Macroscale water fluxes
1. Quantifying errors in the estimation of basin mean precipitation

P. C. D. Milly and K. A. Dunne

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

Received 9 July 2001; revised 15 April 2002; accepted 15 April 2002; published 23 October 2002.

[1] Developments in analysis and modeling of continental water and energy balances are hindered by the limited availability and quality of observational data. The lack of information on error characteristics of basin water supply is an especially serious limitation. Here we describe the development and testing of methods for quantifying several errors in basin mean precipitation, both in the long-term mean and in the monthly and annual anomalies. To quantify errors in the long-term mean, two error indices are developed and tested with positive results. The first provides an estimate of the variance of the spatial sampling error of long-term basin mean precipitation obtained from a gauge network, in the absence of orographic effects; this estimate is obtained by use only of the gauge records. The second gives a simple estimate of the basin mean orographic bias as a function of the topographic structure of the basin and the locations of gauges therein. Neither index requires restrictive statistical assumptions (such as spatial homogeneity) about the precipitation process. Adjustments of precipitation for gauge bias and estimates of the adjustment errors are made by applying results of a previous study. Additionally, standard correlation-based methods are applied for the quantification of spatial sampling errors in the estimation of monthly and annual values of basin mean precipitation. These methods also perform well, as indicated by network subsampling tests in densely gauged basins. The methods are developed and applied with data for 175 large (median area of 51,000 km^2) river basins of the world for which contemporaneous, continuous (missing fewer than 2% of data values), long-term (median record length of 54 years) river discharge records are also available. Spatial coverage of the resulting river basin data set is greatest in the middle latitudes, though many basins are located in the tropics and the high latitudes, and the data set spans the major climatic and vegetation zones of the world. This new data set can be applied in diagnostic and theoretical studies of water balance of large basins and in the evaluation of performance of global models of land water balance. *INDEX TERMS:* 1854 Hydrology: Precipitation (3354); 1869 Hydrology: Stochastic processes; 1894 Hydrology: Instruments and techniques; *KEYWORDS:* error, sampling, topography, interpolation

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

1. Introduction

1.1. Critical Role of Observations in Large-Scale Hydrology

[2] Regional- to global-scale fluxes of water and energy at the land surface couple the land to the global climate system [*Eagleson*, 1994; *U.S. National Research Council*, 1991]. To what extent, then, and by what processes might spatial and temporal variations, including long-term trends, in terrestrial water fluxes reveal, modulate, or even drive variability in the climate system? It has been recognized that a balanced research portfolio of observation, theory, and modeling is the key to development of improved description, understanding and prediction of the global water cycle and its connection with the global climate system [*U.S. National Research Council*, 1990, 1998].

[3] It sometimes seems, however, that we have no shortage of theory or model sophistication for describing land processes, and that the weak scientific link is the use of observational data as a constraint on theory and model development. And while it is indeed important, even crucial, to continue efforts to enhance global observational capabilities [*U.S. National Research Council*, 1998], it is equally important to integrate available observational data into ongoing theoretical research and model development. This is especially true in the case of long-term (spanning decades to a century) observational data sets, whose retrospective value cannot be challenged by nascent technologies. In hydrology, the foremost of such data sets are the global records of point precipitation (input) and measurements of river discharge (output) for basins of widely varying spatial scales. These conventional data contain a wealth of long-term information on the response of river basins to atmospheric variability.

1.2. Need to Quantify Observational Data Errors

[4] Meaningful use of observational data in tests of hydrologic hypotheses and in assessments of model per-

formance requires an understanding of uncertainties in the data used. Accuracy increases the probability that the data will enable us to distinguish between competing hypotheses or, equivalently, to detect errors in models. For example, errors in atmospheric forcing can seriously impede the process of hypothesis testing with respect to physical processes in a land model; errors in the precipitation estimates used to drive a water balance model can produce errors in runoff as large as those caused by conceptual errors in the model itself [*Milly*, 1994]. Failure to recognize data inaccuracy can lead to inappropriate conclusions, because errors in the data can easily be misinterpreted as indicators of real physical processes.

[5] One approach to control of errors in forcing is to screen out data sets judged to have unacceptable errors. *Oki et al.* [1999], in evaluating the performance of 11 global models of land water and energy balance, suggested use of a critical precipitation-gauge density as an index of error. Using flow records for more than 200 river basins, they examined the dependence of model runoff error statistics on the choice of the critical density. They found that the standard deviation of error was maximum at zero density and asymptotically approached a minimum at a density of about 30–50 gauges/10^6 km^2. They concluded that this density of rain gauges was sufficient "to prevent the effect of poor forcing precipitation." However, no direct, quantitative estimates of the precipitation error or its contribution to model runoff error were made.

[6] The introduction of simple data screening rules is a step forward in the evaluation of macroscale land models. The use of more rigorous, objective methods would presumably increase the efficiency of model evaluation. The most efficient screening criteria will filter out data with unacceptable errors (which could be mistaken for model errors) and will accept data with sufficiently small errors (in order to maximize the power of the evaluation data set). Of course, descriptors of error may be useless or even misleading if they are not accurate. Overestimation of data errors can cause missed opportunities to reject invalid hypotheses or models. For error estimates to be acceptable, the methods of their estimation must be shown to be reliable.

1.3. Objectives of These Papers

[7] This is the first in a series of three papers analyzing controls on water balances of large land areas. This first paper (part 1) describes the development of the data set upon which the subsequent papers are based, with special attention to assessment of errors in estimates of precipitation. In part 2 [*Milly and Dunne*, 2002], these data are employed to analyze the control of interannual water balance variations by fluctuations in supplies of water (precipitation) and energy (surface net radiation). In part 3 [*Milly and Wetherald*, 2002], the data of part 1 and the results of part 2 are used to develop and quantify a conceptual picture of land-process controls of monthly stream flow variability. In a related series of papers [*Milly and Shmakin*, 2002a, 2002b; *Shmakin et al.*, 2002], this paper's estimates of precipitation and, crucially, associated errors are used to evaluate the performance of a new global model of land water and energy balances.

[8] More specifically, the objectives of this first paper are (1) to develop methods for precipitation error estimation and (2) to create a river basin data set of precipitation and discharge for use in analyses of large-scale water and energy balances. Our analysis of errors focuses on precipitation; this is presumably the more uncertain of the two variables, because it is derived by spatial averaging of (sometimes sparse) point samples of a spatially variable process. The data set is global in scope, and contains information on precipitation and basin discharge at a monthly timescale. Basins considered are those with horizontal length scale on the order of 100 to 1000 km. Record length varies from basin to basin, ranging from about 20 to 200 years.

2. Data

2.1. Precipitation

[9] We used precipitation gauge data from the beta release of the Global Historical Climatology Network (GHCN) version 2 data set, produced jointly by the National Climatic Data Center and the Carbon Dioxide Information Analysis Center. GHCN version 2 includes monthly precipitation data from 20,790 stations. All GHCN data have been subjected to several simple quality control procedures. Where such procedures identified duplicate records for a single site, we have used the longest record.

[10] Because it is based on surface observations, the coverage of global land area by GHCN is not uniform in space. Major regions with lowest densities of gauges include not only the Sahara, Australian, and central Asian deserts, but also wetter (but sparsely populated and less developed) regions of northern Asia and North America, as well as much of South America.

[11] Precipitation data were analyzed separately for each month of the year. For a given month of the year, a precipitation gauge was used in our analyses only if it had data in at least half of the years of a common time period (1951–1980) for that month of the year. This constraint was applied, in conjunction with our analysis method, in order to avoid spurious generation of trends in basin mean precipitation due to changes in the network over time. Of the 20,790 stations in the data set, 3864 stations are within our chosen set of river basins (described later). This number is reduced to 3212 when those stations with insufficient data during 1951–1980 are excluded.

[12] The GHCN data are raw precipitation data, unadjusted for gauge bias. Gauge bias is associated with several processes, including wind-induced catch errors, especially for snow, and evaporative losses from gauges. In the annual mean, such errors typically require an upward adjustment of measurements by 2% to 20%, with significantly larger values in some cold regions [*Legates and Willmott*, 1990]. Legates and Willmott adjusted gauge data for estimated biases, considering catch error (as function of wind speed and precipitation form) and wetting and evaporation errors. They interpolated the raw and adjusted estimates of long-term precipitation at gauges, for each month of the year, to a global 0.5° by 0.5° grid. Herein we employ the monthly ratios, f_m, of basin means of their adjusted fields to basin means of their raw fields as bias-adjustment factors for our precipitation estimates. Legates and Willmott also produced monthly and annual global fields of estimated standard error of the gauge adjustments,

and we use these to characterize that component of precipitation uncertainty.

2.2. River Discharge

[13] Most discharge time series were selected from the databases of the U. S. Geological Survey (USGS) for gauges in the United States and the Global Runoff Data Centre (GRDC) for gauges outside the United States. USGS data inventories and station information were obtained from *EarthInfo, Inc.* [1995a, 1995b, 1995c, 1995d], and daily flow data were obtained directly from USGS; flow time series from more than 20,000 sites were available for consideration. The GRDC data set contains several thousand daily and monthly time series from gauges worldwide, along with pertinent station information [*Global Runoff Data Center*, 1998]. The sources of data for a few additional rivers are listed in the acknowledgments.

[14] Regrettably, we are unable to characterize errors in the discharge data systematically. Where rating curves are stable but poorly known, systematic bias is possible. Where channels are unstable and detailed channel surveys are infrequent, significant random errors may arise. As is apparent from our analyses here and in subsequent papers, we believe discharge errors are generally small in comparison with precipitation errors. This belief is supported by analyses of *Milly and Shmakin* [2002a], in which our indices of precipitation error are found to be predictive of differences between modeled and measured river discharges, suggesting relatively minimal bias in discharge.

