Regional Regression Models of Annual Streamflow for the United States

By Richard M. Vogel,[1] Ian Wilson,[2] and Chris Daly[3]

Abstract: Estimates of annual streamflow volumes are needed in many different types of hydrologic studies. Usually a streamgauge is unavailable at the location of interest, hence regional methods that relate streamflow to readily measured geomorphic and climate characteristics provide a practical solution. Hydrologic, geomorphic, and climatic characteristics of 1,553 undeveloped watersheds across the United States are used to develop regional regression equations that relate the first two moments of annual streamflow to readily measured basin and climate characteristics. These relations are summarized for each of 18 major U.S. water resource regions. The relationships are remarkably precise, with adjusted R^2 values ranging from 90.2–99.8% and an average value of 96.2% across the continent. The usefulness of these relationships is evaluated by deriving their information content in terms of equivalent record length. These results indicate that regional models of annual streamflow, including runoff maps, are less accurate than suggested by traditional goodness-of-fit statistics. We also provide estimates of precipitation and temperature elasticity of streamflow, by region.

INTRODUCTION

Estimates of annual watershed runoff volumes are required in preliminary studies relating to water supply, irrigation, hydropower, navigation, recreation, and watershed management. Typically, the watershed of interest is ungauged and hence a flow record is unavailable. Thus a means of estimating the water yield or a streamflow trace from available data is needed.

Review: Models of Annual Streamflow for Individual Watersheds

Methods for estimation of annual watershed runoff that do not exploit regional information tend to involve some type of deterministic rainfall-runoff relationship. Such relationships tend to be either empirical or theoretical. For example, Duell (1994) and Revelle and Waggoner (1983) use multivariate regression procedures to develop empirical relationships between annual streamflow in a given year and temperature and precipitation during that year, for basins in the western U.S. Risbey and Entekhabi (1996) show that such multivariate (at-site) models can be misleading because they tend to show greater senstitivity to temperature than either observations or more physically detailed modeling approaches suggest.

Dooge (1992) suggests the use of the expectation of the continuity equation

$$\mu_Q = \mu_P - \mu_E \tag{1}$$

where μ_Q, μ_P, and μ_E represent the mean (long-term) runoff, precipitation, and evapotranspiration, respectively. Such an approach can be easily implemented if combined with a relationship between long-term actual and potential evapotranspiration such as those summarized by Dooge (1992). For example, Yates (1997) and others have combined (1) with the simple relationship introduced by Turc (1954) and Pike (1964):

$$\mu_E = \frac{\mu_P}{\sqrt{1 + \left(\frac{\mu_P}{\mu_{PE}}\right)^2}} \tag{2}$$

Eq. (2) relates long-term actual evapotranspiration μ_E to long-term potential evapotranspiration μ_{PE} and precipitation μ_P. Turc (1954) used data from river basins all over the world to validate (2). Kuhnel et al. (1991) summarize the historical development of relations of the form given in (2). Combining (1) and (2) leads to a model of long-term streamflow as a function of long-term potential evapotranspiration and precipitation.

Other types of annual water balance accounting models have been introduced by Fiering (1967) and Frind (1969). Eagleson (1978), Milly (1994a,b), and others have developed process-oriented annual runoff models with both theoretical and empirical underpinnings. Fiering (1967) and Frind (1969) introduced simplistic water balance models that enabled them to derive important statistical properties of the modeled streamflow.

Review: Regional Models of Annual Streamflow

Perhaps the simplest regional approach for estimating annual runoff is to transfer streamflow from a nearby catchment by assuming the runoff per square mile is constant. As is shown in this and many other studies (see Thomas and Benson 1970, for example) drainage area alone does not explain regional differences in annual streamflow. Additional information on land use, geomorphology, and climate is required.

Several regional approaches have attempted to incorporate streamflow information from many sites in the neighborhood of a particular watershed. Perhaps the most common regional approach to estimating watershed runoff is through the use of runoff maps, which report average annual runoff depth in inches or other units (Busby 1963; Gebert et al. 1987; Bishop and Church 1992; Arnell 1995) or runoff maps that report annual runoff as a fraction of normal runoff for particular water years (Busby 1963; Langbein and Slack 1979). Milly (1994a,b) also developed maps of average annual runoff using an analytic water balance model, which compared favorably with the maps produced by Gebert et al. (1987) and generated from actual runoff data.

Other regional approaches exploit spatial relationships among annual average streamflow, climate, land use, and geomorphology. Langbein (1949) introduced graphical relationships between average annual streamflow, temperature, and precipitation for the United States. Orsborn (1974) documents

[1]Prof., Dept. of Civ. and Envir. Engrg., Tufts Univ., Medford, MA 02155. E-mail: rvogel@tufts.edu

[2]Hydro., P.O. Box 2656, Taos, NM 87571.

[3]Forest Sci. Lab, Oregon State Univ., 3200 SW Jefferson Way, Corvallis, OR 97331-2209.

