WATER RESOURCES RESEARCH, VOL. 30, NO. 7, PAGES 2143–2156, JULY 1994

Climate, soil water storage, and the average annual water balance

P. C. D. Milly

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

Abstract. This paper describes the development and testing of the hypothesis that the long-term water balance is determined only by the local interaction of fluctuating water supply (precipitation) and demand (potential evapotranspiration), mediated by water storage in the soil. Adoption of this hypothesis, together with idealized representations of relevant input variabilities in time and space, yields a simple model of the water balance of a finite area having a uniform climate. The partitioning of average annual precipitation into evapotranspiration and runoff depends on seven dimensionless numbers: the ratio of average annual potential evapotranspiration to average annual precipitation (index of dryness); the ratio of the spatial average plant-available water-holding capacity of the soil to the annual average precipitation amount; the mean number of precipitation events per year; the shape parameter of the gamma distribution describing spatial variability of storage capacity; and simple measures of the seasonality of mean precipitation intensity, storm arrival rate, and potential evapotranspiration. The hypothesis is tested in an application of the model to the United States east of the Rocky Mountains, with no calibration. Study area averages of runoff and evapotranspiration, based on observations, are 263 mm and 728 mm, respectively; the model yields corresponding estimates of 250 mm and 741 mm, respectively, and explains 88% of the geographical variance of observed runoff within the study region. The differences between modeled and observed runoff can be explained by uncertainties in the model inputs and in the observed runoff. In the humid (index of dryness <1) parts of the study area, the dominant factor producing runoff is the excess of annual precipitation over annual potential evapotranspiration, but runoff caused by variability of supply and demand over time is also significant; in the arid (index of dryness >1) parts, all of the runoff is caused by variability of forcing over time. Contributions to model runoff attributable to small-scale spatial variability of storage capacity are insignificant throughout the study area. The consistency of the model with observational data is supportive of the supply-demand-storage hypothesis, which neglects infiltration excess runoff and other finite-permeability effects on the soil water balance.

1. Introduction

A basic problem of hydrology is to describe and explain the geographic and interannual variability of the annual water balance, i.e., the splitting of precipitation into evapotranspiration and runoff [*Brutsaert*, 1982, pp. 241–243]. It has long been observed that annual evapotranspiration approaches annual precipitation in regions where the annual input of energy to the surface (as measured by the potential evapotranspiration) greatly exceeds the amount needed to vaporize the annual precipitation (segment A in Figure 1). Conversely, where energy input is a small fraction of the necessary amount, the annual evapotranspiration approaches the annual potential evapotranspiration (segment B in Figure 1). These asymptotic relations under arid and humid conditions are understood as situations *limited* by the annual supply of water and energy, respectively. Evapotranspiration from most land areas is less than both the water and energy limits, as indicated in Figure 1 by curve C, which shows *Budyko*'s [1974] empirical fit to a large number of experimental points. The departure from the asymptotes (about 30% of precipitation) is large in the region where the aridity index is approximately unity.

A physical theory describing the departure of curve C from asymptotes A and B in Figure 1 has not been advanced in the hydrologic literature. One might ask why the transition region from A to B exists at all or why it is not larger. Scatter around curves such as C is considerable [*Budyko*, 1974, p. 326]. Are there factors other than the index of dryness that affect the water balance in a systematic way? Indeed, *Budyko and Zubenok* [1961] have noted that deviations from curve C tend to be positive when precipitation and potential evapotranspiration have seasonal variations that are in phase and negative when they are out of phase. No quantification or physical explanation of this and other effects on annual water balance has been given.

Asymptotes A and B define the upper water and energy limits on evapotranspiration and therefore the minimum possible runoff from a basin. The task that remains therefore is to explain runoff in excess of this minimum. Two related characteristics of land surfaces, finite water storage capacity and finite permeability, may be identified as possible causes

Paper number 94WR00586.

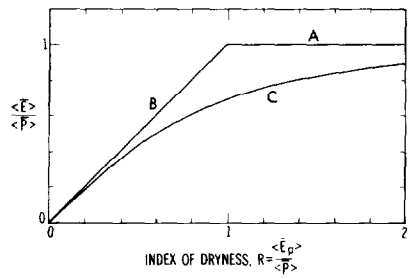

Figure 1. Diagram of average annual water balance, showing evapotranspiration ratio as a function of the index of dryness. $\langle \bar{E} \rangle$ is evapotranspiration, $\langle \bar{P} \rangle$ is precipitation, and $\langle \bar{E}_p \rangle$ is potential evapotranspiration. Curve C is *Budyko*'s [1974] relation, and segments A and B are its asymptotes.

of such additional runoff. For example, if the water storage capacity of soil is too small, temporary excesses of water supply will be lost as runoff, even though the index of dryness of the area may exceed 1. Finite-permeability effects enter in two main ways, both of which tend to increase runoff at the expense of evapotranspiration. If precipitation rates exceed the rates at which water can infiltrate the soil, then runoff will occur regardless of the long-term water and energy supplies. Second, if potential evapotranspiration rates exceed the rates at which water within the root zone can flow the short distances to the plant roots (or to the surface of bare soils), then evapotranspiration may fall below its water and energy supply limits.

The relative importance of finite-capacity and finite-permeability factors in determining the annual water balance has not been analyzed in the literature. *Eagleson* [1978a, b, c] formulated the water balance problem in terms of the Richards equation for soil water, allowing soil permeability to enter the problem. The only allowance for seasonality of climate was the definition of wet and dry seasons, with no water fluxes during the latter; interseasonal water storage was effectively neglected; and intraseasonal (i.e., within-season) storage was implicitly assumed to be unlimited by any capacity. The relationship of Eagleson's work to the empirical work described above has not been explored.

Some aspects of the problem of annual water balance, including seasonality, have been analyzed in the framework of simple models by *Schaake and Liu* [1989] and by *Dooge* [1992]. They used single-store water balance models having a small number of empirical parameters meant to represent unresolved physical processes, including spatial and intraseasonal variabilities. These studies were carried out in a spirit of simplicity and exploration, and it is understandable that some crucial aspects are questionable. In particular, both studies assumed that outflow from the reservoir of water available for evapotranspiration is a linear function of storage and is observable in stream discharge. Physical theory and field evidence suggest, however, that the plant root zone drains in a highly nonlinear way and that stream base flow is more closely tied to water storage in the saturated zone, which has a longer time constant. Given the strong nonlinearity of root zone drainage, it is arguably better described by the threshold concept of a soil water field capacity.

Some recent work has explored the possibility that annual water balance may be explained mainly by the finite-capacity model, with no restriction on infiltration. *Milly* [1993] treated the special case in which storminess (that is, random intraseasonal variability of precipitation) is the only form of hydrologic variability over time, and *Milly* [1994] dealt with the case in which seasonality is the only source of variability over time. In neither case did it appear to be possible to explain all of the observed large-scale runoff.

This paper is a continuation of earlier investigations [*Milly*, 1993, 1994] of annual water balance. The working hypothesis is that water balance can be described as the simple interaction of water supply, demand, and storage in a finite soil water reservoir. Variabilities in time associated with both seasonality and storminess are now included, as is spatial variability of storage capacity. Finite-permeability effects on water balance are ignored. Although the framework for the analysis is oversimplified by most current standards of hydrologic modeling, it nevertheless seems to be the appropriate approach for the scientific question at hand. By explicitly resolving intraseasonal, interseasonal, and spatial variabilities, the approach avoids the introduction of empirical parameters. The conceptual simplicity of the approach allows the development of analytic solutions for particular cases and facilitates the nondimensionalization of the problem even in its most general case. This work follows *Milly* [1993] in exploiting the idea of *Eagleson* [1978a] that statistical integration across hydrologic events can describe the average water balance.

Section 2 presents the framework for the study, which consists of a simple nonlinear conceptual model of water storage and compact representations of the variability over time and space of atmospheric forcing and land surface characteristics. Section 3 describes a dimensionless, Monte Carlo solution of the general water balance problem and presents several exact analytic solutions applicable under certain limiting conditions. In section 4, the theory is tested in an application to a large land area. Effects of input uncertainties are estimated, and the causation of runoff by different types of variability is evaluated. In the concluding section, the results of the study are summarized, some conclusions are drawn, and important limitations are noted.

2. Theoretical Framework

2.1. Overview and Notation

This paper addresses the long-term average water balance of a finite land area under a stationary climate. The starting point for the analysis is a storage model of the local (1- to 100-m horizontal length scale) water balance of the plant root zone of the soil. The average balance for this local scale is obtained by integration of a continuous water balance equation with respect to time. This integration recognizes both the deterministic, seasonal variability and the random, intraseasonal variability of atmospheric forcing of the surface. The areal mean of the long-term balance is subsequently determined by integration in space, using a distribution function to describe spatial variability of surface characteristics and ignoring spatial variability of climate.

The lowercase symbols p, e_p, and y denote the instantaneous, local values for precipitation, potential evapotranspiration, and root zone drainage (discharge) for runoff production, respectively. (We shall refer to y simply as runoff, but it should be kept in mind that actual river or groundwater discharge would lag in time relative to y because of subsurface storage outside the control volume of the model.) These variables fluctuate on daily (and shorter) timescales because

of the random nature of weather and also have underlying seasonal cycles. The corresponding uppercase symbols P, E_p, and Y denote the local expected values for these variables; they vary smoothly through the seasons. Angle brackets (for example, $\langle Y \rangle$) are used to denote the areal average value of a quantity over some area A larger than the local control volume. The overbar (for example, $\bar{Y}$) denotes an annual average.