2.3. Radiation

[15] To aid in evaluation of our precipitation estimates, we also estimated energy availability at basin scale, using the surface radiation budget (SRB) data set of the NASA Langley Research Center. The SRB data set is based on satellite observations and parameterized broadband radiative transfer model calculations, with approximate inclusion of the effects of aerosols [*Darnell et al.*, 1988; *Gupta et al.*, 1992]. Satellite data include top of atmosphere radiances, atmospheric soundings, and cloud information from the International Satellite Cloud Climatology Project, and top of atmosphere clear-sky albedo from the Earth Radiation Budget Experiment. The SRB data set includes monthly, global fields of the components of the surface radiation balance for the period July 1983 through June 1991. An evaluation of the SRB data set has been presented by *Garratt et al.* [1998] in the context of an assessment of various global estimates of the surface radiation balance. All estimates evaluated in that study appeared to have a positive bias in surface net radiation over land. This bias was among the smallest (about 10% on average) for the SRB data set; the bias should not present a serious problem for the particular application of the data herein.

3. River Basin Selection and Discharge Data Analysis

3.1. Preliminary Selection of River Basins

[16] From the extensive data sets of river discharge data, monthly or daily time series were chosen for possible use in this study only if they satisfied the following criteria: (1) at least 30 years in length (occasionally relaxed to 19 years, mainly north of 55°N. due to dearth of data for high-latitude regions), (2) no more than 2% of records missing, and (3) drainage area greater than 10,000 km^2. The set of records so identified was further reduced. For the most part, nesting of basins was not permitted, in order to minimize redundancy. Some exceptions were made, however, as described here. A drainage area of 200,000 km^2 was used to divide the basins into two sets, "small" and "large." Inside either of these sets, no nesting of basins was permitted; drainage areas were nonoverlapping. Nesting of small basins within large basins was permitted, however. Where the no-overlap condition required exclusion of gauges, the time series retained usually were those with the longest and most complete record or the largest drainage area. An attempt was also made in the selection process to favor those basins whose flows would be least affected by reservoirs or other water resource projects that were apparent in *The Times Atlas of the World* [*Times Books*, 1988]. Discharge records were inspected both visually and automatically for overt errors, and suspect time series were excluded. This selection process led to a set of 186 basins.

3.2. Filling of Missing Values

[17] Missing discharge values were filled through an interpolation scheme that provided continuity in time with the measured discharges, and that relaxed toward an estimated mean seasonal cycle on a timescale of 2 months. Let ξ denote the normalized flow anomaly (difference between actual flow and expected flow for that time of year, divided by expected flow for that time of year). When data were missing between times t_1 and t_2, we filled the time series by means of the relation

$$\xi(t) = c_1 e^{-(t-t_1)/\tau} + c_2 e^{-(t_2-t)/\tau}, \quad (1)$$

in which τ is 2 months, and c_1 and c_2 are determined to match observations at both ends of the data gap. The form of (1) reduces to linear interpolation for data gaps much shorter than τ, and reduces to use of the long-term means at times far from the endpoints if the gap is much longer than τ. The estimated mean seasonal cycle was based on the unfilled time series. The timescale of 2 months is an average based on preliminary analyses of stream flow persistence. Filling was performed at either the daily or the monthly timescale, depending on the timescale of the original data. Subsequently, all daily time series were time-averaged to produce a uniform data set of continuous monthly time series.

3.3. Test for Homogeneity of Time Series

[18] As already mentioned, some attempt was made to exclude time series where human intervention may have caused direct, substantial modification of the flow regime. To provide an additional check against inhomogeneities during the period of record, we analyzed normalized flow duration curves (plots of flow normalized by decadal median flow against the fraction of time that flow is exceeded) for each basin. For each decade of record, a normalized flow-duration curve was constructed, and the set of curves for each basin was inspected visually for significant changes. Systematic long-term changes in the shape of the flow-duration curve were interpreted as evidence of significant artificial disturbance within the basin. When a disturbance was identified, the record was either discarded or trimmed to include only the inferred predisturbance record. For basins in the United States, we found that most

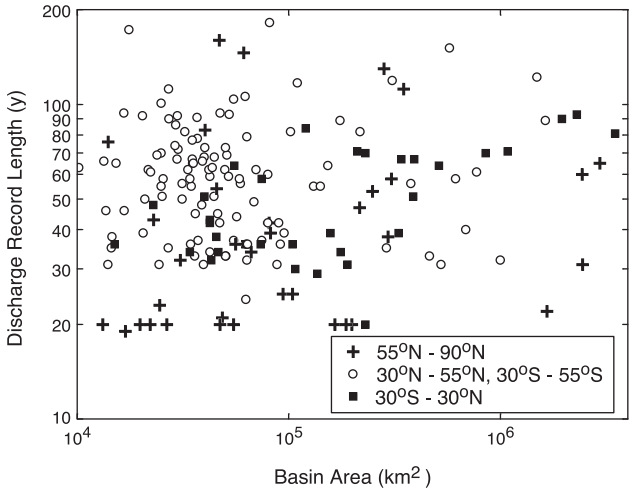

Figure 1. Scatterplot of discharge record length against basin area. Each symbol represents one of the 175 basins in the final data set. Type of symbol indicates latitude range that contains the center of the basin.

inferred inhomogeneities in the record were attributable to known overt hydrologic disturbances documented by *Ruddy and Hitt* [1990] and by *Slack et al.* [1993].

[19] Elimination of those basins with identified disturbances to the flow regime reduced the set of stations to 175. The distribution of basin area and record length is shown in Figure 1. Basin drainage areas range from 10,200 to 3.5×10^6 km^2, with a median value of about 51,000 km^2, and periods of record range from 19 to 182 years, with a median value of about 54 years. The large basins are relatively well distributed across the latitude zones. The largest number of small basins is found in the middle latitudes. As noted earlier, the constraint on minimum record length was relaxed in high latitudes, and this explains the concentration of small high-latitude basins with only about 20 years of record. The number and length of records from small tropical basins are not as great as for the middle latitudes, but are nevertheless substantial.

3.4. River Basin Delineation

[20] For each discharge record, the corresponding drainage basin was identified, at 1°-by-1° resolution, by use of information from *Oki and Sud* [1998] and from *Times Books* [1988]. Oki and Sud produced a global grid of river drainage directions. We used longitude-latitude information provided by USGS or GRDC for each discharge site to locate the gauge on the grid. The basin was then defined as the set of Oki-Sud grid cells draining through the cell containing the gauge. The implied river basins were checked against those shown in the atlas and against basin areas reported by USGS or GRDC for the gauge. In some cases, it was necessary to relocate the gauge to a neighboring cell on the Oki-Sud grid in order to capture the right basin. The resulting basin domains are mapped in Figure 2.

4. Precipitation: Analytic Framework

4.1. Conceptual Model of the Precipitation Process

[21] We view precipitation as a random spatial process and attempt to infer properties of the process from the finite set of measurements made available by the precipitation gauge network [*Rodríguez-Iturbe and Mejía*, 1974]. At any point in the basin, the precipitation amount P_{mn} (mass of water per unit area) during any month m of any year n is expressed as the sum of a long-term mean p_m for that month of the year and a monthly anomaly δP_{mn} for that particular year,

$$P_{mn} = p_m + \delta P_{mn}. \tag{2}$$

The time mean p_m is defined over the fixed time period; here we use the period 1951–80. The anomaly can be expressed as the product of a standard deviation of precipitation over this reference time period s_m and a normalized anomaly π_{mn},

$$\delta P_{mn} = s_m \pi_{mn}. \tag{3}$$

We shall assume that π_{mn} is uncorrelated in time (i.e., negligible autocorrelation at the monthly timescale) and has spatial correlation between two points that depends only on distance between points. In contrast, both the precipitation mean p_m and standard deviation s_m are allowed to vary spatially, and no assumptions are made regarding the statistical structure of those variations.

[22] The neglect of monthly autocorrelation of π_{mn} is clearly an approximation. Precipitation statistics can be sensitive to persistent anomalies in the state of the climate system, though generally a large component of monthly precipitation will be uncorrelated in time. The acceptability of our approximation can most readily be judged in terms of the success of the resultant error indices.

[23] It will be helpful to introduce additional notation for annual mean values. The precipitation amount in year n is denoted by P_n, the long-term annual precipitation is denoted by p_a, and the annual anomaly of precipitation in year n is denoted by δP_n. Thus

$$P_n = p_a + \delta P_n, \tag{4}$$

$$p_a = \sum_{m=1}^{12} p_m, \tag{5}$$

$$\delta P_n = \sum_{m=1}^{12} \delta P_{mn}. \tag{6}$$

4.2. Interpolation, Basin Areal Averaging, and Gauge Bias Adjustment

[24] We use angle brackets to denote the areal average of any variable over a basin (the basin mean). From (2),

$$\langle P_{mn} \rangle = \langle p_m \rangle + \langle \delta P_{mn} \rangle. \tag{7}$$

We use "hats" to denote our gauge-derived estimates of any variable. In view of (7), we shall require

$$\langle \hat{P}_{mn} \rangle = \langle \hat{p}_m \rangle + \langle \delta \hat{P}_{mn} \rangle. \tag{8}$$

Furthermore,

$$\langle \hat{P}_n \rangle = \langle \hat{p}_a \rangle + \langle \delta \hat{P}_n \rangle, \tag{9}$$