Note. Discussion open until November 1, 1999. To extend the closing date one month, a written request must be filed with the ASCE Manager of Journals. The manuscript for this paper was submitted for review and possible publication on August 6, 1998. This paper is part of the ***Journal of Irrigation and Drainage Engineering***, Vol. 125, No. 3, May/June, 1999. ©ASCE, ISSN 0733-9437/99/0003-0148–0157/$8.00 + $.50 per page. Paper No. 18974.

a graphical relation between the average annual flow, precipitation, and drainage area for stations near Vancouver, Washington. Investigators have more often used multivariate regression procedures to develop empirical equations relating mean annual streamflow to readily measured drainage basin characteristics. Such multivariate regional statistical relationships between climate, geomorphology, and streamflow have been developed by many investigators for the purpose of estimating floodflows and lowflow statistics at ungauged sites. For example, such regional relationships are so well developed for floodflows that a computer program is now available to implement them for all regions in the United States (Jennings et al. 1994). Multivariate regional regression models that relate the average annual streamflow to geomorphic and climate characteristics have been developed by Thomas and Benson (1970) for three selected regions in the western, central, and southern U.S., by Hawley and McCuen (1982) for the western U.S., by Vogel et al. (1997) for the northeastern U.S., by Majtenyi (1972) for areas of South Dakota, and by Lull and Sopper (1966) and Johnson (1970) for the New England region. The above list of studies in the United States is not exhaustive. Multivariate regression models have been developed by regional offices of the U.S. Geological Survey for other regions as well. Regional regression models of annual streamflow have received considerable attention from a global perspective, as evidenced by the books of Kalinin (1971), McMahon et al. (1992), and Finlayson and McMahon (1992).

Hawley and McCuen (1982) cite numerous advantages to the use of regional regression methods over runoff maps for estimating annual streamflow. Regional regression methods, unlike maps, produce objective equations that are easily programmed into comprehensive watershed planning procedures. Regional regression equations that relate annual average streamflow (or some other streamflow statistic) to geomorphic, land use, and climatic basin characteristics are easily integrated into and implemented using geographic information systems. Regression methods offer the opportunity to document the accuracy and uncertainty associated with water yield estimates, including estimation of confidence intervals and information content, as is shown here. Regression equations document the relationship between climate, geomorphology, and streamflow; hence they may be used to evaluate the impacts of climate change on water yield. Perhaps the most important advantage documented in this study is that regional models may be developed to quantify both the mean and variance of annual streamflow, thus providing the complete probability distribution of streamflow for any watershed in a region. Given the success of other studies by Hawley and McCuen (1982), Vogel et al. (1997), and others to develop regression equations for mean annual streamflow for selected regions in the United States, it is our goal to develop regional models for all regions of the country and to assess the accuracy of the resulting models on a nationwide basis. By developing regional regression models for all regions of the United States, we quantify the hydrology of the entire nation at an annual level.

Vogel and Wilson (1996) found that the gamma and lognormal distributions provide a good fit to the distribution of annual streamflow throughout the continental United States. This study is a sequel to that study, because we approximate the first two moments of the distribution of annual streamflow for any watershed greater than about 2 km^2 for all regions of the U.S. Another goal is to assess the precision of the resulting regression equations by deriving the information content associated with each equation in terms of the number of equivalent years of streamflow record that would be required to duplicate the regression equations. Although goodness-of-fit statistics suggest that regression equations for annual streamflow are remarkably accurate, particularly in the more humid regions of the country, this study documents that for many western regions of the United States the equations are only equivalent to a few years worth of streamflow data. The regional regression equations used here are based on more than drainage area alone, as is the case for annual runoff maps; hence, the regression equations introduced here can provide more accurate estimates of annual streamflow volumes at ungauged sites than can runoff maps. Runoff maps provide comparable accuracy to a regional regression when variables other than drainage area are not included in the regression analysis.

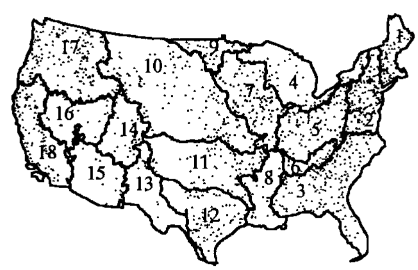
REGIONAL DATABASES

Our approach is empirical, and our ability to relate annual streamflow to climate and geomorphology results from extensive use of regional data for the United States, so we begin by describing those information resources. All annual streamflow and climate records are based on calendar years.

Streamflow Database

The streamflow dataset consists of records of average daily streamflow at 1,553 sites located throughout the continental United States (see Fig. 1). This dataset, termed the hydroclimatologic data network (HCDN), was developed by the U.S. Geological Survey (Slack et al. 1993) for the purpose of studying surface-water conditions throughout the land. The dataset is available on CD-ROM from the U.S. Geological Survey and is accessible on the World Wide Web at the following URL: *http://www.rvares.er.usgs.gov/hcdn_cdrom/1st_page.html*.