2.2. Physical Idealization at Local Scale

The control volume for the water balance is bounded above by the soil-atmosphere interface and has sufficient vertical extent to contain essentially all of the water readily available to vegetation for uptake and transpiration. It is assumed that the vegetation cover is sufficiently extensive that direct evaporation from the soil need not be considered. The vertical extent corresponds approximately to the average depth of rooting of the predominant plants, which is usually of the order of 1 m. The horizontal extent of the control volume is sufficiently large to allow the divergence of horizontal root zone water fluxes, induced by soil heterogeneity and topographic curvature, to be neglected in the water balance. Such flows have not been well characterized by hydrologists, but a subjective estimate of the necessary horizontal scale might be of the order of 1 m (the scale of the pedon of soil science) in relatively flat terrain to 100 m or more (the scale of a hillslope) in sloping terrain. The mass balance of water for such a control volume, expressed in terms of equivalent liquid water depth and volumetric flux rates, is

$$\frac{dw}{dt} = i - e - y, \tag{1}$$

where w is the depth of water stored, i is the rate of infiltration of liquid precipitation and snowmelt into the soil, e is the rate of uptake and transpiration by plants (herein referred to as evapotranspiration), and y is the efflux, or drainage, of water from the control volume.

It is assumed that (1) the soil is sufficiently permeable to allow all liquid precipitation (p_l) and snowmelt (m) to infiltrate, (2) all soil water stored at potentials greater than the permanent wilting point of the vegetation [*Hillel*, 1980] is readily depleted at the potential evapotranspiration rate (e_p), (3) all water stored in excess of a well-defined field capacity [*Hillel*, 1980] is rapidly removed from the control volume by drainage (requiring that any water table or other downstream control of soil moisture be sufficiently far removed from the control volume to be ignored), and (4) no drainage occurs when the average soil moisture content is less than the field capacity. Under this set of assumptions, and with the convention that w includes only plant-available water, equation (1) becomes

$$\begin{aligned} &\frac{dw}{dt} = 0 \qquad p > e_p \text{ and } w = w_0 \\ &\frac{dw}{dt} = 0 \qquad p < e_p \text{ and } w = 0 \\ &\frac{dw}{dt} = p - e_p \qquad \text{otherwise} \end{aligned} \tag{2}$$

where p is the total liquid water supplied to the surface from above (the sum of p_l and m) and w_0 is the plant-available water-holding capacity of the soil (henceforth called storage capacity).

The analysis in the remainder of this paper holds as long as p is understood to represent the liquid input. However, we shall refer to p simply as the precipitation, and in the application to observational data, we shall use precipitation data to estimate p. This means that we implicitly ignore the phenomena of frozen precipitation, snowpack, and snowmelt. This limitation of the study is discussed briefly at the close of the paper.

For a well-developed vegetation cover the storage capacity w_0 may be interpreted as a depth integral over the root zone of the difference between the volumetric moisture contents of the soil at field capacity (θ_f) and at the wilting point ($\theta_{w'}$). Let $r(z)$ denote the fraction of area at depth z that is affected by the root system of the vegetation; in principle, this fraction depends in a complicated way on the rooting density, the hydraulic properties of the soil, and the timescale (i.e., seasonal versus storm/interstorm) of the uptake process. Then

$$w_0 = \int_0^\infty [\theta_f(z) - \theta_{w'}(z)] r(z)\, dz. \tag{3}$$

Typically, it is assumed that $r(z)$ steps down from 1 to 0 at some well-defined rooting depth Z_r, in which case

$$w_0 = \int_0^{Z_r} [\theta_f(z) - \theta_{w'}(z)]\, dz \tag{4}$$

2.3. Variability Over Time

A simple parametric representation of $P(t)$ and $E_p(t)$ is employed. The dominant mechanism that controls the seasonality of climate is the seasonality of the solar irradiance normal to the top of the atmosphere. At extratropical locations, this produces a strong signal with a dominant period of 1 year in most climatic variables, as shown, for example by *Trewartha* [1968, pp. 42–46, 156–161]. Therefore we assume

$$P(t) = \bar{P}(1 + \delta_P \sin \omega t) \tag{5}$$

$$E_p(t) = \overline{E_p}(1 + \delta_E \sin \omega t) \tag{6}$$

where δ_P and δ_E are the ratios of the amplitudes of the annual harmonics to the annual averages of P and E_p, respectively. With $2\pi/\omega$ equal to 1 year, these expressions capture the essential features of the annual land surface hydrologic forcing outside the tropics. Near the equator the noon Sun passes overhead twice per year, and these representations may be used with a period of one-half year.

The random component of variability of atmospheric forcing must also be described. It is assumed that precipitation arrives in discrete events that we shall call storms, that the arrival of these storms in time is a Poisson process, and that the amount of precipitation in any storm is governed by the exponential distribution [*Benjamin and Cornell*, 1970]. The mean storm arrival rate is allowed to vary seasonally

[*Todorovic and Yevjevich*, 1969], with only the annual harmonic retained:

$$N(t) = \bar{N}(1 + \delta_N \sin \omega t). \tag{7}$$

The expected value of storm depth at any time of year is simply P/N. The random variability of potential evapotranspiration is ignored, so e_p is identical to E_p.

In the application presented later, seasonal extrema of $P(t)$, $N(t)$, and $E_p(t)$ are assumed to occur at the ends of January and July, with the time origin at the end of April. It must be acknowledged that (5), (6), and (7) are rather crude approximations. Later in this paper, water balance solutions computed using these approximations are tested against similar solutions based on actual monthly climatologies for P, N, and E_p.

2.4. Variability Over Space

When an area A of horizontal scale significantly larger than the control volume for (2) is considered, it is desirable to average the solution of (2) spatially. In general, spatial variability of both climatic and soil factors must be considered. In this analysis, however, the variability over space of the statistics of point atmospheric forcing is ignored. Essentially, this amounts to an assumption of uniform climate but not uniform weather within the area A. Thus the local values of P, E_p, and N are everywhere equal to their areal means, $\langle P \rangle$, $\langle E_p \rangle$, and $\langle N \rangle$.

It is well known, however, that soil hydraulic characteristics vary greatly at relatively small scales [*Warrick and Nielsen*, 1980]. The nonlinear dependence of water balance on w_0 suggests the need for explicit consideration of spatial variability [*Milly and Eagleson*, 1987]. It is assumed here that the frequency distribution of water-holding capacity within A, $f_w(w_0)$, is given by the gamma distribution. This distribution satisfies the physical requirement that values be nonnegative, is very flexible, and is well suited for analytical work. Thus

$$f_w(w_0) = \frac{\lambda(\lambda w_0)^{\kappa-1} e^{-\lambda w_0}}{\Gamma(\kappa)}, \tag{8}$$

where the mean value of w_0 is given by κ/λ and its coefficient of variation is given by $\kappa^{-1/2}$. The spatial mean of any function $Z(w)$ (such as evapotranspiration or runoff) can be found by integrating over the density function (8), using the relation

$$\langle Z \rangle = \int_0^\infty Z(w) f_w(w)\, dw. \tag{9}$$

3. Theoretical Analysis

An analytic solution of the general water balance problem formulated above has not been found. However, when the variability over time of atmospheric forcing is ignored or is limited either to its seasonal component or to its random, intraseasonal component, then analytic solutions can be derived. In this section we first outline a method of construction of the general solution of the water balance problem by simulation; the dimensionless form of this solution is also noted. We then present the analytic solutions for the three special cases in their dimensionless forms.

For the sake of brevity, solutions only for the evapotranspiration ratio (ratio of local or areal mean long-term average evapotranspiration to long-term average precipitation, $\bar{E}/\bar{P}$ or $\langle \bar{E} \rangle / \langle \bar{P} \rangle$) are presented in this section. The solutions for runoff ratio, similarly defined, follow immediately from the fact that the two fractions must sum to 1.

3.1. General Solution by Monte Carlo Simulation

A general solution for the local water balance under the presented assumptions can be constructed by simulation. Equation (2) is integrated forward in time numerically from an arbitrarily chosen initial condition on w, using the descriptions of $p(t)$ and $e_p(t)$ already given. In general, many years of integration are required to remove the effect of arbitrary initial conditions and to remove the noise associated with the randomness of $p(t)$. The duration of the initial transient period can be found by sensitivity analyses. Following this initial period, the integration is continued for a sufficient number of years to obtain the associated water balance to any desired degree of precision. It can be shown that the resulting solution may be expressed in dimensionless form as

$$\frac{\bar{E}}{\bar{P}} = F(R,\ W,\ \bar{N}\tau,\ \delta_P,\ \delta_N,\ \delta_E) \tag{10}$$

where τ is $2\pi/\omega$, R is a climatological index of dryness analogous to that of *Budyko* [1974],

$$R = \overline{E_p}/\bar{P}, \tag{11}$$

and W is the dimensionless ratio of the storage capacity to the expected precipitation amount for one period (1 year, or 1/2 year near the equator),

$$W = \frac{w_0}{\bar{P}\tau}. \tag{12}$$

The simulation approach may be readily generalized to include spatial variability of w_0. The integral (9) may be expressed in terms of dimensionless W as

$$\langle Z \rangle = \int_0^\infty Z(W) f_w(\bar{P}\tau W)(\bar{P}\tau\, dW). \tag{13}$$

This integral can be approximated numerically as

$$\langle Z \rangle = \frac{1}{n} \sum_{i=1}^{n} Z(W_i), \tag{14}$$

where the W_i correspond to values w_i of w_0 that are located at the centers of n equal-probability sections of $f_w(w)$,

$$\int_0^{w_i} f_w(w)\, dw = \frac{i - \frac{1}{2}}{n}. \tag{15}$$

The integration over time already described, and represented by (10), is thus repeated n times, and the results are averaged according to (14); any desired degree of precision can be obtained by using a sufficiently large n. The final solution can be written

$$\frac{\langle \bar{E} \rangle}{\langle \bar{P} \rangle} = G(R, \langle W \rangle, \bar{N}\tau, \delta_P, \delta_N, \delta_E, \kappa), \quad (16)$$

where κ is the shape parameter introduced in (8) and $\langle W \rangle$ is given by $\langle w_0 \rangle / \bar{P}\tau$.