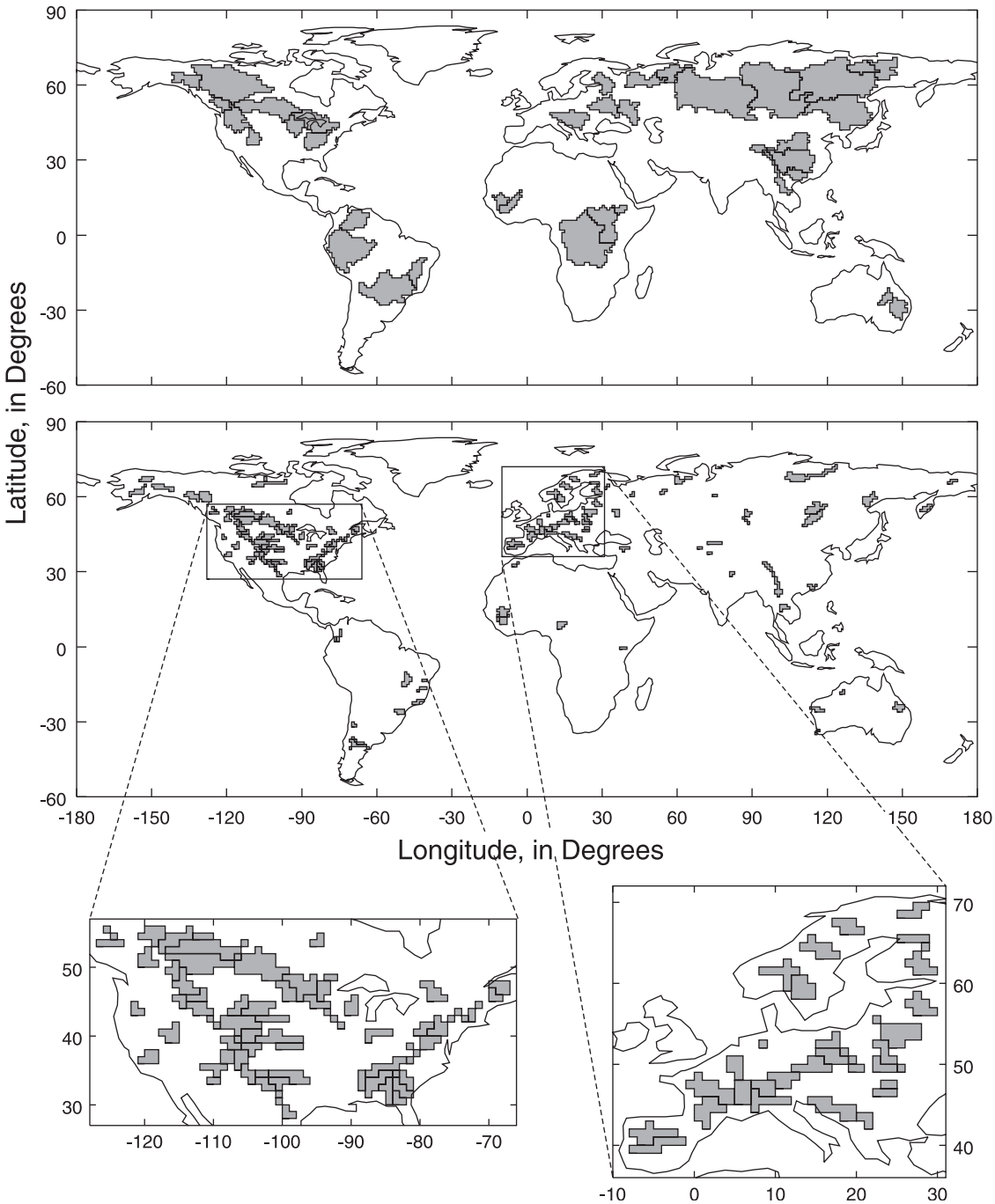

Figure 2. Maps showing basins used in this study. The top panel shows large basins (area greater than 200,000 km^2); the bottom panel shows small basins, with insets for central North America and Europe.

where

$$\langle \hat{p}_a \rangle = \sum_{m=1}^{12} \langle \hat{p}_m \rangle, \tag{10}$$

$$\langle \delta \hat{P}_n \rangle = \sum_{m=1}^{12} \langle \delta \hat{P}_{mn} \rangle. \tag{11}$$

We describe first the computation of $\langle \hat{p}_m \rangle$, and then the similar computation of $\langle \delta \hat{P}_{mn} \rangle$.

[25] Our spatial analyses are based on the approximation of continuous spatial fields by a uniform longitude-latitude grid of small cells; values are assumed to be constant in space within a cell. In this sense, our approach is similar to that of *Morrissey et al.* [1995]. For results presented here, we used a 0.25° longitude-latitude cell size; sensitivity studies indicated that results were essentially unchanged when finer grids were used. (Uniform longitude-latitude spacing may not be the most efficient approach at high latitudes, but it is convenient and presumably converges when sufficiently small cells are used.) To estimate the basin mean of any variable, we first interpolate from

observations at gauges to cells on the grid, and then compute area-weighted means of the grid-cell values. Our chosen interpolation scheme assigns to each cell the value of the interpolated variable at the nearest available gauge. Thus the estimated value $\hat{p}_{mk}$ of p_m at cell k is given by

$$\hat{p}_{mk} = p_m^{i(k)}, \tag{12}$$

where the superscript index function $i(k)$ specifies the index i of the gauge nearest to the center of cell k; $p_m^{i(k)}$ denotes the estimated precipitation at that gauge. If our gauge observations were unbiased, the basin mean precipitation could then be found by areal averaging,

$$\langle \hat{p}_m \rangle = \sum_{k=1}^{K} a_k p_m^{i(k)}, \tag{13}$$

in which a_k is the fraction of basin area contained in cell k, and K is the number of cells in the basin. This approach is equivalent to the well-known use of Thiessen polygons for areal averaging when the grid cells are sufficiently small. To account for gauge bias, though, we modify (13) by introducing estimated values of the monthly, basin-specific gauge-adjustment factors, f_m, mentioned in the Data section,

$$\langle \hat{p}_m \rangle = f_m \sum_{k=1}^{K} a_k p_m^{i(k)}. \tag{14}$$

Implicit in this application of quantitative results from *Legates and Willmott* [1990] is the assumption that bias characteristics of gauges included in our analysis are similar, on average within a given basin, to those included in the former analysis.

[26] The areal averaging of monthly anomalies over the basin follows a similar procedure. Because values of standard deviation are computed from raw gauge precipitation values, they also require adjustment for gauge bias. Thus

$$\langle \delta \hat{P}_{mn} \rangle = f_m \sum_{k=1}^{K} a_k s_m^{i(k)} \pi_{mn}^{i(k)}, \tag{15}$$

in which $s_m^{i(k)}$ and $\pi_{mn}^{i(k)}$ are defined in analogy to $p_m^{i(k)}$. The gridded estimates of s_m and π_{mn} are analyzed separately. The s_m field is time-invariant and is estimated here over the 1951–1980 period. In general, the s_m field is based on a larger set of gauges than the π_{mn} field in any given year, because some gauges used in the analysis do not have data in every month and year. In this sense, the index functions $i(k)$ generally differ between these two variables (and vary over time, in the case of π_{mn}), but, for simplicity, we have avoided including this in the notation. Also, as a result of the variations in the observational network over time, the $\langle \delta \hat{P}_{mn} \rangle$ defined by (15) have a time mean that differs slightly from zero, so the $\langle \delta \hat{P}_{mn} \rangle$ series is subsequently adjusted by subtracting that mean after application of (15).

4.3. Characterization of Errors in Precipitation Estimates

[27] Errors in our estimates of long-term monthly means and individual monthly anomalies are defined, respectively, by

$$\varepsilon_m = \langle \hat{p}_m \rangle - \langle p_m \rangle, \tag{16}$$

$$\varepsilon_{mn} = \langle \delta \hat{P}_{mn} \rangle - \langle \delta P_{mn} \rangle, \tag{17}$$

and the error in the estimate of monthly precipitation is the sum of these two errors,

$$\varepsilon_m + \varepsilon_{mn} = \langle \hat{P}_{mn} \rangle - \langle P_{mn} \rangle. \tag{18}$$

We define errors in annual totals by

$$\varepsilon_a = \sum_{m=1}^{12} \varepsilon_m, \tag{19}$$

$$\varepsilon_n = \sum_{m=1}^{12} \varepsilon_{mn}. \tag{20}$$

Together, these relations quantify the overall error in estimates of annual total precipitation,

$$\varepsilon_a + \varepsilon_n = \langle \hat{P}_n \rangle - \langle P_n \rangle. \tag{21}$$

[28] In the next two sections, we present methods for characterizing the magnitudes of ε_a and ε_n. We do not treat ε_m, but instead work directly with ε_a. Having made no statistical assumptions about the nature of spatial variations in long-term mean fields, we employ heuristic methods for characterization of ε_a. For ε_n, we first work at the monthly timescale to characterize ε_{mn} statistically by means of standard correlation-based techniques, and then use (20) to characterize error in the annual totals.

5. Errors in Estimates of Long-Term Mean Precipitation

5.1. Identification of Components

[29] Errors in long-term basin mean annual precipitation estimates (ε_a) arise as a result of imperfect adjustments for measurement errors at the gauges and of inadequate spatial sampling of the field by the gauge network. We distinguish two types of spatial sampling error. Because it is widely recognized that orographic precipitation influences are the source of major systematic errors in precipitation networks, we give special attention to them. Even in the absence of orographic effects, however, sampling errors are present. Herein we develop separate models for nonorographic and orographic sampling errors, denoted by ε_{as} and ε_{ao}, respectively, and we assume that these errors are additive and independent. Furthermore, we make the additional approximation that both are independent of the error associated with adjustment for gauge measurement errors, which we denote by ε_{ag}. Thus

$$\varepsilon_a = \varepsilon_{as} + \varepsilon_{ao} + \varepsilon_{ag}. \tag{22}$$

Generally, one can think of ε_{as} and ε_{ag} as random constant errors that could be of either sign, while ε_{ao}, if significant, typically reflects a negative bias due to insufficient sampling of precipitation at higher elevations, where forced lifting and cooling of air masses tends to enhance precipitation.