The development of the HCDN was a large undertaking that included screening the data in a variety of ways. Data specialists at the U.S. Geological Survey (USGS) district offices, as well as data specialists at the national headquarters, reviewed records on the basis of the following criteria: (1) Availability of data in electronic form; (2) record lengths in excess of 20 years unless site location is underrepresented; (3) ac-



Region	Region Name	Region	Region Name
1	New England	10	Missouri
2	Mid-Atlantic	11	Arkansas-hite-Red
3	South Atlantic Gulf	12	Texas-Gulf
4	Great Lakes	13	Rio Grande
5	Ohio	14	Upper Colorado
6	Tennessee	15	Lower Colorado
7	Upper Mississippi	16	Great Basin
8	Lower Mississippi	17	Pacific Northwest
9	Souris-Red-Rainy	18	California

FIG. 1. Location of HCDN Stations and Water Resources Regions in United States

curacy ratings of records are at least ''good'' as defined by USGS standards; (4) no overt adjustment of ''natural'' monthly streamflows by flow diversion, augmentation, groundwater pumping, or other forms of regulation; and (5) only measured discharge values are tabulated, reconstructed, or estimated records are not used. Sites included contained periods of record that ranged from 6 to 115 years, with an average record length of 45.5 years per site.

With all of the efforts at developing a database free of anthropogenic influences, the streamflows will still exhibit trends. Lettenmaier et al. (1994) use Mann-Kendall tests to document that significant trends in both annual and monthly streamflows are apparent for broad regions of the United States. They document the fact that annual streamflow in the Northwest has tended to decrease while annual streamflow in the Midwest has tended to increase. This study treats annual streamflows as if they arise from a stationary process, even though Lettenmaier et al. (1994) and Lins (1985) observe that annual streamflows are probably not stationary time series.

Geomorphic Characteristics

Table 1 lists the geomorphic characteristics used in this study. Drainage areas are obtained from the HCDN. Each basin was also outlined using the standard algorithms available in the geographic information system Arcview 3.1 (''Arcview'' 1998) along with a 1 km digital grid elevation map (DEM) for the continental U.S. Average spatial estimates of the geomorphic characteristics (other than drainage area) listed in Table 1 are obtained using standard algorithms within the Arcview 3.1 Spatial Analyst (Arcview 1998).

TABLE 1. List of Climate and Geomorphic Basin Characteristics

Variable Name	Definition	Units	Source
Geomorphic Variables			
Area	Drainage area	Square kilometers	HCDN
Perim	Perimeter of basin	meters	Arcview
Slope	Average basin slope	degrees	Arcview
A/Perim	Drainage area divided by basin perimeter	meters	Arcview
Elev	Mean basin elevation	meters	Arcview
Climatic Variables			
μ_T	Mean annual temperature	Fahrenheit*10	PRISM US_TAVG.14
μ_P	Mean annual precipitation	mm/yr.	PRISM US_PPT.14
σ_P	Standard deviation of annual precitation	mm/yr.	PRISM US_PPT_STD.14
Solar	Solar radiation	mm/yr.	
Annevap	Annual pan evaporation	inches	NOAA
Mayevap	Pan evaporation - May through October only	inches	NOAA
Snow	Annual Snowfall	cm	PRISM US_SNOW.14
%Snow	Percent of precipitation as snow	in tenths of a percent	PRISM US_PPS.14
Tmax	Annual mean maximum temperature	Fahrenheit*10	PRISM US_TMAX.14
T-Aug and T-Jan	Mean August temperature Mean January temperature	Fahrenheit*10	PRISM US_TAVG.14
T-Jan-min	Mean minimum January temperature	Fahrenheit*10	PRISM US_TMIN.01
HeatD	Heating degree days, base = 65°F	Days	PRISM US_HDD.14
CoolD	Cooling degree days, base = 65°F	Days	PRISM US_CDD.14
P-Jan through P-Dec	Mean monthly precipitation for January through December	mm/month	PRISM US_PPT.01 through US_PPT.12