3.2. Analytic Solution in Absence of Variability Over Time ($\delta_P = \delta_N = \delta_E = 0$; $\bar{N}\tau \to \infty$)

For the situation in which neither $p(t)$ nor $e_p(t)$ varies in time, evapotranspiration will equal the lesser of p and e_p regardless of the soil water-holding capacity and its spatial variability. Thus

$$F(R, W, \infty, 0, 0, 0) = \min [1, R] \quad (17)$$

$$G(R, \langle W \rangle, \infty, 0, 0, 0, \kappa) = \min [1, R]. \quad (18)$$

These solutions correspond to the asymptotes A and B in Figure 1. They specify the maximum possible evapotranspiration (and minimum possible runoff) for a given index of dryness R.

3.3. Analytic Solutions in Absence of Seasonal Cycle ($\delta_P = \delta_N = \delta_E = 0$)

Milly [1993] addressed the special case in which $\kappa \to \infty$ without a seasonal cycle and obtained

$$F(R, W, \bar{N}\tau, 0, 0, 0) = \frac{[e^{W\bar{N}\tau(1-R^{-1})} - 1]}{[e^{W\bar{N}\tau(1-R^{-1})} - R^{-1}]}. \quad (19)$$

The insertion of (19) and (8) into (13), followed by integration, yields the spatial average runoff. (Details of the integration are not given here, but it should be noted that a necessary formula,

$$\int_0^\infty \frac{x^{\nu-1}e^{-\mu x}}{1 - \beta e^{-x}}\, dx = \Gamma(\nu) \sum_{n=0}^{\infty} (\mu + n)^{-\nu} \beta^n, \quad (20)$$

which is valid for certain combinations of values of β, μ, and ν, is given incorrectly by the standard reference *Gradshteyn and Ryzhik* [1980, p. 325]. The correct formula shown here is given by *Erdelyi* [1953].) The result of integration is

$$G(R, \langle W \rangle, \bar{N}\tau, 0, 0, 0, \kappa) = 1 - (1 - R) \cdot \sum_{n=0}^{\infty} [1 + n\langle W \rangle \bar{N}\tau(R^{-1} - 1)\kappa^{-1}]^{-\kappa} R^n \quad (21a)$$

$$R < 1$$

$$G(R, \langle W \rangle, \bar{N}\tau, 0, 0, 0, \kappa) = 1 - (1 - R^{-1}) \cdot \sum_{n=0}^{\infty} [1 + (n+1)\langle W \rangle \bar{N}\tau(1 - R^{-1})\kappa^{-1}]^{-\kappa} R^{-n} \quad (21b)$$

$$R > 1.$$

3.4. Analytic Solutions for Large Storm Arrival Rate ($\bar{N}\tau \to \infty$)

For the case in which intraseasonal variability of precipitation and spatial variability of w_0 are ignored, the evapotranspiration ratio is [*Milly*, 1994]

$$F(R, W, \infty, \delta_P, \delta_N, \delta_E) = \min \left\{ 1, R, W + \frac{1+R}{2} + \frac{1-R}{\pi} \arcsin\left(\frac{R-1}{S}\right) - \frac{S}{\pi}\left[1 - \left(\frac{R-1}{S}\right)^2\right]^{1/2} \right\} \quad (22a)$$

$$\left|\frac{R-1}{S}\right| \le 1$$

$$F(R, W, \infty, \delta_P, \delta_N, \delta_E) = \min [1, R] \quad (22b)$$

$$\left|\frac{R-1}{S}\right| > 1$$

where the seasonality index S is given by

$$S = |\delta_P - \delta_E R|. \quad (23)$$

The areal mean evapotranspiration ratio, found by substitution of (22) and (8) into (13), is

$$G(R, \langle W \rangle, \infty, \delta_P, \delta_N, \delta_E, \kappa) = \min (R, 1) - \frac{W_c}{\Gamma(\kappa)} \gamma\left(\kappa, \frac{\kappa W_c}{\langle W \rangle}\right) + \frac{\langle W \rangle}{\Gamma(\kappa+1)} \gamma\left(\kappa + 1, \frac{\kappa W_c}{\langle W \rangle}\right), \quad (24)$$

where $\gamma(\ ,\ )$ is the incomplete gamma function and W_c is the smallest nonnegative value of W for which local runoff is minimized for given values of R and S. When $|(R-1)/S|$ is not greater than 1, the expression for W_c is

$$W_c = \max 0, \frac{S}{\pi}\left[1 - \left(\frac{R-1}{S}\right)^2\right]^{1/2} - \tfrac{1}{2}|1 - R| - \frac{1}{\pi}(1 - R) \arcsin\left(\frac{R-1}{S}\right)\Bigg\}. \quad (25)$$

Otherwise, W_c is zero.

3.5. Illustration of Solutions

The general solutions (10) and (16) are plotted in the top panel of Figure 2 for representative values of the input parameters. The choice of $\delta_e = 1$ implies a midlatitude location (where seasonality of E_p is strong), and the choice $\delta_P = \delta_N = 0$ (no precipitation seasonality) is an average for midlatitudes, where precipitation may have either winter or summer maxima. Representative values for annual mean storm arrival rate, annual mean precipitation rate, and water-holding capacity are 100 yr^{-1}, 1000 mm yr^{-1}, and 200 mm, respectively. These give a typical value for $\langle W \rangle$ of 0.2; values of 0.1 and 0.4 are also included for comparison. Corresponding special-case solutions for no seasonality and for large storm arrival rates are also plotted in the middle and bottom panels, respectively, of Figure 2.

In the general case (top panel) without spatial variability, a dimensionless storage capacity of 0.4 is almost large enough to absorb all fluctuations of water and energy supply

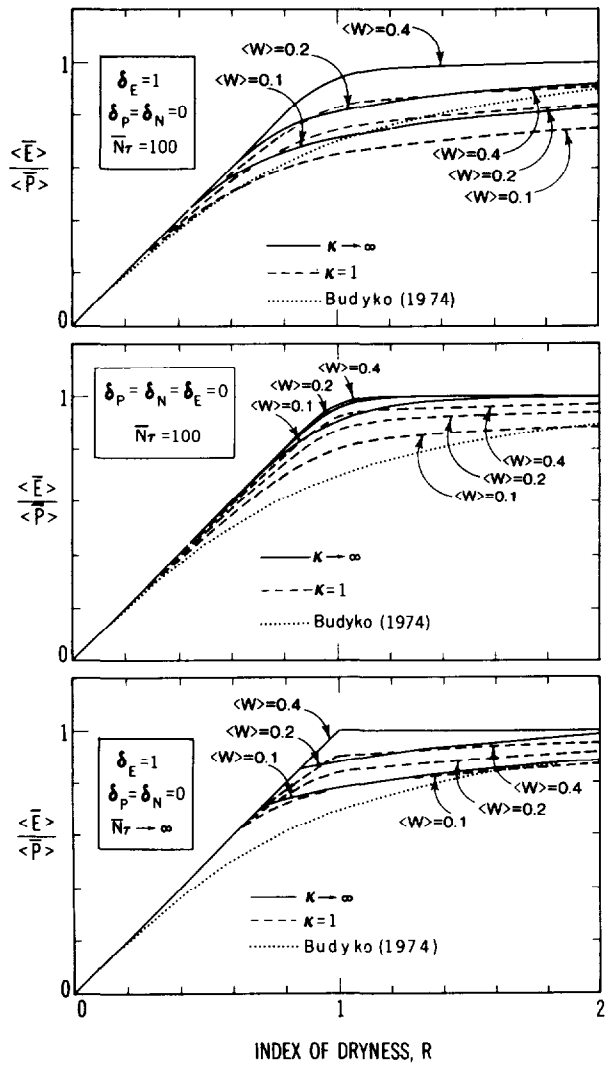

Figure 2. Evapotranspiration ratio as a function of index of dryness for various values of mean dimensionless storage capacity. Solid lines, $\kappa \to \infty$ (no spatial variability of w_0); dashed lines, $\kappa = 1$ (exponential distribution of w_0). (top) Episodic precipitation and annual cycle of potential evapotranspiration considered. (middle) Same parameters as in the top panel, but no annual cycle ($\delta_E = 0$). (bottom) Same as top panel, but precipitation not episodic ($\bar{N}\tau \to \infty$).

over time; the evapotranspiration curve departs from the asymptotes A and B of Figure 1 only slightly near $R = 1$. Halving the value of W from 0.4 to 0.2 (or from 0.2 to 0.1) yields a decrease in evapotranspiration and increase in runoff of the order of 10% of precipitation for R greater than about 0.8. The case of strong spatial variability of w_0 is considered with $\kappa = 1$. Because an increase in w_0 enhances evapotranspiration less than an equal decrease in w_0 reduces it, it follows that increasing dispersion of w_0 around some fixed mean decreases the mean evapotranspiration and increases runoff. *Budyko*'s [1974] curve lies among the illustrative cases of the general solution in Figure 2.

Comparison of the general and special cases illustrated in Figure 2 leads to the conclusion, for the chosen parameter values, that both interseasonal and intraseasonal variabilities of forcing contribute to reductions of annual evapotranspiration below its energy and water supply limits.

