to optimize with respect to R_c because we considered it more uncertain than either R_o or ζ, and because we wanted to minimize the dimensionality of the problem.

It will be informative to know not only the overall value of information on r_s, A_n, R_o, and ζ, but also the individual contributions of these parameters to predictive power. For this reason, we design three single-parameter comparisons. In each of these comparisons, all parameters except one (or two in the case of R_o and ζ) are kept spatially varying according to their usual model distributions. In each case, the corresponding single-scale factor is tuned to optimize model performance. Finally, we also perform similar single-parameter optimizations for the other model parameters: z_o, AWC, C, and λ/C. The complete series of analyses is summarized in Tables 1 and 2.

c. Quantifying model performance

Our measure of model performance Φ is the root-mean-square deviation between modeled and observed runoff ratios during the year 1988,

$$\Phi = \left[\frac{1}{K}\sum_{k=1}^{K}\left(\frac{y_{m,k} - \hat{y}_k}{\hat{p}_k}\right)^2\right]^{1/2}, \tag{4}$$

in which K is the number of basins with suitable observational data, and $y_{m,k}$, $\hat{y}_k$, and $\hat{p}_k$ are modeled and observed runoff and observed precipitation, respectively, all for basin k. Following Part I, we use the ‘‘hat’’ notation as a reminder that observations are subject to

TABLE 1. Summary of experimental design. Parameter assignments (V = variable and C = constant) are explained in Table 2. Whenever V is indicated for A_n, Z_R, and/or r_s, the global distribution includes application of the optimal scale factors from the all-var series. Here A_n is snowfree surface albedo, Z_R is effective depth of the root zone, r_s is non-water-stressed bulk stomatal resistance, z_o is surface roughness length, AWC is available water capacity, C is bulk heat capacity of the ground, and λ is thermal conductivity of the ground.

Series	Tuned parameters	A_n	Z_R	r_s	z_o	AWC	C	λ/C
All-var	A_n, Z_r, r_s	V	V	V	V	V	V	V
All-const	A_n, Z_r, r_s	C	C	C	V	V	V	V
A_n-const	A_n	C	V	V	V	V	V	V
Z_r-const	Z_r	V	C	V	V	V	V	V
r_s-const	r_s	V	V	C	V	V	V	V
z_o-const	z_o	V	V	V	C	V	V	V
AWC-const	AWC	V	V	V	V	C	V	V
C-const	C	V	V	V	V	V	C	V
λ/C-const	λ/C	V	V	V	V	V	V	C

TABLE 2. Parameter assignments. Parenthetic v and s indicate dependences on vegetation and soil distributions, respectively. (Subsequent to given formulas for Z_R, a restriction is applied that Z_R not be less than 0.01 m). In each row, f indicates a distinct tuning factor.

Parameter	V (variable)	C (constant)
A_n	$f_A A_n(v)$	$f_A \bar{A}_n$
Z_R	$\zeta(v) \ln[R_o(v)/f_Z R_c]$	$\bar{\zeta} \ln[\bar{R}_o/f_Z R_c]$
r_s	$f_r r_s(v)$	$f_r \bar{r}_s$
z_o	$z_o(v)$	$f \cdot \bar{z}_o$
AWC	AWC(s)	$f \cdot \overline{AWC}$
C	$C(s)$	$f \cdot \bar{C}$
λ/C	$(\lambda/C)(s)$	$f \cdot \overline{(\lambda/C)}$

error. We use the year 1988 in order to allow the use of the 1987–88 global forcing of the International Satellite Land Surface Climatology Project (ISLSCP) Initiative I CD-ROM; the model is run in stand-alone mode, forced by 6-hourly data. The first year allows spinup of the model, and the second year is used for model performance evaluation. The ISLSCP precipitation and radiation fields are adjusted for consistency with precipitation analyses of MIDUa and radiation estimates from the Surface Radiation Budget dataset, as described in Part I.