Climate Characteristics

Slack et al. (1993) include a file containing some climate and watershed characteristics for each of the basins; however, that computer file is now several decades old and the only climate variables it includes are average minimum January temperature, average annual precipitation, and intensity of 24-h, 2-year storm. Since climate plays such an important role in the hydrologic cycle, a special effort was made to obtain estimates of climate characteristics that are as accurate as possible using the current state of the art. Table 1 summarizes the climate characteristics computed for this study. Spatially weighted estimates of each climate characteristic in Table 1 are derived using a digital elevation map (DEM), a digital grid for each climate characteristic, and a geographic information system (GIS). The DEM is used to outline the boundaries of all 1,553 drainage basins. The climate characteristics are estimated for the 1,553 watersheds from 2.5-min digital grids using the PRISM (Daly et al. 1994, 1997) climate analysis system. Table 1 reports the file names for each PRISM grid used in our analyses. We used standard algorithms available in the spatial analyst of Arcview 3.1 to estimate each climate characteristic over each basin using the PRISM digital climate maps. The climate datasets used here are based on the period 1961–90 with a 2.5-min (~4 km) resolution grid encompassing the lower 48 states. The observational data consisted of approximately 7,000 National Weather Service and cooperator precipitation stations, 500 SNOTEL stations (USDA Soil Conservation Service 1988), and 2,000 additional data points from state and local climate networks. The precipitation dataset was generated using the PRISM modeling system (Daly et al. 1994, 1997). PRISM (parameter-elevation regressions on independent slopes model) is a climate analysis system that uses point data, a digital elevation model (DEM), and other spatial information to generate gridded estimates of annual, monthly, and event-based climatic parameters. It has been designed to accommodate difficult climate mapping situations in innovative ways. These include vertical extrapolation of climate well beyond the lowest or highest observation, reproducing gradients caused by rain shadows and coastal effects, and assessing the varying effects of terrain barriers on precipitation. Originally developed in 1991 for precipitation estimation, PRISM has been generalized and applied successfully to temperature, snowfall, growing degree-days, and weather generator parameters, among others (Taylor et al. 1997). The precipitation dataset and other digital PRISM products are available on the World Wide Web at *http://www.ocs.orst.edu/prism/prism_new.html*.

Summary of Annual Streamflow Characteristics

Using the same HCDN database of streamflow used here (Slack et al. 1993), Vogel and Wilson (1996) employed L-moment diagrams to document that annual streamflow is well approximated by either a Gamma or a lognormal distribution throughout the continental United States. Markovic (1965) also documented that the Gamma and lognormal distributions provide good approximations for modeling annual streamflows based on chi-square goodness-of-fit evaluations at 446 basins in the western U.S. Since a two-parameter probability density function is adequate to define the likelihood of annual streamflow, only the mean μ and standard deviation σ of the annual flows are used here to summarize their pdf. Estimates of μ and σ^2 are obtained using the standard sample estimators $\bar{q} = \sum_{i=1}^{n} q_i/n$ and $s_q^2 = \sum_{i=1}^{n} (q_i - \bar{q})^2/(n - 1)$, respectively.

MULTIVARIATE REGIONAL REGRESSION ANALYSIS

This section summarizes multivariate relationships among the streamflow statistics $\bar{q}$ and s_q^2, and the basin characteristics, including drainage area A, mean basin precipitation μ_P, mean basin temperature μ_T, and the other basin characteristics listed in Table 1. Initially, we attempted to fit a single model for the entire continental U.S.; as expected such models performed poorly. To reduce prediction errors, models for smaller regions are developed.

In 1970, the (now defunct) U.S. Water Resources Council defined 18 major regions in the continental U.S., shown in Fig. 1, for the purpose of assessing the state of water resources across the nation. Lins (1997) argues that these 18 regions, which are based on the location of natural drainage divides, do not necessarily reflect regions of uniform within-year streamflow variability. However, Vogel et al. (1998) document that the persistence, and to a lesser extent the year-to-year variability, of streamflow is relatively homogeneous within each of the 18 regions illustrated in Fig. 1.

We used ordinary least squares (OLS) and weighted least squares (WLS) regression procedures to obtain an equation for each of the dependent variables $\theta = \bar{q}$ (m^3/s) and $\theta = s_q^2$ (in m^6/s^2), where

$$\theta = e^{\beta_0} \cdot X_1^{\beta_1} \cdot X_2^{\beta_2} \cdots X_m^{\beta_m} \cdot \nu \tag{3}$$

for each of the 18 regions. Here X_i, $i = 1, \ldots m$ = basin characteristics; β_i, $i = 1, \ldots m$ = model parameters; and ν = lognormally distributed model errors. The model in (3) is termed a log-linear model because taking natural logarithms yields

$$\ln(\theta) = \beta_0 + \beta_1 \cdot \ln(X_1) + \beta_2 \cdot \ln(X_2) + \ldots + \beta_m \ln(X_m) + \varepsilon \tag{4}$$

where now the residuals $\varepsilon = \ln(\nu)$ are normally distributed with zero mean and variance σ_ε^2.

Our goals in model development are to, first, maximize the adjusted R^2, second, minimize the prediction error measured by σ_ε^2, and third, include both geomorphic and climate characteristics while keeping the number of basin characteristics used to a minimum. We used variance inflation factors to assure that none of the models exhibits multicollinearity. The influence statistic Cooks D (Cook and Weisberg 1982) was employed to identify and remove sites that are known to exhibit an unrealistic amount of influence on estimates of model coefficients. We used the probability plot correlation-coefficient hypothesis test to assure that the model residuals ε are approximately normally distributed. The model prediction errors are quantified in terms of the standard deviation of each estimated dependent variable θ. If our interest were in $\ln(y)$ we could then report the prediction error using an estimate of the standard deviation of ε; however, our interest is in the dependent variable θ, not $\ln(\theta)$, and hence it is necessary to estimate the standard deviation of θ using the fact that the errors ν in (1) are lognormally distributed. Since the errors ν in (1) do not have a mean of zero, it is customary to report their coefficient of variation as the average model prediction error, where $C_v(\nu) = SE = \sqrt{\exp(\sigma_\varepsilon^2) - 1}$, which is simply the relationship between the variance (in log space) of a lognormal variable and its coefficient of variation (in real space). The total variance of a model prediction σ_ε^2 is made up of three components—model error, sampling error, and measurement error. Measurement errors are ignored here. Stedinger and Tasker (1985) describe procedures for estimating the model and sampling error component of the total prediction variance. For each regional regression equation, the average model error is computed as the variance of the residuals in the estimated model s_e^2 and the average sampling error is computed as ps_e^2/m, where p is the number of model parameter estimates and m is the number of sites in the region. Therefore the average standard error of a prediction is estimated as