5.2. Gauge Error

[30] Because we use the gauge adjustment factors of *Legates and Willmott* [1990], we base our estimates of the

residual measurement error (i.e., that remaining after adjustment) on the standard errors of adjustments that were also provided as monthly grids by Legates and Willmott. We represent ε_{ag} as a variable having zero expected value and variance equal to σ_{ag}^2. Under the approximation that gauge bias and, hence, error in its estimation are highly correlated regionally (i.e., within a basin), we estimate σ_{ag}^2 as the basin mean value of the squares of standard errors of adjustment reported by Legates and Willmott.

5.3. Nonorographic Spatial Sampling Error

[31] We expect ε_{as} to be small in a basin with a dense gauge network and large in a basin with a sparse network. The relation between the statistics of ε_{as} and network density, however, is expected to depend on the size of the basin and the nature of (nonorographic) heterogeneity of climate within the basin, which will depend on a variety of geographic factors. Rather than try to identify these and their influences on heterogeneity, we instead use precipitation observations from the existing network to develop a heuristic estimate of ε_{as}. We hypothesize that the sensitivity of the computed basin mean to the removal of stations from the existing network is an indicator of the variance of the sampling error ε_{as} associated with that network. For example, if removal of a few stations from the network would change the estimate of the mean, then chances are that the existing network is inadequate. On the other hand, if the estimate does not change when many of the stations are removed, then the estimate is probably a good one.

[32] Quantitative expression of this idea requires additional notation. To make explicit the dependence of the basin mean estimate on any precipitation gauge network N, we now write $\langle \hat{p}_a \rangle$ as $\langle \hat{p}_a(N) \rangle$. Let N_o denote the full available network, and let $N_o^{1/2}$ be a network generated by randomly choosing half of the stations from the full network. We define a characteristic sampling error $S(N_o)$ of the network N_o by

$$[S(N_o)]^2 = E\left\{ \left[\left\langle \hat{p}_a\left(N_o^{1/2}\right) \right\rangle - \langle \hat{p}_a(N_0) \rangle \right]^2 \right\} \qquad (23)$$

in which $E\{\ \}$ denotes the expectation operator. This measure can be estimated directly from the available observations.

[33] We hypothesize that ε_{as} is a random variable of mean zero and standard deviation proportional to $S(N_o)$. To test the hypothesis, and to estimate the hypothesized constant of proportionality, we performed numerical experiments on a few of the basins with the best gauge networks. Quality of the gauge network was assessed subjectively on the basis of gauge density, perceived freedom from orographic effects, and other factors. (Orography-related sampling errors are analyzed separately, as discussed in the next section.) The basin mean precipitation estimated from the full available network was assumed to have negligible error. Then the network was repeatedly degraded by random sampling to a series of subnetworks, with varying numbers of gauges. For each basin, 200 subnetworks having half the total number of gauges were generated randomly and independently. Then 200 additional subnetworks having only a quarter of the total available gauges were created. This process of subnetwork generation was repeated (with rounding of the number of stations to the lower integer) to as many levels as possible. For each subnetwork N at any of these levels, our procedure for estimation of annual-total long-term basin mean precipitation $\langle \hat{p}_a(N) \rangle$ was applied, and $S(N)$ was computed. (To evaluate $S(N)$, we created 200 independent subsubnetworks $N^{1/2}$ of N and found $S(N)$ as the root-mean-square difference between the subsubnetwork means and the full subnetwork mean.) Finally, for each subnetwork N, we formed the ratio

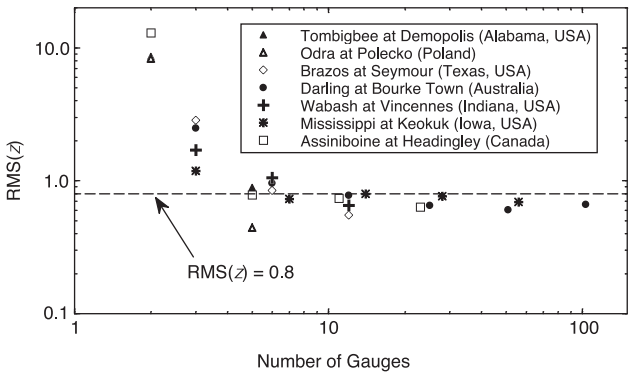

Figure 3. Scatterplot of root-mean-square value of the variate z defined by (24) against the number of gauges retained in the subnetwork N for seven selected river basins believed to have minimal orographic errors. Each symbol represents the root-mean-square value across 200 realizations of a subnetwork of given size. Near-constancy of this statistic supports use of the index $S(N)$ to estimate nonorographic standard error of long-term basin mean annual precipitation.

$$z = \frac{\langle \hat{p}_a(N) \rangle - \langle \hat{p}_a(N_o) \rangle}{S(N)}. \qquad (24)$$

If the full network N_o is sufficiently dense, the numerator of (24) is the sampling error associated with subnetwork N. It would follow from our hypothesis that the variate z is randomly distributed with a mean of zero and a standard deviation that is independent of the basin and the subnetwork.

[34] Figure 3 shows that the root-mean-square value of z (defined for a given network size on a given basin) is relatively constant, except for networks containing very few gauges. The increase in variance of z for small networks is associated with the resulting poor characterization of error by $S(N)$ in the denominator of (24). The expected value of z differs slightly from zero for any given network size on any given basin. We found mean values of z ranging from -0.22 to 0.33, considering all cases tested with five or more gauges. Additionally, for a given basin and network size, the distribution of z differs from the normal distribution by having higher density in the tails, particularly when the number of gauges is small. Overall, however, the constancy of the root-mean-square value of z seen for sufficiently large n supports our hypothesis that $S(N)$ is a measure of network error, except when the network consists of very few gauges. As a crude approximation, we shall assume that the nonorographic spatial sampling error for a given network has a

mean of zero and a standard deviation equal to $0.8S(N)$ for networks with five or more gauges; the coefficient 0.8 is representative of the root-mean-square value of z, as illustrated by Figure 3. Thus

$$\sigma_{as} = 0.8S(N_o), \tag{25}$$

where σ_{as} is the standard deviation of ε_{as}.

[35] This analysis clearly has limitations. When the siting of gauges in a network N is biased toward those locations having only low (or high) precipitation, for example, the index $S(N)$ cannot reflect conditions in the unsampled locations, and error in analyzed precipitation will be underestimated. The one common example of such a situation is the case of orographic precipitation, which is treated next.

5.4. Orographic Spatial Sampling Error

[36] Typically, high-elevation locations, which tend to be associated with elevated precipitation amounts, are undersampled in precipitation networks. Given the failure of the network to capture the small-scale features of orographically induced precipitation, it is anticipated that basin means computed by normal interpolation methods will be subject to substantial negative bias. To examine this problem, we assume, as a gross approximation, that precipitation within a given basin is linearly related to elevation. Thus

$$p_a = \langle p_a \rangle \left[1 + (Z - \langle Z \rangle)/Z_p\right], \tag{26}$$

in which Z is elevation and Z_P is an elevation scale defining the relative sensitivity of precipitation to elevation. If the random errors in the gauge bias adjustments are ignored, it can be shown to follow that

$$\langle \hat{p}_a(N_o) \rangle = \langle p_a \rangle [1 + (\langle \hat{Z}(N_o) \rangle - \langle Z \rangle)/Z_p], \tag{27}$$

in which $\langle \hat{Z}(N_o) \rangle$ is the estimate of $\langle Z \rangle$ obtained by interpolation of gauge elevations (included in the GHCN data set), by means of the same scheme as that used for precipitation. Rearrangement of (27) leads to

$$[\langle \hat{p}_a(N_o) \rangle - \langle p_a \rangle]/\langle p_a \rangle = [\langle \hat{Z}(N_o) \rangle - \langle Z \rangle]Z_p. \tag{28}$$

For convenience in application, however, we scale the error by the estimated mean rather than the true mean, obtaining

$$[\langle \hat{p}_a(N_o) \rangle - \langle p_a \rangle]/\langle \hat{p}_a(N_o) \rangle = [\langle \hat{Z}(N_o) \rangle - \langle Z \rangle]/Z_p. \tag{29}$$

[37] Equation (29) expresses our hypothesis that the relative error in precipitation associated with failure of the network to capture orographic precipitation can be assessed using information about basin mean elevation. For testing (29), we did not feel that the gauge networks were sufficiently dense to evaluate $\langle p_a \rangle$ accurately even in the best gauged of our orographically influenced basins, so the subsampling approach of the previous section was not used. Rather, we obtained an independent estimate of $\langle p_a \rangle$ using the semiempirical water balance theory of *Budyko* [1974]. According to this theory, the evaporation ratio (ratio of long-term basin mean annual evaporation, $\langle E_a \rangle$, to long-term basin mean annual precipitation, $\langle p_a \rangle$) is a unique function of the radiative index of dryness. The index of dryness is the ratio of long-term mean net surface radiation, expressed as an equivalent annual evaporation amount, $\langle R_a \rangle$, to $\langle p_a \rangle$. Thus the value of $\langle p_a \rangle$ predicted by Budyko's theory, $\langle p_a \rangle_B$, obeys the relation

$$\frac{\langle E_a \rangle}{\langle p_a \rangle_B} = \phi_B\left(\frac{\langle R_a \rangle}{\langle p_a \rangle_B}\right), \tag{30}$$

in which [*Budyko*, 1974]

$$\phi_B(x) = \left[x\left(\tanh x^{-1}\right)(1 - \cosh x + \sinh x)\right]^{1/2}. \tag{31}$$

In application of (30), we evaluate $\langle E_a \rangle$ as the difference between $\langle p_a \rangle_B$ and long-term mean gauged discharge per unit basin area at the basin outlet. Values of $\langle R_a \rangle$ are obtained as 8-year averages of net radiation from the SRB radiation data set, averaged over the basin area and divided by the latent heat of vaporization of water. We recognize that substantial error is possible when (30) is used to estimate precipitation, but we expect that any associated bias is small compared to the orographic error under investigation.