The set of basins used in (4) is a subset of those used in Part I for an initial evaluation of the LaD model. Mainly on the basis of data availability for 1988, the set of 175 basins of MIDUa was reduced to a set of 82 in Part I. Basins were excluded from the latter set for the present analysis if one or more of the following conditions were met:

- The characteristic annual precipitation error would induce a large runoff-ratio error. A quantity Δ^* is defined in Part I as the error in apparent 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. In Part I it was shown 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}. (A large value of Ψ implies both a very arid annual-mean climate and the presence of a strong seasonal excess of precipitation over evaporative energy supply.)
- The estimated area under cultivation, according to Matthews (1983), is more than 50% of the total basin area. This constraint was added because the model vegetation field is based entirely on estimated natural conditions, but vegetation type is a critical factor in the present analysis.
- The estimated area of deserts, according to Matthews (1983), is more than 50% of the total basin area. This constraint was used because LaD parameter specifications for desert are quite arbitrary and are not linked to vegetation characteristics in the model.

When all of these constraints were applied, the number of basins decreased to $K = 22$. These were, however, surprisingly well distributed across the major vegetation types that determine the assignment of parameter values (Table 3).

d. *Practical details*

For each experiment series, numerous experiments were run and the search for the optimum was done iteratively. Automated optimization techniques were not used, because the process of checking their performance for the limited number of series run would have required essentially the same amount of computation. For the all-var and all-const cases, we began with uniformly spaced values of f_A (scale factor for A_n) and logarithmically spaced values of f_r and f_Z (scale factors for r_s and Z_R, respectively). The initial coarse grid of Φ results indicated the approximate location of the optimum, and subsequent experiments were run to find the exact location and value of the optimum to a precision more than adequate to quantify differences among the various experiment series. To accelerate the search and to simplify the identification of the optimum, we constrained the optimum to be the minimum value of Φ located on the surface of zero mean error in runoff ratio across all basins. More detailed sensitivity analyses for specific cases suggested that this constrained value did not differ significantly from the global optimum. Approximately 60–70 experiments were run in both the all-var and all-const series.

TABLE 3. Information on individual basins in this study.

Dominant vegetation type	Symbol	River	Gauge location	Basin area (km^2)
Broadleaf evergreen forest	BE	Amazon	Manacapuru (Brazil)	2 300 000
		Magdalena	Puerto Berrío (Colombia)	74 000
Broadleaf deciduous forest	BD	Kanawha	Kanawha Falls (West Virginia)	22 000
		Ohio	Sewickley (Pennsylvania)	50 000
Mixed broadleaf and needleleaf forest	BN	Altamaha	Doctortown (Georgia)	35 000
		Danube	Orsova (Romania)	580 000
		Narva	Narva (Estonia)	56 000
		Neman	Smalininkai (Lithuania)	81 000
		Penobscot	West Enfield (Maine)	17 000
		Potomac	Point of Rocks (Maryland)	25 000
Needleleaf evergreen forest	NE	Northern Dvina	Ust'-Pinega (Russia)	348 000
		Rio Grande	Otowi Bridge (New Mexico)	37 000
		Vuoksi	Tainionkoski (Finland)	61 000
		Warta	Gorzów (Poland)	52 000
Grassland	G	Cooper	Callamurra (Australia)	230 000
		Fitzroy	Fitzroy Crossing (Australia)	45 000
		Fitzroy	The Gap (Australia)	136 000
		Niobrara	Spencer (Nebraska)	31 000
		Ob'	Salekhard (Russia)	2 900 000
		Paraná	Corrientes (Argentina)	1 950 000
		Pecos	Artesia (New Mexico)	40 000
		Powder	Locate (Montana)	34 000

For the single-parameter optimizations, all factors were set initially at the optimum values determined in the all-var series. Subsequently, one factor (corresponding to the globally constant parameter field) was given 3–5 values, Φ was computed for each value, the location of zero mean error in runoff ratio was determined, and the corresponding value of Φ was noted.