$$SE = \sqrt{\exp[s_e^2(1 + p/m)] - 1} \tag{5}$$

where s_e = standard error of the residuals resulting from each regression.

When a regression equation is used to estimate the mean or variance of annual streamflow at an ungauged site, the actual prediction error will also depend on how far removed the independent variables are from the mean values used in developing the regression. In other words, (5) only approximates the average prediction error. Hardison (1971) and Stedinger and Tasker (1986) document expressions that may be used to estimate the variance of a prediction at an ungauged site. OLS regression procedures ignore the cross-correlation among the annual flow traces in a region as well as the variation in the record lengths in a region. Stedinger and Tasker (1986) document how generalized least squares (GLS) procedures can be used to account for both of these additional factors.

With OLS regression procedures, each basin is treated equally, implying that all observations of the dependent variable are "equally reliable." Tasker (1980) introduced WLS regression to account for the fact that each basin has a different streamflow record length leading to estimates of the dependent variable with varying degrees of reliability. Using WLS regression, Tasker (1980) shows that the weight w_j assigned to each set of observations of the dependent variable and its associated independent variables is proportional to the reciprocal of the variance of an estimate of the dependent variable. Ignoring model error, the reciprocal of the variance of the dependent variable $\bar{q}$ (a sample estimate of the dependent variable μ_q) is given by $\mathrm{var}^{-1}(\bar{q}) = n/\sigma^2$. Since the reciprocal of the variance of the dependent variable is proportional to the record length, we use the simple weighting scheme

$$w_i = \frac{n_i}{\sum_{j=1}^{m} n_j} \tag{6}$$

where w_i = weight for site i; n_i = length of streamflow record, in years, for site i; and m = number of sites in region. Eq. (6) places a weight on each dependent variable in proportion to the record length used to estimate those variables. A similar approach is employed for the dependent variable s_q^2.

Results

One goal is to develop regional regressions for use in climate change investigations. Hence, our initial models employ the independent variables drainage area A, mean annual temperature μ_T, and mean annual precipitation μ_P. These regional regression models for the mean and variance of annual streamflow are summarized in Tables 2 and 3, respectively. The tables report the model coefficients along with their t-ratios (in parentheses), the maximum variance inflation factor (VIF) for each model, the average prediction error (SE), and the adjusted R^2. Note that in each Table, 2 and 3, two sets of regression equations are reported for each region. The first regression equation corresponds the regression with drainage area alone, and the second regression includes the explanatory variables precipitation and temperature. The adjusted R^2 values for the regressions for the mean annual streamflow based on drainage area alone vary from 27.3 to 99.1%, with an average value of only 71.4%. When mean annual temperature and precipitation are added to these models, the R^2 values range is 85.2–99.7% with a mean of 94.5%. Clearly climate information is required to obtain an accurate estimate of either the mean or variance of annual streamflow.

TABLE 2. Regression Models for Mean Annual Streamflow $\hat{\mu}_Q = e^a A^b \mu_P{}^c \mu_T{}^d$