[38] Figure 4 is a scatterplot of estimated precipitation error against basin mean elevation error. The few outlier points can be explained readily in terms of the previous analysis of spatial sampling error in the absence of orographic effects. All but one of the outliers represent basins having either a very small number of gauges (fewer than five) or a relatively large value of the index $S(N_o)$. (The exception is one basin that by chance, has a very large value of $\langle \hat{Z}(N_o) \rangle - \langle Z \rangle$ but for which the apparent precipitation error is not remarkably large.) This provides additional support for the hypothesis that the magnitude of $S(N_o)$ is indicative of standard error in basin mean precipitation.

[39] Figure 4 also shows a significant correlation within the central cloud of points. To develop a relation, we performed a regression, excluding those basins expected to have large error even in the absence of orographic effects and the single basin with $\langle \hat{Z}(N_o) \rangle - \langle Z \rangle$ less than -2500 m. Results of the regression suggest that a precipitation-gauge network will tend to underestimate basin mean precipitation by about 10% on average, for every (-105 m $\pm$ 10 m) of elevation error , for elevation error in the range of -1200 m to 200 m. The strength of the correlation (r = 0.65) is indicative of the power of this index as a measure of orographic bias of a precipitation network. The failure of the excluded basin with $\langle \hat{Z}(N_o) \rangle - \langle Z \rangle$ less than -2500 m to fit on this line is not surprising; it is not reasonable to expect a linear regression to hold over such a large range in elevation, because humidity generally decreases exponentially with height. The scatter around the line presumably reflects random errors in both $\langle \hat{p}_a(N_o) \rangle$ and $\langle p_a \rangle_B$.

[40] We use Figure 4 as the basis for our estimate of ε_{ao},

$$\varepsilon_{ao} = \langle \hat{p}_a(N_o) \rangle \left(\langle \hat{Z}(N_o) \rangle - \langle Z \rangle\right)/(1050m). \tag{32}$$

Note that this model differs fundamentally from our models of ε_{ag} and ε_{as}, which describe random errors having zero

expected value. In the case of ε_{ao}, we focus instead on the expected value of the error, which we know is large, and ignore the random component, which we cannot characterize well.

[41] It is desirable to have an independent check on the usefulness of $\langle\hat{Z}(N_o)\rangle - \langle Z\rangle$ as a predictor of basin mean orographic precipitation bias. For this purpose, we obtained a high-resolution analysis of precipitation during 1961–90 over part of the United States from the Oregon Climate Service; the analysis was conducted by use of the expert system of *Daly et al.* [1994] for estimation of climate variables in mountainous terrain. Let $\langle p_a\rangle_D$ denote the basin mean precipitation determined from this data set. We now substitute $\langle p_a\rangle_D$ for $\langle p_a\rangle_B$ in the test of the $\langle\hat{Z}(N_o)\rangle - \langle Z\rangle$ index. Figure 5 shows that $\langle p_a\rangle_B$ implies a significantly larger sensitivity of the bias to the elevation index than does $\langle p_a\rangle_D$; note that the same subset of basins is used for both relations in Figure 5 for consistency. The difference in sensitivities could be explained by elevation-dependent systematic errors in $\langle p_a\rangle_B$ and/or $\langle p_a\rangle_D$. Errors in $\langle p_a\rangle_B$ could arise from the Budyko equation's neglect of nonclimatic factors controlling runoff, such as a possible reduction in soil water holding capacity with elevation, leading to increased tendency for runoff [*Milly*, 1994]. Errors in $\langle p_a\rangle_D$ could arise from the insufficiency of observations at high elevations.

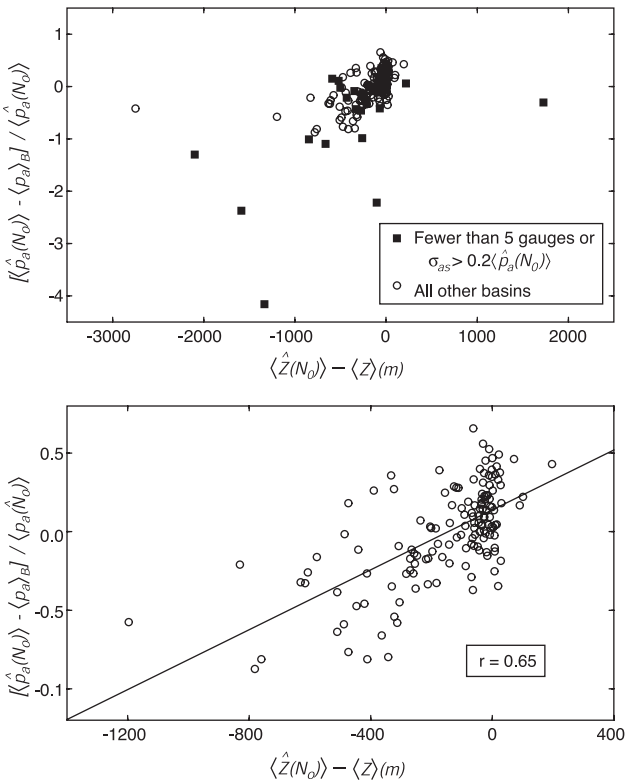

Figure 4. Scatterplots of apparent relative orographic error in precipitation $\left[\langle\hat{p}_a(N_o)\rangle - \langle p_a\rangle_B\right]/\langle\hat{p}_a(N_o)\rangle$ against error in basin mean interpolated elevation $\langle\hat{Z}(N_o)\rangle - \langle Z\rangle$. Top panel shows full data set, with solid symbols representing those basins with fewer than five precipitation gauges or with $\sigma_{as} > 0.2\langle\hat{p}_a(N_o)\rangle$; these and the point for which $\langle\hat{Z}(N_o)\rangle - \langle Z\rangle$ is less than −2500 m are excluded to produce the bottom panel, which shows regression line fitted to the restricted set of points. The true basin mean elevation $\langle Z\rangle$ is evaluated using the ETOPO5 5-min global elevation data set, obtained from the National Geophysical Data Center. The regression equation (with standard errors of estimation) is $\left[\langle\hat{p}_a(N_o)\rangle - \langle p_a\rangle_B\right]\langle\hat{p}_a(N_o)\rangle = (0.139 \pm 0.024) + \left(\langle\hat{Z}(N_o)\rangle - \langle Z\rangle\right)/(1050m \pm 105m)$.

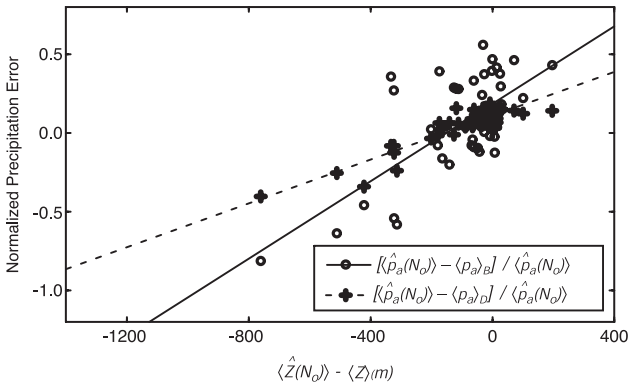

Figure 5. Same as bottom panel of Figure 4 but for the subset of basins for which a detailed expert system analysis of precipitation is available from the Oregon Climate Service. Also shown are similar results obtained using results from the expert system analysis $\langle p_a\rangle_D$ rather than the estimate from *Budyko*'s [1974] equation $\langle p_a\rangle_B$. Slopes are $(810\ m \pm 140\ m)$ for $\langle p_a\rangle_B$ and $(1430\ m \pm 100\ m)^{-1}$ for $\langle p_a\rangle_D$.

[42] By ignoring orography in our interpolation procedure, we have obtained substantial orographic errors in some basins. One could argue for inclusion of topographic information in the interpolation procedure [e.g., *Daly et al.*, 1994] to develop more accurate precipitation estimates. We believe, however, that such an approach might be ill posed for our global-scale application, because many of our basins presumably lack a sufficient gauge network for the analysis. Furthermore, such an approach would considerably complicate some of our error analyses. As a less ambitious alternative, we could have used a relation such as (32) to adjust our estimates for orographic errors. Given the simplistic nature of (32), however, we are more comfortable applying it to characterize precipitation errors in a gross sense than to remove those errors.

6. Errors in Precipitation Anomalies

6.1. Estimation of Sampling Error

[43] We now turn to estimation of the characteristics of the random error in $\langle\delta\hat{P}_{mn}\rangle$ associated with inadequate sampling of π_{mn}. This error is a random variable, defined by (17) as the difference between estimated and actual values of the basin mean anomaly. We can express the actual basin mean anomaly as

$$\langle\delta P_{mn}\rangle = f_m \sum_{k=1}^{K} a_k s_m^{i(k)} \pi_{mnk}, \tag{33}$$

in which π_{mnk} designates the (unknown) actual normalized anomaly at cell k. Implicit in this expression are errors of discretization and errors of interpolation of the standard deviation field, both of which will henceforth be ignored.