3.6. Cross-Checks of Analytic and Monte Carlo Solutions

Each of the curves presented in the middle and bottom panels of Figure 2 was actually computed twice, once by use of the appropriate analytic solution and once by the Monte Carlo method. In each case it was possible to make any difference between the solutions arbitrarily small by increasing the number of years simulated and the number of values of storage capacity (n) used in the Monte Carlo solution. (Typical simulations used 500 years, following a short spin-up, and 20 values of w_0.) The agreement between the analytic and Monte Carlo solutions in these and other tests provides confidence that the simulation technique was properly formulated and that the derivations of the analytic solutions are correct.

4. Application, Testing, and Sensitivity Analysis

The theory presented in section 3 may be tested for its ability to reproduce the observed dependence of the annual water balance on the identified independent variables. The test region is the land area of the United States east of the Rocky Mountains; this is the same area used in an earlier study of the water balance problem [*Milly*, 1994]. The first part of this section describes how the input variables, as well as annual totals of runoff and evapotranspiration, were estimated from available observational data. The remainder of the section compares the modeled annual water balance with observations and addresses related problems.

4.1. Estimation of Inputs and Annual Totals of Runoff and Evapotranspiration

4.1.1. Precipitation ($\bar{P}$, δ_P, $\bar{N}$, δ_N). *Legates and Willmott* [1990a] estimated global fields of monthly mean precipitation, adjusted for gage biases, at a spatial resolution of 0.5 degree. The annual total of these fields was used as an estimate of $\bar{P}\tau$. Fourier analysis of the monthly values was used at each grid point to estimate δ_P. Both $\bar{P}\tau$ and δ_P are mapped in Figure 3.

Unadjusted daily precipitation records from first-order weather stations were used to estimate storm arrival rates. The requirement that the computed means and variances of daily total precipitation amounts equal their theoretical values for the assumed stochastic precipitation model [*Rodriguez-Iturbe et al.*, 1984] gives

$$N = \frac{2(E[h])^2}{T \text{ Var } [h]}, \tag{26}$$

where h is the daily precipitation depth and T is equal to 1 day. This relation was used to estimate N for each month of the year at the first-order observational sites. From these monthly values, corresponding estimates of $\bar{N}\tau$ and δ_N were computed for each station, the latter by Fourier analysis. These site-specific estimates were then objectively interpolated to the 0.5-degree grid. The results are shown in Figure 3. The distribution of $\bar{N}\tau$ is very similar to the distribution of the number of days per year having measured precipitation [*U.S. Department of Commerce*, 1968], although the latter values are slightly larger.

4.1.2. Potential evapotranspiration ($\bar{E}_p$, δ_E). *Milly* [1994] concluded that estimates of potential evapotranspira-

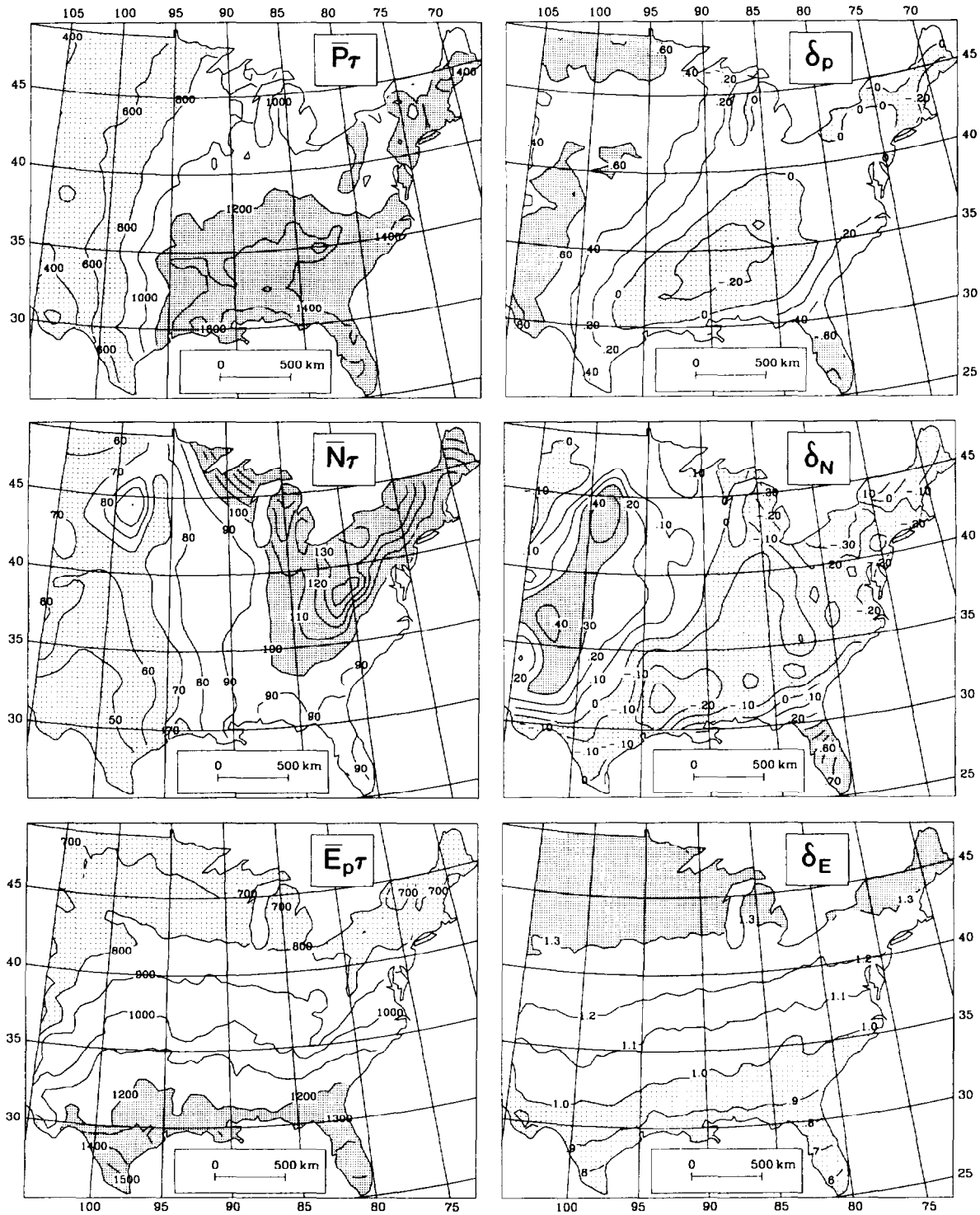

Figure 3. (left) Maps of annual totals of (from top to bottom) precipitation ($\bar{P}\tau$, millimeters), storm arrivals ($\bar{N}\tau$, dimensionless), and potential evapotranspiration ($\bar{E}_p\tau$, millimeters). (right) Corresponding ratios of amplitudes of annual harmonics to respective annual averages (dimensionless); positive δ indicates summer maximum.

tion made by the method of *Thornthwaite* [1948] had a negative bias in the study area. This conclusion was based on the fact that the method gave annual potential evapotranspiration that was less than the apparent evapotranspiration (precipitation minus runoff) over more than half of the study area; in view of the relative reliability of the observations of precipitation and streamflow, the discrepancy was attributed to errors in the potential evapotranspiration. For the present study, monthly values of potential evapotranspiration estimated by the method of *Thornthwaite* [1948], using the temperature data of *Legates and Willmott* [1990b], were all increased by a factor of 1.2, following *Milly* [1994].

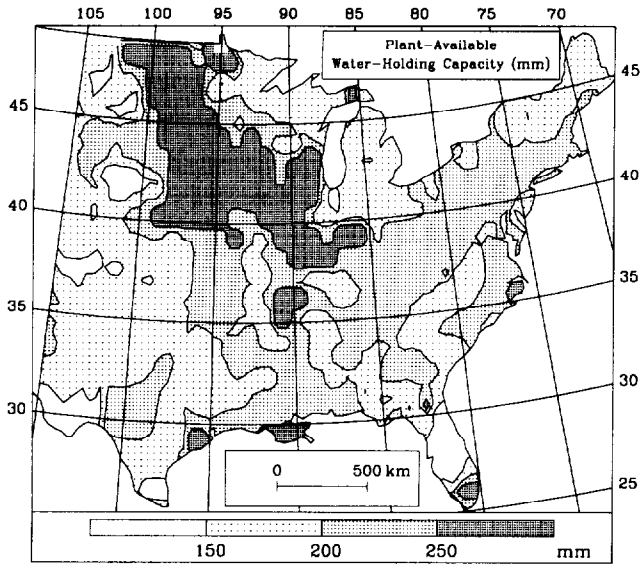

Figure 4. Soil plant-available water-holding capacity (millimeters).

At each grid point, Fourier analysis of the monthly values yielded δ_E. The resultant distributions of $\overline{E_P}$ and δ_E are shown in Figure 3. The implied negative winter values of E_P for $\delta_E > 1$ were a result of including only the annual harmonic to represent the seasonal cycle in (5). These values caused no difficulties and were permitted in the analysis in order to ensure that seasonal integrals of E_P were consistent with observations.

4.1.3. Soil water-holding capacity (w_0, κ). The plant-available water-holding capacity of soils, w_0, was estimated with global coverage at 0.5-degree resolution by *Patterson* [1990], using available global data sets for soils and vegetation (Figure 4). *Patterson*'s [1990] treatment of rooting depths was necessarily highly simplified because data are scarce and the error of estimation of w_0 is probably large.