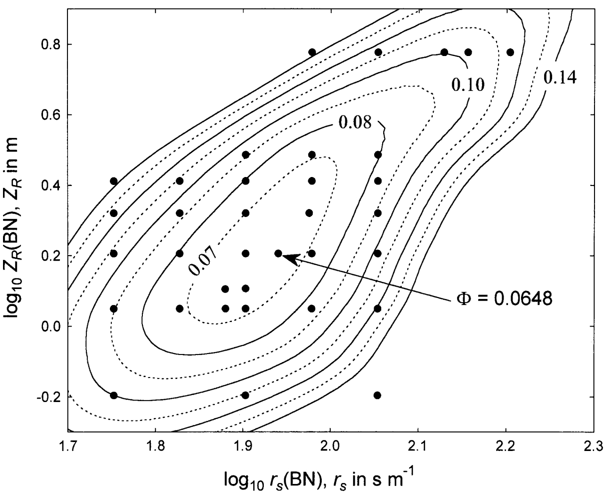

FIG. 1. Contour plot of the performance measure Φ, for the all-var case, as a function of non-water-stressed bulk stomatal resistance and critical rooting density of a mixed broadleaf/needleleaf (BN) forest, for a snowfree albedo-scale factor of 1. Parameters for all vegetation types were varied simultaneously as described in the text; BN was chosen simply to allow familiar physical quantities, rather than the scale factors, as the plotted variables.

3. Results

The dependence of Φ on stomatal resistance and rooting depth, for a fixed level of snowfree albedo, is illustrated in Fig. 1. The Φ surface is well behaved, and there is no difficulty estimating the minimum level of Φ. With a snowfree albedo scale factor of 1 (the case in Fig. 1), the minimum value of Φ is 0.065. For albedo factors of 0.5 and 1.5, the minimum is 0.066. Thus, the overall minimum is estimated to be 0.065.

The optimal adjustment factors for A_n, r_s, and R_c in the all-var case were found to be about 1.0, 0.44, and 0.25, respectively. The last of these three values implies $R_c = 0.125$ kg m^{-3}. The fourfold decrease in R_c from the initial estimate causes only a small change in typical values of rooting depth derived therefrom; resultant Z_R values are mostly in the 1–1.5-m range, depending on vegetation type. The reduction of r_s by a factor of 0.44 is similar to the result of crude tuning in Part I, in which other parameters were not allowed to vary. This represents a significant reduction in our a priori values of r_s, which were obtained from Dorman and Sellers (1989). We have speculated that at least part of the discrepancy may be associated with the lack of evaporation from an interception store and from wet ground in the model (Part I). The absence of these pathways for evaporation in the model would be compensated, in the tuned model, by an artificial reduction in the resistance to vapor flux through foliage.

For the all-const series, optimized globally constant parameter values for A_n, Z_R, and r_s are 0.12, 0.74 m, and 67 s m^{-1}, respectively. These are values consistent with median values obtained in the all-var calibration.

TABLE 4. Optimized model performance for each experiment series and performance of semiempirical water-balance equations. Here Φ is a root-mean-square difference between observed and modeled runoff ratios, given by (4).

Experiment or equation	Φ
All-var	0.065
All-const	0.095
A_n-const	0.065
Z_r-const	0.072
r_s-const	0.098
z_o-const	0.062
AWC-const	0.068
C-const	0.065
λ/C-const	0.065
Modified Turc–Pike	0.107
Budyko	0.110