Region (1)	a (2)	b (3)	c (4)	d (5)	Maximum VIF (6)	Standard error (7)	Adjusted R^2 (8)	Average record length (9)
1	−3.9207 (−61.7) −9.4301 (−9.5)	1.00537 (94.9) 1.01238 (138.7)	 1.21308 (12.5)	 −0.5118 (−4.0)	 1.6	0.15 0.09	99.1% 99.7%	45.7
2	−3.8891 (−45.4) −2.7070 (−2.6)	0.95184 (68.0) 0.97938 (145.2)	 1.6251 (15.6)	 −2.0510 (−19.0)	 1.2	0.28 0.12	96.8% 99.4%	46.7
3	−4.0556 (−36.5) −10.1020 (−8.4)	0.96561 (62.7) 0.98445 (119.5)	 2.2599 (22.5)	 −1.6070 (−9.5)	 1.1	0.33 0.17	95.7% 98.9%	41.0
4	−3.4073 (−11.6) −5.6780 (−3.3)	0.84183 (19.9) 0.96519 (41.6)	 2.2889 (12.1)	 −2.3191 (−8.4)	 1.3	0.32 0.16	87.9% 97.1%	48.2
5	−4.1073 (−28.7) −4.8910 (−4.0)	0.97165 (48.3) 0.99319 (134.8)	 2.32521 (24.7)	 −2.5093 (−13.8)	 1.0	0.32 0.12	96.1% 99.4%	51.9
6	−2.9702 (−16.8) −8.8200 (−3.6)	0.86864 (30.9) 0.96418 (57.7)	 1.3581 (11.9)	 −0.7476 (−2.0)	 1.5	0.27 0.14	95.3% 98.7%	48.5
7	−4.7378 (−21.8) −11.8610 (−14.1)	0.9588 (35.2) 1.00209 (121.6)	 4.5596 (33.7)	 −3.8984 (−18.4)	 2.5	0.45 0.14	88.4% 98.8%	51.4
8	−4.2883 (−14.4)	1.01724 (25.1) 0.98399 (32.3)	 3.1570 (4.3)	 −4.1898 (−5.1)	 2.3	0.22 0.15	96.7% 98.5%	42.8
9	−6.4370 (−4.8)	0.9432 (5.3) 0.81629 (17.6)	 6.4222 (21.0)	 −7.6551 (−24.2)	 1.1	1.80 0.36	48.4% 95.7%	36.3
10	−2.8316 (−5.9) −10.9270 (−6.8)	0.59199 (9.5) 0.89405 (27.7)	 3.2000 (22.8)	 −2.4524 (−9.3)	 1.3	1.61 0.58	40.2% 86.4%	40.7
11	−3.6498 (−4.3) −18.6270 (−13.7)	0.7657 (6.6) 0.96494 (36.4)	 3.8152 (37.6)	 −1.9665 (−7.2)	 1.7	1.96 0.30	27.3% 96.2%	39.9
12	−3.1838 (−5.4)	0.59662 (8.0) 0.84712 (21.6)	 3.8336 (16.9)	 −4.7145 (−18.5)	 1.4	1.53 0.56	41.3% 86.9%	38.0
13	−3.0703 (−3.8)	0.5129 (4.0) 0.77247 (10.9)	 1.9636 (6.9)	 −2.8284 (−8.5)	 2.7	1.39 0.51	38.9% 87.4%	39.5
14	−3.2483 (−9.5) −9.8560 (−2.6)	0.76625 (15.5) 0.98744 (24.6)	 2.4690 (7.8)	 −1.8771 (−3.9)	 1.5	0.63 0.42	84.9% 93.5%	42.0
15	−4.5617 (−5.0)	0.6964 (6.2) 0.8663 (8.1)	 2.5065 (2.9)	 −3.4270 (−3.7)	 1.4	1.21 0.81	72.0% 85.2%	42.6
16	−2.7240 (−4.9)	0.57681 (6.3) 0.83708 (16.5)	 2.1672 (9.9)	 −3.0535 (−11.5)	 1.6	1.23 0.50	56.0% 89.6%	43.0
17	−2.4177 (−6.1) −10.1800 (−7.3)	0.75318 (13.4) 1.00269 (53.6)	 1.86412 (39.7)	 −1.1579 (−4.9)	 1.2	1.54 0.35	52.9% 95.6%	42.2
18	−5.2420 (−12.8) −8.4380 (−3.6)	1.13814 (16.5) 0.97398 (33.7)	 1.99863 (22.5)	 −1.5319 (−4.6)	 1.2	1.59 0.48	67.5% 94.7%	41.4

A final attempt was made to obtain the best possible multivariate regression for the mean annual streamflow for each of the 18 regions, using the independent variables in Table 1. For this analysis, we employed both forward and backward stepwise regression procedures to select the independent variables from Table 1 having the greatest explanatory power. Table 4 summarizes the final multivariate models for estimating the mean annual streamflow μ_Q. In all cases, the models listed in Table 4 are an improvement over the models in Table 2, which are based on the independent variables A, μ_P, and μ_T; however, the improvement is often only marginal.

Though the goodness-of-fit of the regression equations reported in Tables 2–4 is remarkably good overall, improvement is always possible. In an effort to identify the type of information required to improve these relations, Fig. 2 illustrates the standard error for each model computed from (5) versus the dryness ratio, defined as actual annual evapotranspiration (ET) divided by annual precipitation. Actual ET was computed as the difference between precipitation and runoff. Annual precipitation for each region was obtained from the PRISM grid described in Table 1. Long-term runoff is obtained for each of the watersheds within the 18 regions using the regression equations based on A, μ_P, and μ_T in Table 2. Fig. 2 documents quite clearly that regional models with higher standard errors correspond to regions with high dryness ratios. Large dryness ratios correspond to the arid and semiarid regions. Therefore, improvements in the models are likely to come from improvements in our ability to describe watershed aridity, because Fig. 2 documents that it is differences in watershed aridity that seem to explain the differences in goodness-of-fit of the regional models.