Local variability of w_0, characterized by κ, is associated with variability of the soil hydraulic properties and of the plant root systems. The contribution due to soil variability was estimated crudely for five counties in the study area on the basis of soil surveys [*U.S. Department of Agriculture*, 1992a, b, c, d, e]. The surveys provide tabulations of the mean and standard deviation of the "available water capacity" (essentially $\theta_f - \theta_w$) for various horizons of each mapped soil unit, as well as the area covered by each unit. These were used to derive the coefficient of variation of depth-integrated (0–36 inches (0–91 cm)) available water capacity. The estimates are biased upward by the assumption that local deviations of available water capacity are perfectly correlated vertically through the profile. Estimated values of the coefficient of variation of w_0 were 0.1 (Wright County, Iowa), 0.2 (Rockingham County, North Carolina), 0.4 (Pendleton County, West Virginia), 0.5 (Walker County, Alabama), and 0.8 (Chippewa County, Michigan). The geometric mean is 0.32, and the corresponding value of κ is 10. Variability of rooting densities could contribute further to variability of w_0. In situations of water scarcity, however, there tends to be an inverse correlation between available water capacity and rooting density, which would tend to reduce the variability of w_0 relative to that of available water capacity. For the present analysis the value of κ is taken to be 10. Clearly, this estimate is subject to considerable uncertainty and geographic variability. Sensitivity runs are described later to evaluate the effect of this uncertainty on the water balance computations.

4.1.4. Annual runoff and evapotranspiration ($\langle\bar{Y}\rangle$, $\langle\bar{E}\rangle$). *Gebert et al.* [1987] analyzed the geographic variability over the United States of average annual runoff for the period 1951–1980. Their analysis employed records of measured streamflow from 5951 gaging stations that were judged to be representative of local conditions and unaffected by upstream storage reservoirs or unknown diversions or return flows. The runoff map of *Gebert et al.* [1987] was subjectively interpolated to the nearest inch of runoff, at 0.5-degree resolution (middle panel of Figure 5). Qualitatively, the distribution of runoff is quite similar to that of precipitation, which is one of its main determinants. However, the runoff ratio (fraction of precipitation that runs off) is much smaller in the west than in the east (bottom panel of Figure 8). An empirical analysis of runoff from this area was given by *Langbein et al.* [1949].

In the long term, over an area with no net groundwater inflow or outflow, the difference between precipitation and streamflow equals evapotranspiration. For this study we define an "observed" mean annual evapotranspiration (bottom of Figure 6) as the difference between the observed annual totals of precipitation and runoff.

4.2. Comparisons With Observational Data

Annual runoff was calculated by means of (16). The modeled and observed annual runoff distributions are mapped in Figure 5, along with their difference. The geographic distribution of calculated runoff shares, at least qualitatively, the large-scale features apparent in the observations. In both computations and observations, runoff is lowest in the western region, where its gradient is predominantly east-west. The computations also reproduce the large-scale band of high observed runoff extending northeastward from the northern coast of the Gulf of Mexico, through the Appalachian Mountains, and into the northeastern United States. Mean values of P, Y, and E estimated from observations over the study area are 991 mm, 263 mm, and 728 mm, respectively; mean modeled values of Y and E are 250 mm and 741 mm, respectively. The area-weighted, root-mean-square difference between the observed and modeled values of 0.5-degree runoff and evapotranspiration is 78 mm.

In the west the major departure from the prevailing east-west gradient in the observations is found in the northern plains, where the westward bulge in the 25-mm runoff contour marks the area of anomalously high runoff in the Sand Hills of Nebraska. The anomaly appears also in the computed runoff; the sandy texture of the soil there inhibits water retention, as seen in Figure 4.

The runoff error (negative evapotranspiration error) is positive over much of the south central part of the study area and negative over much of the Appalachians. Regional errors in potential evapotranspiration and storage capacity are possible causes of these errors. It appears that *Patterson*'s [1990] estimates of w_0 fail to include the inhibiting effect of mountain topography on soil development, at least in the study area (Figure 4), and this may explain the runoff

deficit in the Appalachians. A simple error analysis is described in the next section.

Modeled and observed evapotranspiration are mapped in Figure 6. Overall, the distributions are very similar. The maps are nearly identical in the west, where the evapotranspiration ratio is almost unity in both model and observations.

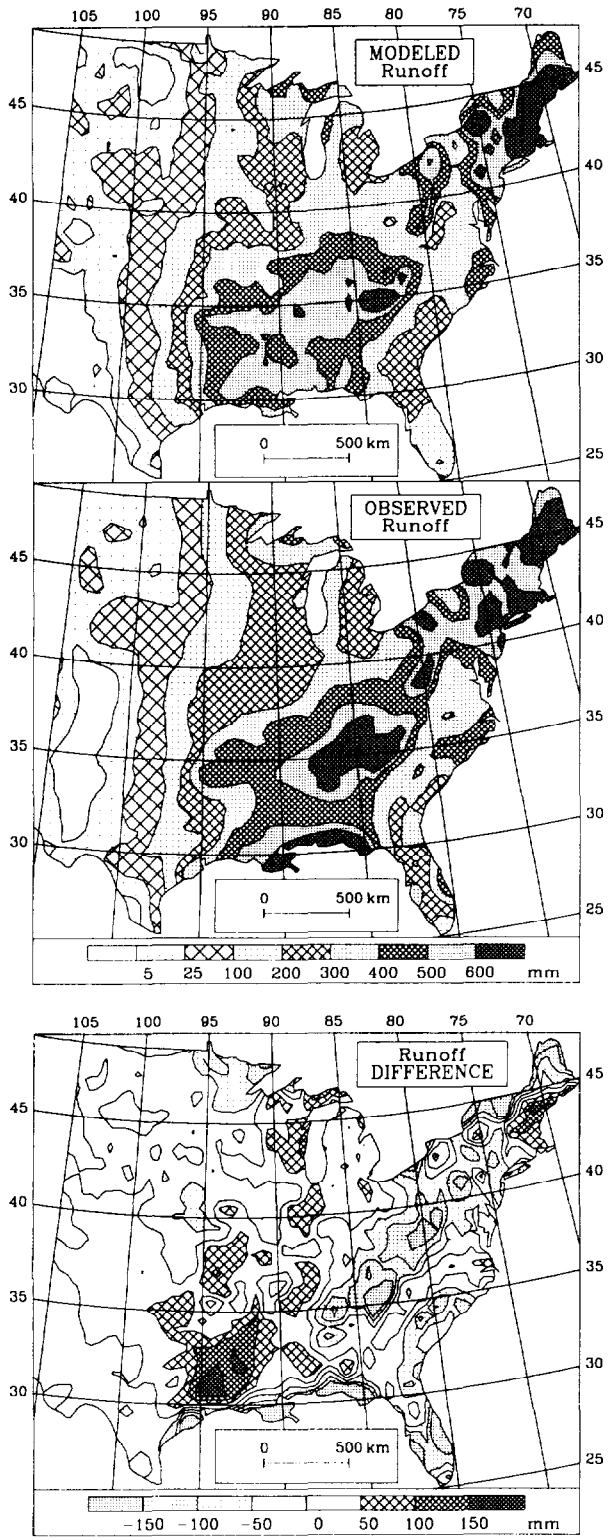

Figure 5. Distributions of (top) modeled and (middle) observed annual runoff and (bottom) their difference.

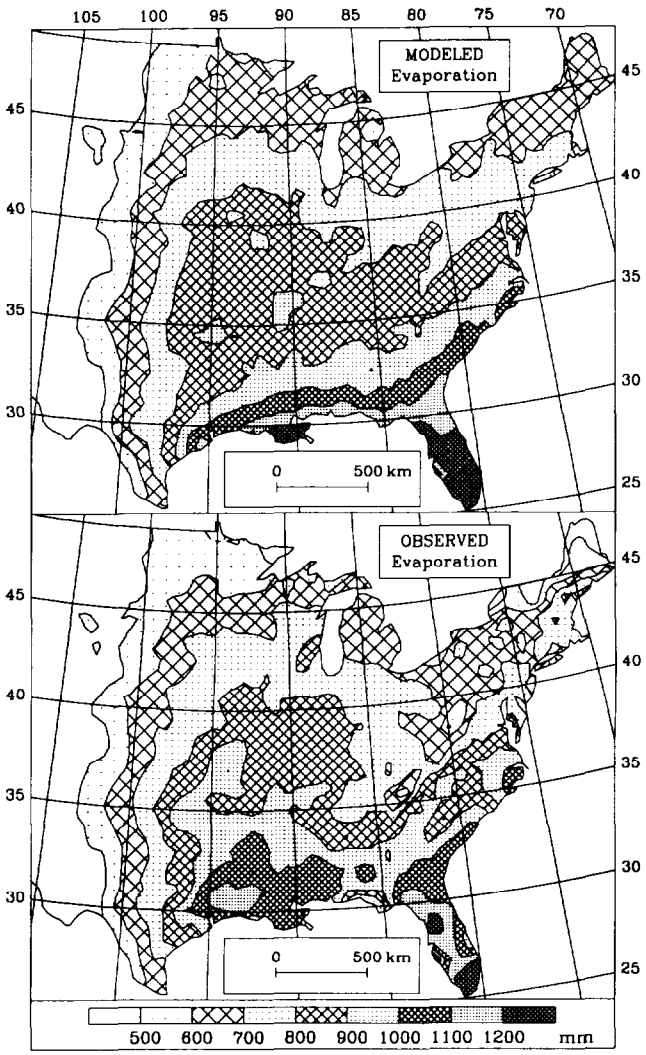

Figure 6. Distributions of (top) modeled and (bottom) observed annual evapotranspiration.