The factor f_m is included in recognition of the fact that gauge values of standard deviation are computed without adjustment for gauge bias. Combination of (15), (17), and (33) yields

$$\varepsilon_{mn} = f_m \sum_{k=1}^{K} a_k s_m^{i(k)} \pi_{mn}^{i(k)} - f_m \sum_{k=1}^{K} a_k s_m^{i(k)} \pi_{mnk} = f_m \sum_{k=1}^{K} a_k s_m^{i(k)} \left(\pi_{mn}^{i(k)} - \pi_{mnk} \right), \quad (34)$$

We can now derive the variance σ_{mn}^2 of ε_{mn} as its expected squared value,

$$\sigma_{mn}^2 = f_m^2 E\left\{ \left[\sum_{k=1}^{K} a_k s_m^{i(k)} \left(\pi_{mn}^{i(k)} - \pi_{mnk} \right) \right]^2 \right\}, \quad (35)$$

which leads to

$$\sigma_{mn}^2 = f_m^2 \sum_{k=1}^{K} \sum_{l=1}^{K} a_k a_l s_m^{i(k)} s_m^{i(l)} \left[E\left\{ \pi_{mn}^{i(k)} \pi_{mn}^{i(l)} \right\} - 2E\left\{ \pi_{mn}^{i(k)} \pi_{mnl} \right\} + E\{ \pi_{mnk} \pi_{mnl} \} \right], \quad (36)$$

or

$$\sigma_{mn}^2 = f_m^2 \sum_{k=1}^{K} \sum_{l=1}^{K} a_k a_l s_m^{i(k)} s_m^{i(l)} \left[\rho_m\left(d_{i(k),i(l)} \right) - 2\rho_m\left(d_{i(k),l} \right) + \rho_m\left(d_{k,l} \right) \right], \quad (37)$$

in which $\rho_m(d)$ is the correlation of π_{mn} as a function of distance d during month m of the year. The three distances in (37) denote, respectively, the distance from the gauge nearest cell k to the gauge nearest cell l, the distance from the gauge nearest cell k to cell l, and the distance from cell k to cell l. Under our assumption of negligible intermonthly correlation, the variance of the error in $\langle \delta \hat{P}_n \rangle$ can be computed as the sum of the monthly error variances,

$$\sigma_n^2 = \sum_{m=1}^{12} \sigma_{mn}^2. \quad (38)$$

Finally, we note that the monthly and annual variances of basin mean precipitation are given by

$$Var(\langle P_{mn} \rangle) = f_m^2 \sum_{k=1}^{K} \sum_{l=1}^{K} a_k a_l s_m^{i(k)} s_m^{i(l)} \rho_m\left(d_{k,l} \right), \quad (39)$$

$$Var[\langle P_n \rangle] = \sum_{m=1}^{12} Var[\langle P_{mn} \rangle]. \quad (40)$$

[44] The application of (37) and (39) requires knowledge of the spatial correlation function $\rho_m(d)$ of monthly precipitation for each month of the year. The conventional approach is to estimate it on the basis of precipitation records from the basin in question. In our application, many basins do not have an adequate network and period of record to support such computations. Instead, we have adopted an alternate approach. Using the GHCN data set, we performed a global correlation analysis with the function

$$\rho_m(d) = \exp[-(d/d_{om})^{\mu_m}] \quad (41)$$

and developed a 2°-by-2° global grid of the two parameters d_{om} and μ_m for each month, filling missing 2° cells by nearest-neighbor interpolation. Subsequently, the values of the parameters were area-averaged over each basin, and the basin mean values were used in conjunction with (41) to carry out the computations in (37) and (39).

6.2. Test of Error Estimates

[45] To assess the accuracy of our error estimates, we again used subsampling techniques on basins having dense gauge networks and minimal orographic errors. We used the same seven basins that were selected earlier for the evaluation of $S(N)$. For each basin, we estimated the time series of basin mean anomalies and their standard errors, using the full network N_o and using subnetworks N composed of one-half, one-quarter, and one-eighth of the full network. For each subnetwork, we formed the monthly and annual time series w_{mn} and w_n, defined as normalized errors in the estimates of the monthly and annual time series,

$$w_{mn} = \frac{\langle \delta \hat{P}_{mn}(N) \rangle - \langle \delta \hat{P}_{mn}(N_o) \rangle}{\delta_{mn}(N)}, \quad (42)$$

$$w_n = \frac{\langle \delta \hat{P}_n(N) \rangle - \langle \delta \hat{P}_n(N_o) \rangle}{\delta_n(N)}. \quad (43)$$

If N_o is sufficiently dense to give a good estimate of the true anomaly, and if our expressions for σ_{mn}^2 and σ_n^2 are sufficiently good estimates of the variances of estimation errors, then the measures w_{mn} and w_n should have means of zero and standard deviations of unity. For each of the three (1/2, 1/4, and 1/8) subnetworks on each of the seven basins considered, we computed the mean and standard deviation of these measures over the period of record. For w_{mn}, the extreme values of the 21 standard deviations were 0.74 and 1.24, and 14 of the values were in the range from 0.9 to 1.1. For w_n, the extreme values were 0.80 and 1.15, and 14 values were in the range from 0.9 to 1.1. These results suggest that σ_{mn} and σ_n provide accurate measures of sampling errors in our estimated anomalies.

[46] As a second, less direct, test of our correlation-based method for estimating ε_n, we tested the ability of the same method to estimate the variance of basin mean precipitation. To do this, we compared two estimates of the standard deviation of basin mean annual precipitation. The first estimate was computed as the standard deviation of the estimated time series $\hat{P}_n$. The second estimate was computed as $[Var(\langle P_n \rangle)]^{1/2}$ from (40). Note that the latter quantity depends only on the estimated correlation structure, and not on any specific observations. The comparison is presented in Figure 6. The two measures of variability are very well correlated, although there is a consistent tendency for the values derived from time series of $\hat{P}_n$ to exceed the theoretical values $[Var(\langle P_n \rangle)]^{1/2}$. This bias could result from

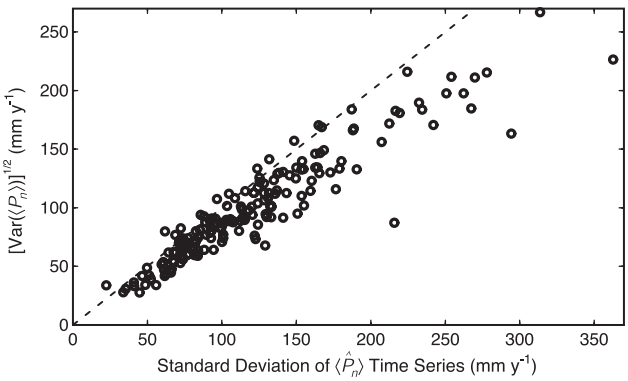

Figure 6. Scatterplot of $[Var(\langle P_n \rangle)]^{1/2}$ estimated from (40) against the standard deviation of the time series $\langle \hat{P}_n \rangle$.

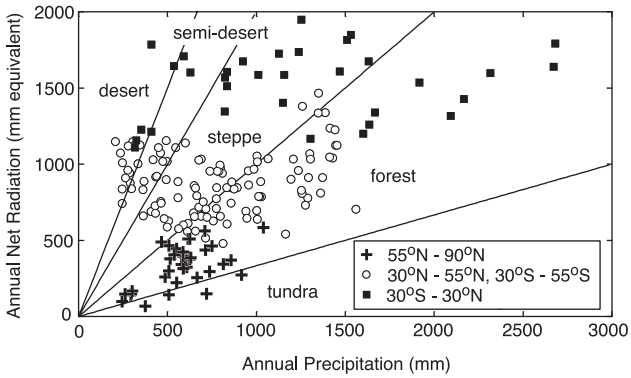

Figure 7. Scatterplot of annual net radiation (expressed as equivalent evaporative flux) against annual precipitation estimated by (30). Each symbol represents one basin in the data set. Type of symbol indicates latitude range that contains the center of the basin. Biome boundaries are defined as by *Budyko* [1974].

a systematic error in the procedure for estimation of $[Var(\langle P_n \rangle)]^{1/2}$, but it seems more likely that it may result from excessive variance in the $\hat{P}_n$ time series caused by allowing the variability at any gauge point to extend coherently to all of its nearest-neighbor cells.

7. Data Set Overview and Sample Results

7.1. Overview of Data Set

[47] An overview of time series length and basin area dependent on major latitude zones has already been presented in Figure 1, and the geographical distribution of basins was shown in Figure 2. Here we present the distribution of basins in climate space, using precipitation and net radiation as the two main climatic factors (Figure 7). Also shown in Figure 7 are approximate climatic boundaries of major world biomes, following definitions of *Budyko* [1974]. The data set spans climate space and associated biomes reasonably well. The very small number of points in tundra environment is consistent with the small areal coverage of global land by tundra. The heaviest concentration of basins is, appropriately, in the forest and steppe (grassland, savanna) biomes, which are the major participants in the global water cycle. Fewer basins are located in the more arid desert and semidesert biomes.

[48] To provide an overview of precipitation errors in our data set, we use two parameters. The first (ψ_a) is a dimensionless measure of the typical size of ε_a relative to the basin mean precipitation,

$$\psi_a = \left[E\{\varepsilon_a^2\}\right]^{1/2} / \langle \hat{p}_a \rangle. \tag{44}$$

This characteristic relative error in the mean, evaluated by use of the various assumptions and models that have been developed herein, can be expressed as

$$\psi_a = \left[\left(\frac{0.8 S(N_o)}{\langle \hat{p}_a \rangle} \right)^2 + \left(\frac{\langle \hat{Z}(N_o) \rangle - \langle Z \rangle}{1050\, m} \right)^2 + \left(\frac{\sigma_{ag}}{\langle \hat{p}_a \rangle} \right)^2 \right]^{1/2}. \tag{45}$$

The second error parameter used in presentation of the results is the mean over the period of discharge record (denoted by overbar) of the variance of errors in annual anomalies of basin mean precipitation, normalized by the variance of basin mean precipitation,

$$\overline{\psi_n} = \overline{\sigma_n^2} / Var(\langle P_n \rangle) \tag{46}$$

this index characterizes the accuracy of estimates of interannual variability of precipitation. Figure 8 provides an overview of estimated basin mean precipitation errors in the data set, showing the distribution of ψ_a and $\overline{\psi_n}$, stratified by three latitude zones. Median values of these measures are 0.11 for the characteristic relative error in the mean (ψ_a) and 0.076 for the normalized anomaly error variance ($\overline{\psi_n}$). Characteristic relative errors in the mean are greater than 32% for one quarter of the basins. Most of the basins having the most accurate precipitation estimates are located in the middle latitudes, although a midlatitude location is in no way a guarantee of high accuracy. High-latitude basins generally have the largest errors.