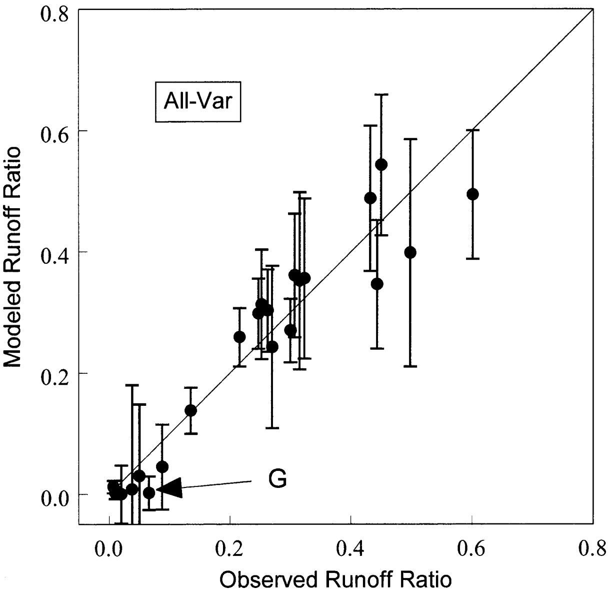

FIG. 2. Scatterplot of modeled against observed runoff ratio in 1988, for all-var. Each symbol represents one of the 22 basins having suitable data. Error bars represent $\pm 2\Delta^*$, where Δ^* is the error in apparent runoff ratio that would be caused by the characteristic error in basin-mean annual precipitation, if the model were perfect. For one predominantly grassland basin (indicated by G), the model results differ by more than $\pm 2\Delta^*$ from the observations.

We noted that the general topology of the Φ function did not differ qualitatively between the all-var and all-const cases.

Of main interest for this study are the optimal values of Φ determined in each experimental series (Table 4). As already noted, the optimal value of Φ for all-var is 0.065. This value implies that the 1988 basin runoff in the model typically differs from the observed runoff by 6.5% of estimated 1988 precipitation. When the major vegetation parameters are constrained to be globally constant, with tuning (all-const), Φ degrades seriously to a value of 0.095. This value implies that over half of the error variance in the latter case can be explained by the geographical variation of vegetation parameters in the former case.

The relative contributions of A_n, Z_R, and r_s to the difference between all-var and all-const results can be inferred from results of the single-parameter calibrations (Table 4). For the A_n-const case, the optimal value of Φ is identical to that for all-var. Thus, when albedo was prescribed as a global constant, the model output was not degraded. For Z_R-const, we obtained an optimal Φ value of 0.072. This value indicates some decrease in model performance from the all-var case, implying that spatial information on rooting depth may contribute to model performance. The most significant degradation in model performance for a single-parameter series was in the r_s-const case. When stomatal resistance was made to be globally constant, the optimal Φ was 0.098, approximately equal to the all-const value. This result implies that the dominant difference between all-var and all-const is that associated with information on global distribution of r_s.

The degradation of model results from all-var to all-const is illustrated by comparison of Figs. 2 and 3. In the all-var case (Fig. 2), only one of the 22 basins has a modeled runoff ratio that differs from the observation by more than $\pm 2\Delta^*$. In all-const, calibrated model runoff ratio differs from the observations by more than $\pm 2\Delta^*$ in seven of the basins. The two largest underestimates of runoff occur for basins in which the vegetation is predominantly broadleaf deciduous forest. As seen in Fig. 4, prescription of globally constant r_s leads to an underestimate of r_s in broadleaf deciduous forests, causing the model to allow excessive evaporation, and thus suppressing runoff. Conversely, the largest overestimate of runoff in the calibrated all-const experiment is for a basin covered by broadleaf evergreen forest, consistent again with Fig. 4.

The apparent lack of sensitivity of Φ to information on the global distribution of albedo was investigated further. For the four forest types represented in this analysis (Table 3), A_n ranges only from 0.11 to 0.13 in all-var, so changing to a constant value of 0.12 in A_n-const causes little difference in absorbed shortwave radiation of forests; hence, little change in runoff ratio results. However, the grassland A_n changes from 0.20 in all-var to 0.12 in A_n-const. This change implies a 10% increase in absorbed solar radiation under snowfree conditions. The corresponding change in net all-wave radiation typically was 15% for grassland (larger than 10% because of a relatively unchanged net loss of longwave radiation between the two experiments). This significant change in energy supply, however, produced almost no change in the modeled runoff ratio. The failure of increased energy supply to convert to decreased runoff is easily understood. The runoff ratio already was smaller than 0.03 in most grassland basins in the all-var experiment, so any further reductions induced by enhanced energy availability inevitably would be small compared to typ-