However, consistent with the runoff error already noted, the modeled evapotranspiration fails to reproduce the topographic dependence that is observed in the east.

Figure 7 contains scatterplots of modeled 0.5-degree grid point values of runoff and evapotranspiration against the observational data. The correlation of grid point values of runoff is 0.938; the model thus explains 88% of the variance in observed grid point runoff. For evapotranspiration the correlation coefficient is 0.924; the model explains 85% of the variance.

To a great extent, the spatial patterns common to both the modeled and the observed distributions in Figure 5 are attributable to precipitation patterns. Therefore a more telling comparison is given in Figure 8, which shows modeled and observed runoff ratios; these vary from less than 0.01 to more than 0.5. The patterns are similar, with typical differences of the order of 0.05; the most significant errors are still in the Appalachians.

4.3. Runoff Uncertainty Associated With Input Uncertainty

The runoff differences seen in Figure 5 can be attributed to errors in the model, errors in the various inputs, and errors

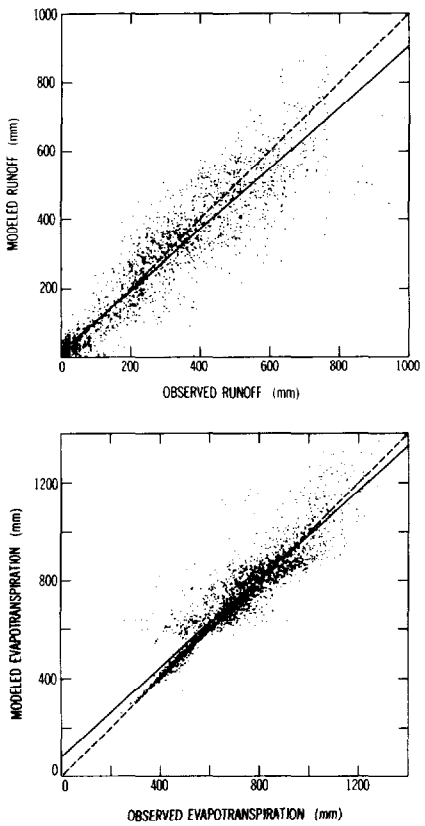

Figure 7. Plots of modeled versus observed 0.5-degree grid point values of (top) runoff and (bottom) evapotranspiration. Dashed lines are 1:1; solid lines are least squares fit.

in the observed runoff. Here we make an estimate of the size of differences that could be expected to result from errors in the most important inputs and from errors in observed runoff. The standard errors of estimation of grid point annual precipitation, annual potential evapotranspiration, mean water-holding capacity, and observed annual runoff were subjectively assigned values of 5%, 10%, 25%, and 5% of their estimated values, respectively; errors in $\bar{N}\tau$, δ_P, δ_N, and δ_E were neglected; and κ was considered separately (see below). The three mutually independent input errors were propagated through (16) by simulation, and the corresponding error variance of modeled annual runoff at each grid point was derived. To this was added the observation variance. No attempt was made to describe the spatial correlation of errors. However, it was expected that many errors would correlate over large distances (10^2–10^3 km) because of neglected climatic dependences of precipitation gage biases and E_p estimates, as well as topography- and vegetation-dependent errors in the estimates of w_0. The square root of the modeled total variance of difference between modeled and observed runoff is mapped in Figure 9; its root-mean-square value over the study area is 66 mm. (Independent contributions from precipitation, potential evapotranspiration, water-holding capacity, and observed runoff are 40 mm, 39 mm, 32 mm, and 16 mm, respectively.) The 66-mm figure compares favorably with the actual root-mean-square difference of 78 mm, and the map of computed differences between modeled and observed runoff in Figure 5 appears to be consistent with the map of modeled standard error of estimation of runoff (Figure 9).

The effect of uncertainty in κ was evaluated separately. The value of 10 was replaced by values of 4 and ∞, corresponding to coefficients of variation of w_0 of 0.5 and 0, respectively. Areal mean runoff changed little, increasing by 10 mm (4%) with $\kappa = 4$ and decreasing by 7 mm (3%) with $\kappa \to \infty$.

It may be concluded that differences between modeled and observed water balances could be explained entirely by input errors and errors in the field of observed runoff. In the absence of more precise estimates of these variables, there is no compelling reason, on the basis of the present comparison alone, to reject the hypothesis underlying the model formulated herein. This issue will be discussed further in the concluding section.

4.4. Sensitivity of Water Balance to Storage Capacity

When the uncertainty analysis was conducted, it was noted that an increase in w_0 yielded a smaller change in the water balance than did an equal decrease in w_0. To explore this at the largest spatial scale, water balances were recomputed using (16) with all values of w_0 scaled by a constant

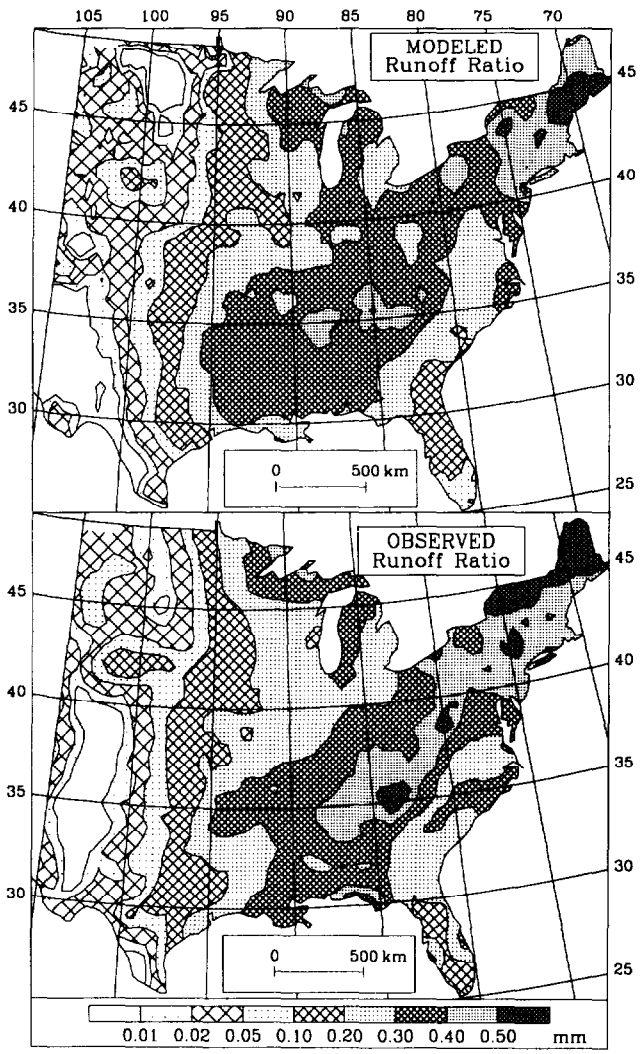

Figure 8. Distributions of (top) modeled and (bottom) observed runoff ratios.

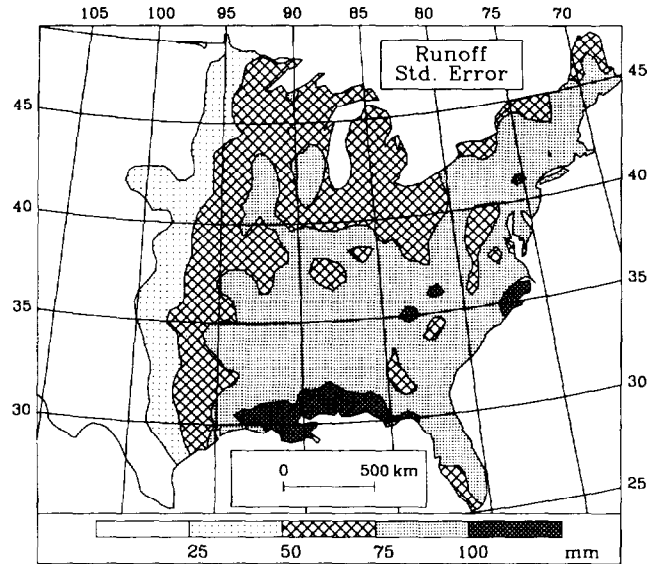

Figure 9. Distribution of modeled standard error of difference between modeled and observed runoff.

factor. The factor was varied over many orders of magnitude, and the resulting areal mean runoff is plotted in Figure 10. As the storage capacity approaches zero, runoff approaches the total water supply, which is essentially equal to precipitation. With increasing storage capacity the evapotranspiration increases and the runoff decreases. At sufficiently large capacity, evapotranspiration is limited by the lesser of water and energy supply and is insensitive to the capacity.

The sensitivity of water balance to plant-available water-holding capacity diminishes markedly at a scale factor on the order of 1. In the absence of input and model errors this implies that the (estimated) actual values of capacity are almost large enough to maximize evapotranspiration. *Milly and Dunne* [1994] noted a similar phenomenon in a model of the global water cycle and speculated that it could indicate that the rooting depths of plants (a crucial determinant of plant-available water-holding capacity) reflect ecologically optimized responses to the relative timing and magnitude of water and energy supplies. The present analysis shows, in a crude way, that observational data support this proposition. A more rigorous test of this concept would consider geographic variability and ideally would use more accurate estimates of w_0 than those used here.

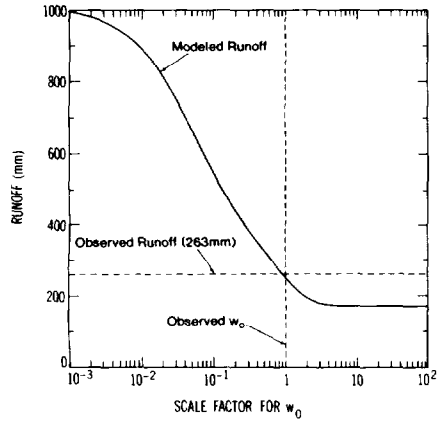

Figure 10. Sensitivity of area-average runoff to the scale factor for the plant-available water-holding capacity.