[49] Relative magnitudes of individual terms in (45) indicate relative importance of various types of errors. In

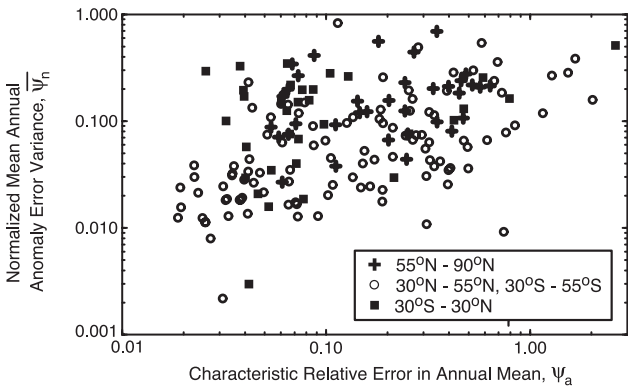

Figure 8. Scatterplot characterizing errors in estimates of annual, basin mean precipitation in the data set. Each symbol represents one basin. Type of symbol indicates latitude range that contains the center of the basin. The characteristic relative error in the mean is defined by (44), and the normalized anomaly error variance is defined by (46).

the basins with the larger errors (i.e., ψ_a above the median), the orographic error is almost always the dominant term, and the spatial sampling error generally is also significant, while the error in gauge bias adjustment is relatively small. In the basins with the smaller errors, any one of the three types of error may dominate or all three may be of the same order of magnitude. Across the data set, therefore, orographic sampling bias is the source of the largest errors in long-term mean precipitation estimates.

7.2. Gauge Density as an Index of Precipitation Error

[50] Gauge density is often used as an indirect measure of the quality of precipitation estimates obtained from a gauge network. Figure 9 shows the relation of characteristic relative error in annual basin mean precipitation to gauge density for the basins considered here. The large scatter implies that gauge density is very poorly correlated with error in the mean. It does appear, however, that the errors tend to decrease, on average, as gauge density increases from 100 gauges/10^6 km^2 to 1000 gauges/10^6 km^2. If gauge density is the only available measure of error, and if a characteristic error of 10% is considered acceptable, Figure 9 implies that a network of 600–1000 gauges/10^6 km^2 will usually be sufficient. Such a density is very high, however, and is found only in a very small fraction of the global land area.

[51] Gauge density, when combined with basin area, does provide a measure of anomaly error variance, as can be seen in Figure 10. Indeed, such a result is expected theoretically if interbasin differences in such additional factors as basin shape, distribution of gauges in the network, and precipitation spatial correlation structure among basins are ignored [*Rodríguez-Iturbe and Mejía*, 1974].

7.3. Sample Time Series

[52] Sample time series of precipitation and discharge data are presented in Figure 11. The Tanana River in Alaska has a drainage area of about 66,000 km^2 above the gauge at Nenana. Estimation errors of the anomalies are rather large, but are sufficiently small that the precipitation anomalies correlate fairly well with the discharge anomalies, especially on the interdecadal timescale. The correlation between precipitation and discharge provides credibility for both records, and also suggests the possibility of estimating stream flow that occurred before the stream gauge record began in the 1960s. The anomaly estimation errors, however, are greater for the first half of the century than for more recent years, as a result of the absence of critical precipitation gauges before about 1940.

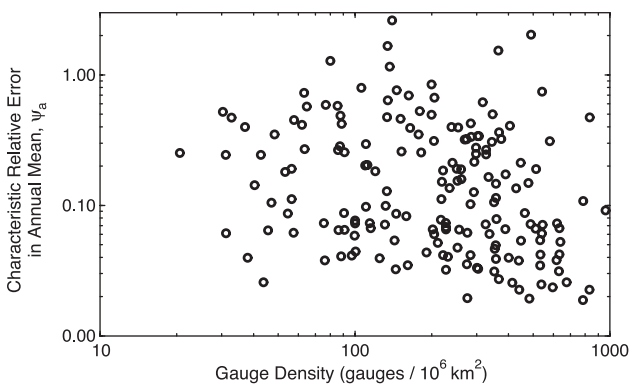

Figure 9. Scatterplot of characteristic relative error in annual basin mean precipitation against gauge density.

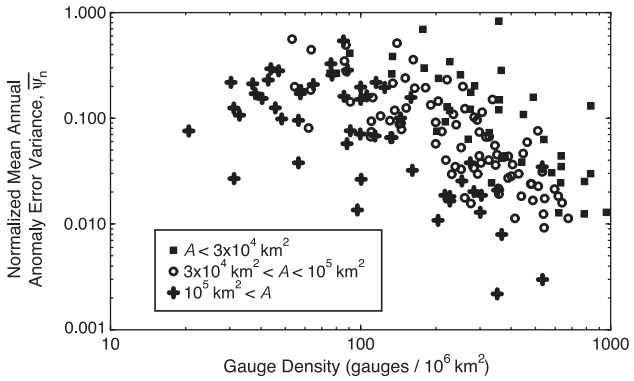

Figure 10. Scatterplot of normalized mean annual anomaly error variance against gauge density, stratified by basin area A.

[53] The Apalachicola River in the southeastern United States drains an area of about 45,000 km^2 above the gauge at Chattahoochee. As a result of the high density of precipitation measurements in this region, the anomaly errors are much smaller than the variance of annual anomalies. Interdecadal changes in precipitation correlate very strongly with discharge, though the two variables may diverge greatly in any given year. The flow estimation errors imply that these departures are real physical events, rather than artifacts of sampling error. The joint precipitation-discharge record suggests, by correlation, that stream flow during the first third of the century may have been lower than would be expected from the available stream flow record alone.

[54] The Niger River in western Africa drains an area of about 340,000 km^2 above the gauge at Dire in Mali. The peak flow during the 1950s and subsequent multidecadal drop in flow are confirmed by the precipitation record. Any backward extension of the stream flow record on the basis of the precipitation record is hindered by the smaller number of precipitation gauges (hence, larger anomaly error variances) prior to 1920, but would apparently support the occurrence of a high-flow regime back to the start of the century.

8. Summary and Discussion

8.1. Summary

[55] Progress in understanding and modeling of large-scale land water and energy balances might be accelerated by increased integration of observations into modeling studies. Such integration must be done with the recognition that observational errors in precipitation are typically large enough to mask model errors. Optimal use of observations therefore requires that precipitation errors be quantified accurately. Here we have developed, tested, and applied methods for the characterization of various types of basin mean precipitation errors. Our main focus has been on spatial sampling errors in the long-term mean and in temporal anomalies. With these methods, along with results of *Legates and Willmott* [1990] to characterize gauge bias, we have synthesized a set of long-term, monthly precipita-