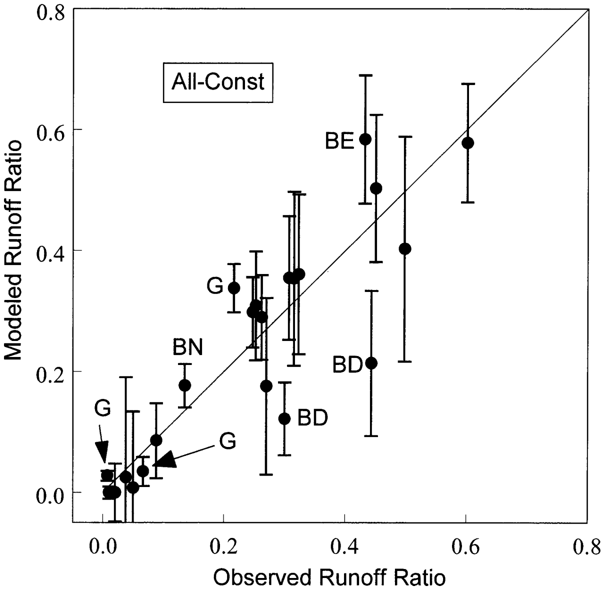

FIG. 3. Scatterplot of modeled against observed runoff ratio in 1988, for all-const. Each symbol represents one of the 22 basins having suitable data. Error bars represent $\pm 2\Delta^*$, where Δ^* is the error in apparent runoff ratio that would be caused by the characteristic error in basin-mean annual precipitation, if the model were perfect. Labeled points are those for which model output differs by more than $\pm 2\Delta^*$ from observations: BD, broadleaf deciduous forest; BE, broadleaf evergreen forest; BN, mixed broadleaf and needleleaf forest; G, grassland.

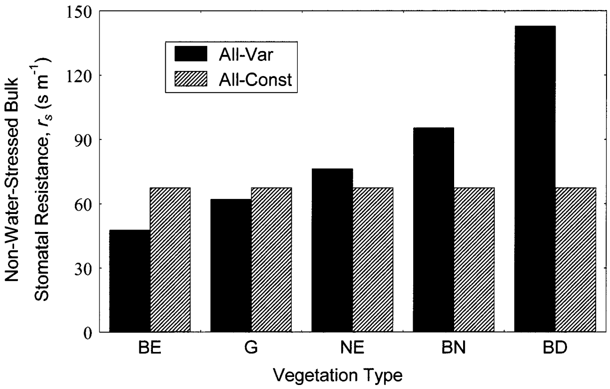

FIG. 4. Calibrated values of non-water-stressed bulk stomatal resistance as a function of vegetation type.

ical differences between observed and modeled runoff ratios. Thus, the value of the optimal Φ could change only very little between all-var and A_n-const.

The results of the single-parameter calibrations for z_o, AWC, C, and λ/C are given in Table 4. None of the optimal Φ values for these cases differs much from the optimal all-var value of Φ. We conclude that information on spatial variability of these parameters adds no predictive power to the model in its present form.

In the spirit of the investigation by Koster et al. (1999), mentioned in the introduction, we compared the LaD model performance to that of Budyko's (1974) equation. We also considered a more general equation of the same basic type as that of Budyko (Turc 1954; Pike 1964; Choudhury 1999; Milly and Dunne 2002b, manuscript submitted to *Water Resour. Res.*, hereafter MIDUb),

$$\frac{y}{p} = 1 - \left[1 + \left(\frac{p}{r}\right)^{\upsilon}\right]^{-1/\upsilon}, \qquad (5)$$

where y is annual runoff, p is annual precipitation, r is surface net radiation divided by the latent heat of vaporization of water, and υ is a fitting parameter. We used (5) in addition to Budyko's equation, because it allows for tuning of υ to our dataset, maximizing the performance of this approach. For both the Budyko and the generalized Turc–Pike equations, we used values of r that had been produced by the optimal all-var model experiment.

The performance of the Budyko equation, the generalized Turc–Pike equation, and the LaD model runs is summarized in Table 4. Using the Budyko equation to predict runoff, we found a value of 0.110 for Φ. With the generalized Turc–Pike equation, where we chose υ to minimize the sum of squares of deviations in runoff ratios, the optimal value of Φ was only marginally better, at 0.107; the optimal value of υ was 2.26, similar to values found in previous investigations (MIDUb). Both the Budyko and the Turc–Pike Φ values are larger than the all-const (0.095) optimum obtained with the LaD model. It is reasonable to infer that the superior performance of the LaD model, even for the case of globally constant parameters, may be attributable to knowledge of the subannual temporal variability of forcing. The inferiority of the optimized Turc–Pike equation, relative to the all-const case, is weakly evident also in Fig. 5, which shows nine basins (compared to seven in the all-const case) for which observations differ by more than $\pm 2\Delta^*$ from predictions of the equation.