Table 1. Partitioning of Runoff Into Its Components

Runoff Component	Runoff Amount, mm		
	$\kappa = 10$	$\kappa \to \infty$	Difference
$\langle \bar{Y} \rangle_a$	168	168	0
$\langle \bar{Y} \rangle_\delta$	25	19	6
$\langle \bar{Y} \rangle_N$	14	11	3
$\langle \bar{Y} \rangle_{\delta N}$	43	45	−2
$\langle \bar{Y} \rangle$	250	243	7

4.5. Analysis of Contributions to Runoff

In order to understand the cause of runoff according to (16), it is helpful to decompose total runoff into several terms associated with various types of variability over space and time:

$$\langle \bar{Y} \rangle = \langle \bar{Y} \rangle_a + \langle \bar{Y} \rangle_\delta + \langle \bar{Y} \rangle_N + \langle \bar{Y} \rangle_{\delta N} \tag{27}$$

$$\langle \bar{Y} \rangle_a = \langle \bar{P} \rangle [1 - G(R, \langle W \rangle, \infty, 0, 0, 0, \kappa)] \tag{28}$$

$$\langle \bar{Y} \rangle_\delta = \langle \bar{P} \rangle [G(R, \langle W \rangle, \infty, 0, 0, 0, \kappa) - G(R, \langle W \rangle, \infty, \delta_P, \delta_N, \delta_E, \kappa)] \tag{29}$$

$$\langle \bar{Y} \rangle_N = \langle \bar{P} \rangle [G(R, \langle W \rangle, \infty, 0, 0, 0, \kappa) - G(R, \langle W \rangle, \bar{N}\tau, 0, 0, 0, \kappa)] \tag{30}$$

$$\langle \bar{Y} \rangle_{\delta N} = \langle \bar{P} \rangle [G(R, \langle W \rangle, \infty, \delta_P, \delta_N, \delta_E, \kappa) - G(R, \langle W \rangle, \bar{N}\tau, \delta_P, \delta_N, \delta_E, \kappa)] - \langle \bar{P} \rangle [G(R, \langle W \rangle, \infty, 0, 0, 0, \kappa) - G(R, \langle W \rangle, \bar{N}\tau, 0, 0, 0, \kappa)]. \tag{31}$$

Note that substitution of (28) through (31) into (27) returns (16) and that the first three runoff components on the right side of (27) may be evaluated analytically using solutions already given. The runoff $\langle \bar{Y} \rangle_a$ is the runoff that would occur in the absence of variability over space and time; this is the runoff corresponding to the asymptotes in Figure 1. The term $\langle \bar{Y} \rangle_\delta$ is the runoff in excess of $\langle \bar{Y} \rangle_a$ that would be caused by storage capacities insufficient to overcome seasonal variations in the difference between water and energy supplies if storm arrival rates were large. The term $\langle \bar{Y} \rangle_N$ is the runoff in excess of $\langle \bar{Y} \rangle_a$ that would be induced by the inability of the storage capacities to compensate for the episodic delivery of water and energy to the surface, in the absence of seasonality. The final term $\langle \bar{Y} \rangle_{\delta N}$ accounts for nonlinear interactions between seasonality and storminess; it can be viewed as the effect of seasonality on the production of runoff by the episodic nature of storms. It is also possible to divide further the total runoff, and each of the four terms identified above, into components that would arise in the absence of spatial variability ($\kappa \to \infty$) and those that are induced by spatial variability.

The various components of expected annual runoff, aver-

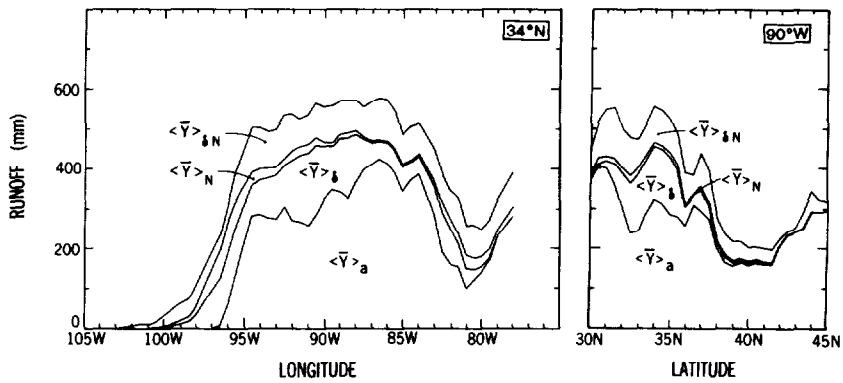

Figure 11. Division of annual runoff $\langle \bar{Y} \rangle$ (top curve) into its components $\langle \bar{Y} \rangle_a$, $\langle \bar{Y} \rangle_\delta$, $\langle \bar{Y} \rangle_N$, and $\langle \bar{Y} \rangle_{\delta N}$ (left) at latitude 34°N and (right) at longitude 90°W. $\langle \bar{Y} \rangle_a$ is height of lowest curve, and other components are vertical distances between successive curves.

aged over the study area, are given in Table 1. About two thirds of the runoff can be explained simply by the local excesses of expected annual precipitation over annual potential evapotranspiration, reflected in $\langle \bar{Y} \rangle_a$. Significant fractions are also represented by $\langle \bar{Y} \rangle_\delta$, $\langle \bar{Y} \rangle_N$, and $\langle \bar{Y} \rangle_{\delta N}$, which contribute 10%, 6%, and 17% of the total, respectively, over the study area (for $\kappa = 10$). The difference between runoff produced with $\kappa = 10$ and that with $\kappa \to \infty$ is quite small. Overall, spatial variability of w_0 contributes only about 3% of the total runoff.

Figure 11 shows how the total runoff is divided into the four components defined above, for an east-west transect at 34°N and a north-south transect at 90°W. The east-west transect shows that $\langle \bar{Y} \rangle_a$ is the dominant runoff term in the east. In the west, however, that term is zero, because the dryness index exceeds unity there. Runoff $\langle \bar{Y} \rangle_\delta$ caused by seasonality adds substantially to $\langle \bar{Y} \rangle_a$ and penetrates slightly farther west than does $\langle \bar{Y} \rangle_a$. It is zero west of 100°W because potential evapotranspiration exceeds precipitation there throughout the year; most precipitation in the central United States falls during summer, as was seen in Figure 3. The two runoff terms associated with the episodic nature of precipitation add about 100 mm over most of the east-west transect; in the western region these terms are dominant. It can be seen from the 90°W transect that the importance of variability of forcing over time for runoff generation is greatly reduced in the northern half of the study area; there the main addition to $\langle \bar{Y} \rangle_a$ comes from $\langle \bar{Y} \rangle_{\delta N}$.

4.6. Evaluation of Seasonality Approximation

All solutions presented in this paper rely on the simple model of seasonality embodied in (5), (6), and (7). The accuracy of this approximation was evaluated by performing a similar Monte Carlo analysis in which the variables $P(t)$, $E_p(t)$, and $N(t)$ were allowed to vary stepwise from month to month, with each grid point value obtained by spatial interpolation of monthly station data. The resulting mean areal runoff was 254 mm, and the map of computed runoff was almost identical to the one in Figure 5. Thus the adoption of (5), (6), and (7) introduced an areal mean bias of only −4 mm, which is less than 2% of runoff or 0.5% of precipitation.

5. Concluding Remarks

5.1. Summary

This paper describes the mathematical development and the testing of a hypothesis concerning the control of the average annual water balance of the land surface. It is hypothesized that the local annual water balance is controlled by the distributions in time of water supplies (precipitation) and demands (potential evapotranspiration), which are balanced, to the extent possible, by storage of water in the root zone of the soil. The storage capacity of the root zone is characterized by using conventional concepts of soil water availability to plants. It is assumed that permeability is sufficiently large that infiltration of water into the soil and uptake of available soil water by roots are unrestricted. The hypothesis was stated mathematically as a simple water balance model driven by time series of precipitation and potential evapotranspiration. Simplified parametric representations of the variability (random and seasonal) in time of precipitation and potential evapotranspiration were used. Spatial variability of the storage capacity of the soil was assumed to follow the gamma distribution.

Given these assumptions the time-mean partitioning of precipitation into runoff and evapotranspiration is determined by seven dimensionless numbers. The first of these, the ratio of annual potential evapotranspiration to annual precipitation (index of dryness), has long been recognized as an important determinant of the annual water balance. The second factor is the ratio of water-holding capacity to annual mean precipitation amount; large values of the ratio tend to promote evapotranspiration and suppress runoff. The third factor is the mean number of precipitation events per year; for the same amount of annual precipitation, a few heavy rainfalls will produce more runoff than many small ones. Three more factors are the ratios of seasonal fluctuations to annual means of precipitation, storm arrival rate, and potential evapotranspiration. In general, seasonality tends to generate imbalances in water supply and demand, leading to increased runoff. The final factor is a measure of the spatial variability of water-holding capacity. Spatial variability enhances runoff, but its effect was small in this study.

In the most general case the water balance problem was

solved by Monte Carlo simulation. For important limiting cases, however, it was possible to derive exact analytic solutions for the water balance.