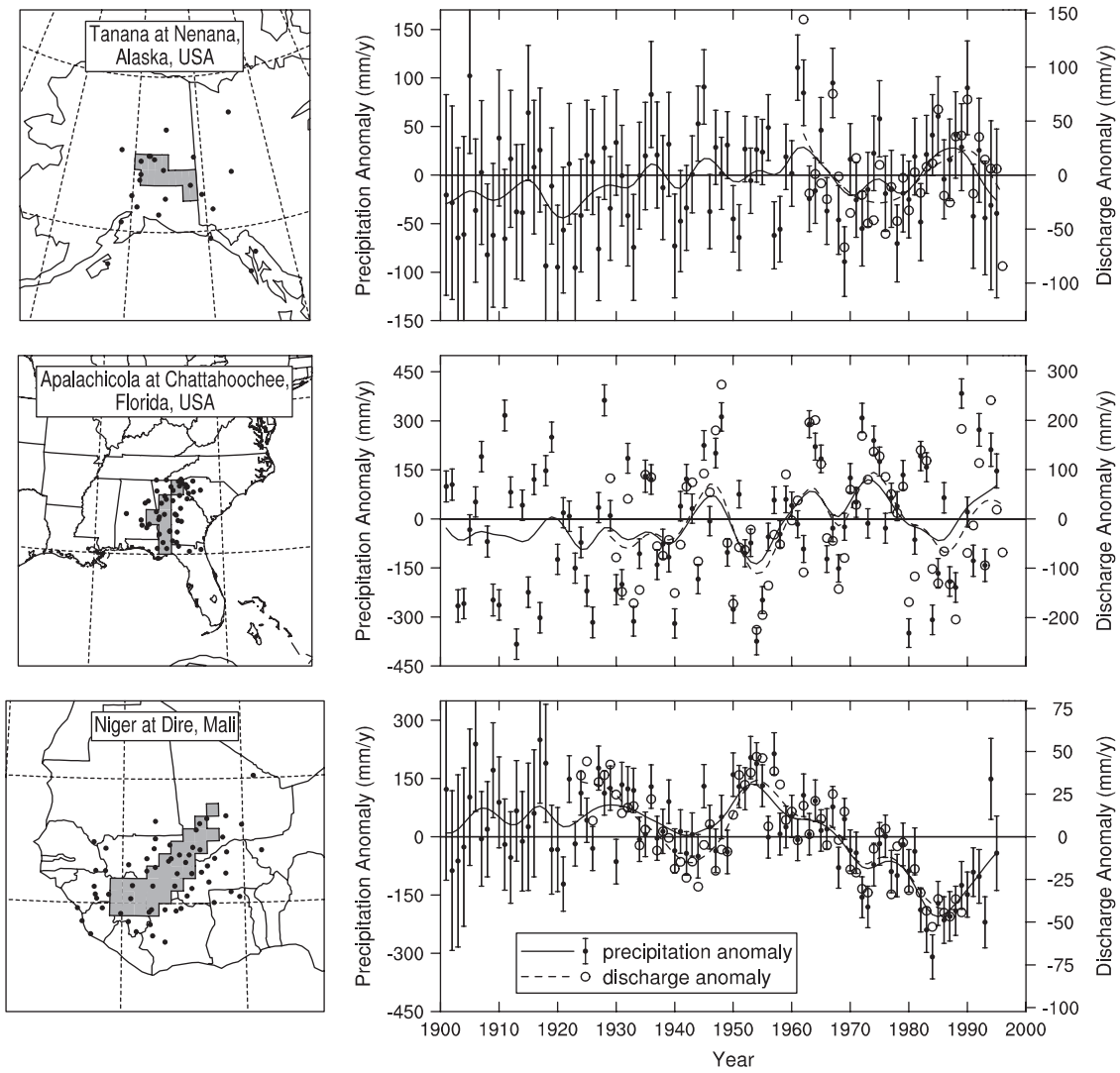

Figure 11. (left) Basin maps and (right) time series of annual precipitation and discharge anomalies for (top) Tanana River at Nenana, Alaska; (middle)Apalachicola River at Chattahoochee, Florida; and (bottom) Niger River at Dire, Mali. Symbols on maps show location of stations used in the analysis. Symbols on time series graphs represent water year anomalies (departures from means over period of discharge record), which were smoothed by a 21-point binomial filter to create the curves. The error bars on precipitation anomaly values correspond to $\pm 2\sigma_n$ uncertainty. Water years begin with November for the Tanana River basin, June for the Apalachicola River basin, and May for the Niger River basin. Scales for precipitation and discharge are defined so that one standard deviation corresponds to the same vertical displacement for both variables.

tion and discharge data for 175 large basins distributed across the major climatic and vegetation zones of the world. The authors may be contacted for information on access to the data set.

8.2. Precipitation Errors as a Factor in Model Testing

[56] The large magnitude of errors in data for many of the basins underlines the importance of error assessment in any application of observational data to diagnostic and modeling analyses of land water balances at large-basin scale. Figure 8 shows that a 10% to 20% bias in precipitation is typical among the basins we selected; had we not adjusted for gauge bias, these values would be even larger. When we consider that runoff from humid basins can be approximated as the difference between precipitation and evaporative potential, we see that a 10% to 20% error in precipitation can easily become a 100% error in runoff. Clearly, some selectivity is required in choosing basins for quantitative testing of land models. In part 2 of this series, we explore the consequences of using estimated precipitation errors in selecting observational data for application in theoretical and modeling analyses.

8.3. Precipitation Error, Model Calibration, and Energy Balance Modeling

[57] Historically, hydrologic modeling studies have had a central focus on fluxes of water mass, giving much less attention to fluxes of energy and the critical mass-energy linkage through evaporation. Because evaporation is rarely measured, and energetic processes are typically represented

by empirical loss functions or demand functions rather than observable energy fluxes, there is a tendency to use model calibration to avoid the reality of serious bias in precipitation estimates. Perhaps this is one factor explaining a general lack of attention by modelers to biases in precipitation estimates, despite the fact that they have been well documented [e.g., *Dawdy and Langbein*, 1960; *Larson and Peck*, 1974; *Legates and Willmott*, 1990]. With the advent of independent, quantitative, physically based estimates of energy forcing, it is possible that precipitation errors will be recognized more frequently. Perhaps recognition of the considerable undersampling of precipitation, especially in mountainous and cold environments, will lead to renewed attention to this problem by the hydrologic community.

8.4. Orographic Precipitation Sampling Error

[58] Orographic precipitation sampling errors appear to be the source of the largest errors in our data set. At the same time, such errors are probably the most poorly quantified by our analysis. The error estimates are based mainly on indirect techniques, and the comparison of inferred errors with those based on more detailed analyses in the United States pointed out significant differences in magnitude. Our index probably gives a reasonable guess as to the order of magnitude of the orographic error, but better methods are needed if quantitatively accurate analyses are to be performed for mountainous basins. Improvements may be achieved through a combination of increased gauging, more representative distribution of gauges, and more advanced treatments of the spatial estimation problem. The latter should include statistical and/or dynamic methods incorporating detailed information (including model-derived and remotely sensed data) on topographic structure, atmospheric motion, and atmospheric thermodynamics. Many such methods have been under development in recent years; their rigorous evaluation and increasing entrainment into global water balance analyses seems advisable.

[59] **Acknowledgments.** Stream flow data were supplied by the Global Runoff Data Centre, D-56002 Koblenz, Germany; by the U. S. Geological Survey; and by EarthInfo, Inc., and Environment Canada (for a few Canadian rivers). Several individuals supplied data for major rivers in South America: Jose Cordova (Orinoco), Jeff Richey (Amazon), Norberto Garcia (Parana), and Jaime Delgado and Pedro Restrepo (Magdalena and Cauca). Lauren Hay, Elena Shevliakova, and two anonymous reviewers gave us very helpful reviews of a draft of the manuscript.

References

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

Daly, C., R. P. Neilson, and D. L. Phillips, A statistical-topographic model for mapping climatological precipitation over mountainous terrain, *J. Appl. Meteorol.*, *33*, 140–158, 1994.

Darnell, W. L., W. F. Staylor, S. K. Gupta, and F. M. Denn, Estimation of surface insolation using Sun-synchronous satellite data, *J. Clim.*, *1*, 820–835, 1988.

Dawdy, D. R., and W. B. Langbein, Mapping mean areal precipitation, *Bull. Int. Assoc. Sci. Hydrol.*, *5*, 16–23, 1960.

Eagleson, P. S., The evolution of modern hydrology (from watershed to continent in 30 years), *Adv. Water Resour.*, *17*, 3–18, 1994.

EarthInfo, Inc., USGS Daily Values Central 1995 [CD-Rom], Boulder, Colo., 1995a.

EarthInfo, Inc., USGS Daily Values East 1995 [CD-Rom], Boulder, Colo., 1995b.

EarthInfo, Inc., USGS Daily Values West1 1995 [CD-Rom], Boulder, Colo., 1995c.

EarthInfo, Inc., USGS Daily Values West2 1995 [CD-Rom], Boulder, Colo., 1995d.

Garratt, J. R., A. J. Prata, L. D. Rotstayn, B. J. McAvaney, and S. Cusack, The surface radiation budget over oceans and continents, *J. Clim.*, *11*, 1951–1968, 1998.

Global Runoff Data Center, GRDC Catalogue Tool 2.2 for Windows 95/NT, computer program, Koblenz, Germany, 1998.

Gupta, S. K., W. L. Darnell, and A. C. Wilber, A parameterization of longwave surface radiation from satellite data: Recent improvements, *J. Appl. Meteorol.*, *31*, 1361–1367, 1992.

Larson, L. W., and E. L. Peck, Accuracy of precipitation measurements for hydrologic modeling, *Water Resour. Res.*, *10*, 857–863, 1974.

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

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

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

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

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

Milly, P. C. D., and R. T. Wetherald, Macroscale water fluxes, 3, Effects of land processes on variability of monthly river discharge, *Water Resour. Res.*, *38*, doi:10.1029/2001WR000761, in press, 2002.

Morrissey, M. L., J. A. Maliekal, J. S. Greene, and J. Wang, The uncertainty of simple spatial averages using rain gauge networks, *Water Resour. Res.*, *31*, 2011–2017, 1995.

Oki, T., and Y. Sud, Design of Total Runoff Integrating Pathways (TRIP): A global river channel network, *Earth Inter.*, *2*, 11, 1998.

Oki, T., T. Nishimura, and P. Dirmeyer, Assessment of annual runoff from land surface models using Total Runoff Integrating Pathways (TRIP), *J. Meteorol. Soc. Jpn.*, *77*, 235–255, 1999.

Rodríguez-Iturbe, I., and J. M. Mejía, The design of rainfall networks in time and space, *Water Resour. Res.*, *10*, 713–728, 1974.

Ruddy, B. C., and K. J. Hitt, Summary of selected characteristics of large reservoirs in the United States and Puerto Rico, 1988, *U.S. Geol. Surv. Open File Rep., 90-163*, 1990.

Shmakin, A. B., P. C. D. Milly, and K. A. Dunne, Global modeling of land water and energy balances, part III, Interannual variability, *J. Hydrometeorol.*, *3*(3), 311–321, 2002.

Slack, J. R., A. M. Lumb, and J. M. Landwehr, Hydro-Climatic Data Network (HCDN): streamflow data set, 1874–1988, *U.S. Geol. Surv. Water Resour. Invest. Rep. 93-4076*, 1993.

Times Books, *The Times Atlas of the World*, 7th comprehensive edition, London, 1988.

U.S. National Research Council, *Research Strategies for the U.S. Global Change Research Program*, Natl. Acad. Press, Washington, D.C., 1990.

U.S. National Research Council, *Opportunities in the Hydrologic Sciences*, Natl. Acad. Press, Washington, D.C., 1991.

U.S. National Research Council, *GCIP, Global Energy and Water Cycle Experiment (GEWEX) Continental-Scale International Project, A Review of Progress and Opportunities*, Natl. Acad. Press, Washington, D.C., 1998.

K. A. Dunne and P. C. D. Milly, Geophysical Fluid Dynamics Laboratory/NOAA, Princeton, NJ 08542, USA. (cmilly@usgs.gov)