4. Summary and discussion

a. Summary

We have tested the hypothesis that information on spatial variations in land characteristics contributes to the capability of a land model to reproduce observed patterns of annual water and energy balances. To this end, we compared observations to model outputs from experiments with and without globally variable parameters. In order to make a fair comparison, the model was calibrated for both cases. The measure of performance of the model was the goodness of fit of the calibration. The performance measure was defined in terms of river discharge. Discharge can be accurately measured, and it is indicative of water and energy balances in general, because it is highly correlated with evaporation, latent heat flux, and sensible heat flux for given precipitation and radiative forcings.

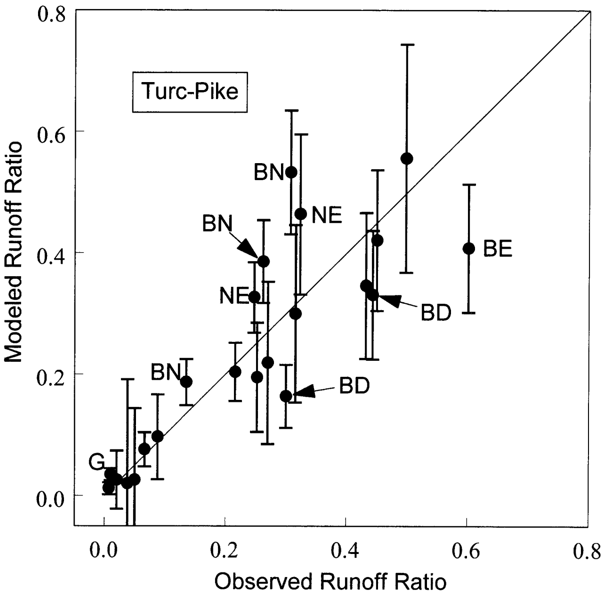

FIG. 5. Scatterplot of modeled against observed runoff ratio in 1988, for generalized Turc–Pike Eq. (5) with $\upsilon = 2.26$. Each symbol represents one of the 22 basins having suitable data. Error bars represent $\pm 2\Delta^*$, where Δ^* is the error in apparent runoff ratio that would be caused by the characteristic error in basin-mean annual precipitation, if the model were perfect. Labeled points are those for which model output differs by more than $\pm 2\Delta^*$ from observations: BD, broadleaf deciduous forest; BE, broadleaf evergreen forest; BN, mixed broadleaf and needleleaf forest; G, grassland; NE, needleleaf evergreen forest.

We find that knowledge of spatial variations in land characteristics does improve model performance. For the LaD model introduced in Part I, with a carefully selected subset of the dataset of MIDUa, the analysis showed that the use of globally variable land characteristics approximately halved the error variance of the annual runoff ratio. The major contribution to this reduction is associated with the vegetation-based specification of global variations of non-water-stressed bulk stomatal resistance. Parameters that define the rooting depth also appeared to make a minor contribution; it is possible that the relatively weak influence of rooting depth in our results is more indicative of our poor knowledge of its distribution than of model insensitivity to rooting depth. For other parameters, globally variable parameter estimates yielded no detectable improvement in model performance over globally constant values.