Climate Elasticity of Streamflow

For the log-linear models developed here, it can be shown that the model coefficients reported in Tables 2–4 represent streamflow elasticities. For the regression models for the mean annual streamflow, $\mu_Q = aA^b\mu_P^c\mu_T^d$. Defining precipitation elasticity as $\varepsilon_P = [d\mu_Q/d\mu_P](\mu_P/\mu_Q)$, one can show that for this simple model $\varepsilon_P = c$. Similarly, the coefficients b and d represent the elasticity of streamflow to drainage area and temperature, respectively. For example, the coefficients for the scale term (drainage area) in Table 2 are always very nearly unity, showing that a 1% increase in scale leads to an approximately 1% increase in streamflow. Similarly, the precipitation elasticities are all positive and often quite large, indicating the highly nonlinear response of streamflow to precipitation. As expected, all coefficients for the temperature terms are negative, indicating that increases in temperature tend to increase evapotranspiration, leading to decreases in streamflow.

TABLE 3. Regression Models for Variance of Annual Streamflow $\hat{\sigma}_Q^2 = e^a A^b \mu_P^c \mu_T^d$

Region (1)	a (2)	b (3)	c (4)	d (5)	Maximum VIF (6)	Standard error *SE* (7)	Adjusted R^2 (8)
1	−10.0772 (−70.4)	1.9175 (80.4)				0.33	98.9%
	−26.3900 (−7.4)	1.97748 (75.9)	1.4482 (4.2)	0.9450 (2.1)	1.6	0.29	99.2%
2	−10.1572 (−96.2)	1.87622 (108.9)				0.36	98.6%
		1.86987 (106.2)	0.4181 (2.2)	−2.1064 (−9.8)	1.2	0.28	99.2%
3	−10.7839 (−58.0)	2.01066 (78.0)				0.62	96.8%
	−41.9100 (−16.5)	1.98535 (114.7)	3.3542 (15.9)	1.1217 (3.2)	1.1	0.40	98.6%
4	−10.5840 (−17.5)	1.83559 (21.0)				0.73	88.6%
		1.91153 (19.5)	1.8639 (2.6)	−3.9008 (−4.5)	1.3	0.51	93.9%
5	−10.9644 (−51.3)	1.99251 (66.5)				0.52	97.7%
	−44.7100 (−15.1)	1.98998 (110.4)	2.7482 (11.9)	2.3381 (5.2)	1.0	0.29	99.2%
6	−9.3426 (−25.1)	1.83746 (31.2)				0.59	95.5%
		2.02185 (34.6)	2.0462 (5.2)	−4.0546 (−8.3)	1.5	0.34	98.4%
7	−10.9258 (−29.7)	1.89786 (41.3)				0.84	91.6%
	−44.4000 (−14.8)	1.94925 (66.2)	3.3290 (6.9)	1.7264 (2.3)	2.5	0.50	96.5%
8	−9.8317 (−27.0)	1.9747 (39.7)				0.26	98.7%
		1.96504 (32.9)	−1.33998 (−22.3)	—	1.0	0.24	98.9%
9	−13.1340 (−6.7)	1.8238 (6.9)				4.45	62.9%
		1.6545 (14.3.)	8.8103 (11.6)	−11.2298 (−14.2)	1.1	0.97	92.2%
10	−9.0091 (−10.7)	1.4427 (13.2)				7.27	56.4%
		1.82874 (18.4)	3.8524 (9.7)	−6.0639 (−12.5)	1.3	1.70	85.3%
11	−8.9230 (−6.5)	1.6120 (8.5)				9.29	38.9%
	−37.3990 (−12.1)	1.89742 (31.4)	5.9209 (25.6)	−2.1974 (−3.6)	1.7	0.82	93.2%
12	−7.0913 (−7.9)	1.2314 (10.8)				4.14	55.4%
		1.6033 (25.2)	5.7729 (15.7)	−7.4473 (−18.0)	1.4	1.01	89.3%
13	−7.7400 (−5.8)	1.1021 (5.3)				3.85	53.8%
		1.5137 (13.1)	2.9978 (6.5)	−4.8379 (−9.0)	2.7	1.13	86.7%
14	−8.7850 (−17.6)	1.54046 (21.2)				0.98	91.9%
		1.78314 (21.7)	1.9821 (4.3)	−3.9825 (−7.0)	1.5	0.72	95.3%
15	−9.9420 (−5.2)	1.3897 (5.9)				10.16	66.5%
		1.7282 (7.4)	5.0770 (2.7)	−7.0460 (−3.5)	1.4	2.52	86.3%
16	−8.0534 (−9.9)	1.3481 (10.2)				2.64	74.3%
		1.72656 (23.6)	3.1127 (9.8)	−5.0781 (−13.2)	1.6	0.70	95.2%
17	−7.9331 (−13.0)	1.57582 (18.2)				4.15	66.1%
		1.91574 (48.6)	2.86024 (28.8)	−5.0060 (−36.5)	1.2	0.63	96.1%
18	−9.6617 (−17.9)	1.95287 (21.5)				2.80	78.1%
		1.7064 (29.6)	2.0952 (12.8)	−3.6283 (−20.9)	1.2	1.04	92.7%