The hypothesis of storage control of the average annual water balance was tested by applying the derived water balance model to the area of the United States east of the Rocky Mountains. For each point on a 0.5-degree grid the seven independent variables were estimated a priori on the basis of published data and methods; there were no free parameters for calibration. The model yielded mean runoff and evapotranspiration values of 250 mm and 741 mm, respectively, over the study area; observations yielded 263 mm and 728 mm, respectively. The modeled and observed runoff distributions were very similar. The root-mean-square difference between observed and modeled grid point values for runoff and evapotranspiration was 78 mm, and the model explained 88% and 85% of the geographic variance in grid point runoff and evapotranspiration, respectively. An uncertainty analysis supported the suggestion that input errors and errors in observed runoff are sufficient to explain the departures of the model from the runoff observations.

An analysis of the sensitivity of the water balance to the plant-available water-holding capacity of the soil provided some support for the suggestion [*Milly and Dunne*, 1994] that average plant rooting depths are close to those that are just large enough to minimize runoff.

In the framework of the model the average annual runoff can be partitioned into runoff associated with the imbalance between long-term means of precipitation and potential evapotranspiration, runoff caused by differing seasonalities of precipitation and potential evapotranspiration, runoff caused by storminess of precipitation, and runoff caused by the interaction of storminess and seasonality. In the areal mean over the study area the term associated with annual mean water and energy supplies was the largest, but no term was negligible. The terms associated with fluctuations of the forcing become increasingly important under increasing aridity. The effect of local spatial variability of the soil water-holding capacity on mean runoff (and evapotranspiration) was negligibly small.

5.2. Discussion

The agreement between model and observations in the present study supports the assumptions underlying the theory, in particular the hypothesis of storage control of the water balance. The differences between model and observations were shown to lie within the range that could be expected, given the uncertainties in potential evapotranspiration, precipitation, and especially water-holding capacity of the root zone of the soil. Of course, these uncertainties leave open the possibility that the annual water balance could be affected by factors not included here, such as restriction of rainwater infiltration by limited soil permeability and restriction of transpiration by moisture diffusion toward plant roots. It can only be concluded that it is not necessary to invoke those factors to explain the observational data.

If the conceptual model formulated here is at all representative of the real sensitivities of water balance to its controlling factors, then model testing with existing large-scale data sets has serious limitations. From a data standpoint the weak point in the analysis is the atmospheric forcing by precipitation and potential evapotranspiration. It appears that the uncertainties associated with these factors may be at least as large as the differences among conceptual models of water balance, making the rejection of any particular hypothesis an improbable event. Without better estimates of these quantities, it is doubtful whether significant advances can be made in the scientific analysis of the annual water balance on large spatial scales.

5.3. Limitations

From a physical standpoint a weakness in the present theory is its adoption of the concept of a water-holding capacity. This concept ignores many of the details of soil water flow [*Hillel*, 1980]. However, in view of our inability to apply more sophisticated theories of soil water physics successfully in natural field situations [*Beven*, 1989], it seems appropriate to have adopted a simplified parametric representation for the present study.

Although the water-holding capacity has been allowed to depend on rooting density and soil hydraulic properties, these dependences have been incorporated only in the crudest fashion. In principle, it should be possible to make better estimates by straightforward application of the standard soil physical theory. In view of the practical difficulties in application of the theory, however, it is arguable that such an analysis would not significantly narrow the range of uncertainty in the effective water-holding capacity.

It should also be noted that the determination of potential evapotranspiration was placed outside the scope of this study. Its proper estimation requires consideration of the energy balance at the land surface and of the non-water-stressed resistance of plants to water loss from the surface.

In order to simplify the description of water supply to the soil, it was convenient here to ignore phenomena associated with frozen water. The obvious effect of such phenomena in the present context is to increase the total storage capacity above that of the root zone of the soil, thereby increasing the amount of winter precipitation that is conserved until the snowmelt season. Because evaporation tends to be energy-limited throughout the affected time period, snow storage should not have a major influence on the annual water balance, although it does affect the timing of runoff. An effect that should possibly be investigated is the snow-induced depression of surface temperatures below those that would occur at snow-free surfaces and the resulting reduction in the potential evapotranspiration rate. Such depression must result both from the high albedo of snow and from the thermostatic constraint that snow temperature cannot exceed the freezing point. It could be expected that this would cause an increase in runoff and a decrease in evapotranspiration.

Acknowledgments. K. A. (Patterson) Dunne provided global estimates of total available water-holding capacity and gave technical assistance. R. Koster, R. T. Wetherald, G. J. Wiche, D. Wolock, and two anonymous reviewers provided technical reviews.

References

Benjamin, J. R., and C. A. Cornell, *Probability, Statistics and Decision for Civil Engineers*, McGraw-Hill, New York, 1970.

Beven, K., Changing ideas in hydrology—The case of physically-based models, *J. Hydrol.*, *105*, 157–172, 1989.

Brutsaert, W. H., *Evaporation into the Atmosphere*, 299 pp., D. Reidel, Norwell, Mass., 1982.

Budyko, M. I., *Climate and Life*, 508 pp., Academic, San Diego, Calif., 1974.

Budyko, M. I., and L. I. Zubenok, Determination of evaporation from the land surface (in Russian), *Izv. Akad. Nauk SSSR, Ser. Geogr.*, no. 6, 3–17, 1961.

Dooge, J. C. I., Sensitivity of runoff to climate change: A Hortonian approach, *Bull. Am. Meteorol. Soc.*, *73*, 2013–2024, 1992.

Eagleson, P. S., Climate, soil, and vegetation, 1, Introduction to water balance dynamics, *Water Resour. Res.*, *14*, 705–712, 1978a.

Eagleson, P. S., Climate, soil, and vegetation, 4, The expected value of annual evapotranspiration, *Water Resour. Res.*, *14*, 731–739, 1978b.

Eagleson, P. S., Climate, soil, and vegetation, 6, Dynamics of the annual water balance, *Water Resour. Res.*, *14*, 749–764, 1978c.

Erdelyi, A. (Ed.), *Higher Transcendental Functions*, vol. 1, 302 pp., McGraw-Hill, New York, 1953.

Gebert, W. A., D. J. Graczyk, and W. R. Krug, Average annual runoff in the United States, 1951–80, *Hydrol. Invest. Atlas HA-710*, U.S. Geological Survey, Reston, Va., 1987.

Gradshteyn, I. S., and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 1160 pp., Academic, San Diego, Calif., 1980.

Hillel, D., *Applications of Soil Physics*, 385 pp., Academic, San Diego, Calif., 1980.

Langbein, W. B., et al., Annual runoff in the United States, *U.S. Geol. Surv. Circ.*, *52*, 14 pp., 1949.

Legates, D. R., and C. J. Willmott, Mean seasonal and spatial variability in gauge-corrected, global precipitation, *Int. J. Climatol.*, *10*, 111–127, 1990a.

Legates, D. R., and C. J. Willmott, Mean seasonal and spatial variability in global surface air temperature, *Theor. Appl. Climatol.*, *42*, 11–21, 1990b.

Milly, P. C. D., An analytic solution of the stochastic storage problem applicable to soil water, *Water Resour. Res.*, *29*, 3755–3758, 1993.

Milly, P. C. D., Climate, interseasonal storage of soil water, and the annual water balance, *Adv. Water Resour.*, in press, 1994.

Milly, P. C. D., and K. A. Dunne, Sensitivity of the global water cycle to the water-holding capacity of soils, *J. Clim.*, *7*, 506–526, 1994.

Milly, P. C. D., and P. S. Eagleson, Effects of spatial variability on annual average water balance, *Water Resour. Res.*, *23*, 2135–2143, 1987.

Patterson, K. A., Global distributions of total and total-available soil water-holding capacities, Master's thesis, Dep. of Geogr., Univ. of Del., Newark, 1990.

Rodriguez-Iturbe, I., V. K. Gupta, and E. Waymire, Scale considerations in the modeling of temporal rainfall, *Water Resour. Res.*, *20*, 1611–1619, 1984.

Schaake, J. C., Jr., and C. Liu, Development and application of simple water balance models to understand the relationship between climate and water resources, in New Directions for Surface Water Modeling (Proceedings of the Baltimore Symposium, May 1989), *IAHS Publ.*, *181*, 343–352, 1989.

Thornthwaite, C. W., An approach toward a rational classification of climate, *Geogr. Rev.*, *38*, 55–94, 1948.

Todorovic, P., and V. Yevjevich, Stochastic process of precipitation, *Hydrol. Pap. 35*, 61 pp., Colo. State Univ., Fort Collins, 1969.

Trewartha, G. T., *An Introduction to Climate*, 4th ed., 408 pp., McGraw-Hill, New York, 1968.

U.S. Department of Agriculture, Soil survey of Chippewa County, Michigan, report, Washington, D. C., 1992a.

U.S. Department of Agriculture, Soil survey of Pendleton County, West Virginia, report, Washington, D. C., 1992b.

U.S. Department of Agriculture, Soil survey of Rockingham County, North Carolina, report, Washington, D. C., 1992c.

U.S. Department of Agriculture, Soil survey of Walker County, Alabama, report, Washington, D. C., 1992d.

U.S. Department of Agriculture, Soil survey of Wright County, Iowa, report, Washington, D. C., 1992e.

U.S. Department of Commerce, Climatic atlas of the United States, report, 80 pp., Rockville, Md., 1968.

Warrick, A. W., and D. R. Nielsen, Spatial variability of soil physical properties in the field, in *Applications of Soil Physics*, edited by D. Hillel, 385 pp., Academic, San Diego, Calif., 1980.

P. C. D. Milly, U.S. Geological Survey, Geophysical Fluid Dynamics Laboratory/NOAA, PO Box 308, Princeton, NJ 08542.

(Received December 13, 1993; revised February 17, 1994; accepted February 22, 1994.)