Additionally, the performance of the model was compared to the performance of simple equations that relate runoff to annual precipitation and surface net radiation. The model performed better than the simpler equations, even when globally constant model parameters were used, but especially when globally variable parameters were used. This ordering of model performance is consistent with the types of information provided as input. The most accurate results are obtained using information on both spatial variations in parameters and temporal structure of forcing; accuracy decreases when the spatial information is withdrawn; and accuracy decreases further when temporal information is withdrawn.

b. Land model versus Budyko-type equation

The results described above are a departure from earlier findings for land models. Koster et al. (1999) voiced the logical expectation that land models should be able to perform better than a simple Budyko-type equation, but presented results implying that they could not. They implicitly posed a challenge in discussing the results of their analysis:

> *As* [*land models*] *become more complex and realistic, they will presumably reach a point at which their generated runoff rates are indeed more accurate than those produced by* [*Budyko's*] *equation. When this happens, we can say that the explicit physics in the* [*model*] *does indeed contribute to the realism of the simulated energy and water budgets at the annual time scale.*

It would be gratifying to claim that the LaD model is the first model to meet this challenge, but we believe some other models would probably perform equally well or better in a similarly designed experiment. We believe that our positive results in this regard are explained mainly by the level of attention given to assessment and control of errors in precipitation. As argued by MIDUa, this is a crucial issue in the evaluation and future development of land models. It appears that the results of Koster et al. (1999) may have contained a substantial distortion of results due to nonnegligible errors in precipitation forcing, despite their reasonable efforts to eliminate this factor. The present success is also partially a result of excluding from analysis several seasonally arid basins, for which the LaD model and similar models are believed to be conceptually flawed. Additionally, the strategy for tuning of the model may have contributed to the success of the experiment.

c. Intrinsic model errors, forcing errors, and parameter errors

The error in a variable computed by a model is determined by the intrinsic errors in the model formulation, the errors in the parameters, and the errors in the forcing. These undoubtedly interact in nonlinear ways, but for purpose of discussion they may be considered to be independent sources of error. When the error associated with any one of these factors (model, parameters, or forcing) is much greater than the error associated with the others, the most fruitful efforts to reduce total model error probably will be those that attack the largest error source.

We believe insufficient characterization and control of errors in model forcing, especially precipitation, is currently the ''bottleneck,'' or limiting factor, in the rigorous testing of land models and, hence, in their further development (Milly 1994). When the forcing data used for model testing contain errors larger than those intrinsic to the model, any improvements in model formulation or parameter specification will be masked by the data errors. Similarly, in model intercomparisons, the differences between models of low and high accuracy can be hidden by the common error in shared forcing data. Forcing errors also can mask errors associated with errors in parameters, leading a globally constant-parameter model to perform no worse than a model with accurate global distribution of parameters. On the basis of results presented here and in Part I, we believe more attention should be directed toward detailed characterization of errors in model forcing as an integral part of the land model evaluation process.

For the sake of completeness, we acknowledge that errors in measurements of the ''output'' variables used for model testing also must be taken into consideration. When models, parameters, and forcing are all more accurate than the measurements with which the models are being tested, then it is time to improve the measurements. In our application, we have focused on discharge observations. Probably, errors in these measurements generally are much smaller than errors in the models, forcing, and parameters. This fortunate state of affairs is brought about by the natural spatial integration of river basins. The situation presently is very different for observationally derived estimates, at the spatial scale of the model grid, of the various components of water storage: snowpack, soil water, groundwater, and surface water. Each of these factors can be measured accurately at a point, but intense spatial variability prohibits easy extrapolation to the spatial scales resolved by models.

d. *Stand-alone analysis*

In interpreting and applying the conclusions of this analysis, is it crucial that the experimental design be kept in mind. Most importantly, it should be recalled that the model was run in stand-alone mode, that is, with prescribed forcing by precipitation, downwelling radiation, and near-surface atmospheric state. In so isolating the land for analysis, we have cut potentially important feedback links, which could amplify or attenuate any surface changes through induced changes in the forcing. It is reasonable to suggest that a similar experiment, if it allowed for atmospheric feedback, might have results that differ quantitatively or even qualitatively from our results.

In particular, we note that the sensitivities to stomatal resistance with prescribed forcing may be much greater than those with atmospheric feedbacks. When evaporation is suppressed by high stomatal resistance, the atmospheric boundary layer is warmed and dried, thereby enhancing the tendency for evaporation. This enhancement implies that a stand-alone design is more sensitive than an atmosphere-coupled design to r_s variations and, hence, is a good starting point for the type of analysis we have introduced in this paper. On the other hand, it also implies that the benefit of knowing the spatial pattern of r_s in a climate model (and, correspondingly, the degradation of accuracy when r_s variations are unknown) might not be so great as implied by the present analysis.

e. *Focus on annual timescale*

We reiterate here a comment made in Part I, noting that the conclusions of these papers are limited by their focus on the annual timescale. Physical processes of land water and energy exchange operate over a wide spectrum of timescales. While acknowledging the importance of many timescales, we find it helpful to approach the problem of data analysis and model development with an initial focus on the annual timescale. Nevertheless, it should be noted that the focus on annual timescale will tend to affect specific aspects of the results. In particular, it would not be surprising to find that the relative importance of individual parameters might change if the study were conducted using seasonally varying parameters as inputs and seasonal runoff as a performance measure.

Acknowledgments. Krista Dunne (USGS) provided invaluable technical assistance and advice. 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 Tony Gordon and Robert Tuleya.

REFERENCES

Budyko, M. I., 1974: *Climate and Life.* Academic, 508 pp.

Choudhury, B. J., 1999: Evaluation of an empirical equation for annual evaporation using field observations and results from a biophysical model. *J. Hydrol.,* **216,** 99–110.

de Rosnay, P., and J. Polcher, 1998: Modelling root water uptake in a complex land surface scheme coupled to a GCM. *Hydrol. Earth Syst. Sci.,* **2,** 239–255.

Dickinson, R. E., J. Jaeger, W. M. Washington, and R. Wolski, 1981: Boundary subroutine for the NCAR global climate model. National Center for Atmospheric Research Tech. Note TN-173+1A, 75 pp.

Dorman, J. L., and P. J. Sellers, 1989: A global climatology of albedo, roughness length and stomatal resistance for atmospheric general circulation models as represented by the Simple Biosphere Model (SiB). *J. Appl. Meteor.,* **28,** 833–855.

Jackson, R. B., J. Canadell, J. R. Ehleringer, H. A. Mooney, O. E. Sala, and E. D. Schulze, 1996: A global analysis of root distributions for terrestrial biomes. *Oecologia,* **108,** 389–411.

Kleidon, A., and M. Heimann, 1998: Optimised rooting depth and its impact on the simulated climate of an Atmospheric General Circulation Model. *Geophys. Res. Lett.,* **25,** 345–348.

Koster, R. D., T. Oki, and M. J. Suarez, 1999: The offline validation of land surface models: Assessing success at the annual timescale. *J. Meteor. Soc. Japan,* **77,** 257–263.

Manabe, S., 1969: Climate and the ocean circulation. 1. The atmospheric circulation and the hydrology of the earth's surface. *Mon. Wea. Rev.,* **97,** 739–774.

Matthews, E., 1983: Global vegetation and land use: New high-resolution data bases for climate studies. *J. Climate Appl. Meteor.,* **22,** 474–487.

Milly, P. C. D., 1994: Climate, soil water storage, and the average annual water balance. *Water Resour. Res.,* **30,** 2143–2156.

——, and A. B. Shmakin, 2002: Global modeling of land water and energy balances. Part I: The Land Dynamics (LaD) model. *J. Hydrometeor.,* **3,** 283–299.

Pike, J. G., 1964: The estimation of annual run-off from meteorological data in a tropical climate. *J. Hydrol.,* **2,** 116–123.

Sellers, P. J., Y. Mintz, Y. C. Sud, and A. Dalcher, 1986: A Simple Biosphere Model (SiB) for use within atmospheric general circulation models. *J. Atmos. Sci.,* **43,** 505–531.

Turc, L., 1954: Le bilan d'eau des sols: Relations entre les précipitations, l'évaporation et l'écoulement. *Ann. Agron.,* **5,** 491–595.

Zeng, X., Y.-J. Dai, R. E. Dickinson, and M. Shaikh, 1998: The role of root distribution for climate simulation over land. *Geophys. Res. Lett.,* **25,** 4533–4536.

Zobler, L., 1986: A world soil file for global climate modeling. National Aeronautics and Space Administration Tech. Memo. 87802, 33 pp.