Comparisons of Models at National Scale

Fig. 3 compares estimates of the mean annual streamflow obtained from each of four different regional models with observed streamflows. Figs. 3(a–c) compare estimated and observed flows corresponding to the regressions developed here with the following predictor variables: (1) A only; (2) A, μ_P, and μ_T; and (3) all variables. Similarly, Fig. 3(d) compares observed and estimated mean flows, using equations (1) and (2), which were developed by others. Here, annual average potential evapotranspiration μ_{PE} in (1) is estimated using an equation developed by Turc (1954) which is $\mu_{PE} = 300 + 25\mu_T + 0.05\mu_T^3$. Kuhnel et al. (1991) cite numerous studies that validated (1) and (2) for over a thousand catchments worldwide. Our regional regression equations appear to be a significant improvement over such empirical relationships in the U.S. Another significant advantage of regional regression equations is that one can estimate the standard error of estimate that allows us to quantify the information content of each regression, as is done in the following section.

EFFECTIVE RECORD LENGTH FOR REGIONAL MODELS

In this section we derive expressions that approximate the information content of each of the regional regressions. Information content is defined as the effective record length that would be required at a gauged site to estimate either $\bar{q}$ or s_q^2 with the same accuracy as the regional regression for that region. For this purpose, we assume that annual streamflows arise from an AR (1) lognormal process (see Vogel and Wilson 1996; Vogel et al. 1998). Assuming that a sample of n annual streamflow observations, q_t, $t = 1, \ldots, n$, arises from an AR (1) lognormal model, then the mean and variance of those samples are described by

$$\mu_q = \exp(\mu_y + \sigma_y^2/2) \tag{7}$$

and

$$\sigma_q^2 = \exp(2\mu_y + \sigma_y^2)[\exp(\sigma_y^2) - 1] \tag{8}$$

where $y = \ln(q)$; μ_q and σ_q^2 = mean and variance of flows; and μ_y and σ_y^2 = mean and variance of natural logarithms of flows. If q is AR (1) lognormal, then $y = \ln(q)$ is AR (1) normal, so that

$$y_t = \mu_y + \rho_y(y_{t-1} - \mu_y) + \upsilon_t\sigma_y\sqrt{1 - \rho_y^2} \tag{9}$$

where υ_t are normally distributed with zero mean and unit variance, and ρ_y is the lag-one serial correlation of the logarithms. The serial correlation in log space ρ_y is related to the serial correlation in real space ρ_q via

TABLE 4. Regression Models for Mean Annual Streamflow Using All Basin Characteristics $\mu_Q = e^a$(1st Predictor)^b(2nd Predictor)^c(3rd Predictor)^d(4th Predictor)^e(5th Predictor)^f

Region (1)	a (2)	b (3)	c (4)	d (5)	e (6)	f (7)	Maximum VIF (8)	Standard error *SE* (9)	Adjusted R^2 (10)
1	—	Area 1.0129 (182.0)	μ_P 1.0177 (6.0)	Solar −1.51261 (−15.7)	*P*-November 0.2937 (2.4)	—	4.2	0.08	99.8%
2	5.0515 (5.2)	Area 0.980224 (150.0)	*T*max −2.7114 (−21.3)	*P*-July 0.8361 (6.4)	*P*-May 0.9352 (6.9)	—	1.9	0.12	99.4%
3	−17.7550 (−15.3)	Area 0.986557 (119.3)	μ_P 2.0653 (20.1)	Annual evaporation −0.6620 (−3.1)	Snow 0.04297 (3.2)	*P*-June 0.2264 (2.2)	6.7	0.16	99.0%
4	−15.0090 (−12.3)	Area 0.96809 (42.9)	μ_P 1.8888 (11.0)	Cool *D* −0.35926 (−8.9)	—	—	1.2	0.15	97.3%
5	3.9260 (2.5)	Area 0.997939 (152.8)	μ_P 1.0030 (4.5)	*P*-January 0.7647 (6.9)	*T*max −2.8710 (−15.5)	—	7.4	0.11	99.5%
6	−11.359 (−9.5)	Area 0.96266 (61.1)	μ_P 1.3746 (12.2)	*T*-January-minimum −0.4196 (−2.4)	—	—	1.4	0.14	98.8%
7	−26.5585 (−46.0)	Area 1.01256 (132.4)	μ_P 3.80518 (38.9)	Cool *D* −0.65629 (−20.5)	—	—	1.5	0.13	98.9%
8	−20.9350 (−6.8)	Area 1.03835 (50.6)	μ_P 2.2359 (3.9)	σ_P −1.5733 (−2.9)	*P*-May 1.7960 (3.6)	% Snow −0.05195 (−2.8)	4.2	0.09	99.5%
9	—	Area 0.84672 (19.5)	Annual evaporation −5.5864 (−28.3)	*P*-September 3.3743 (19.8)	—	—	2.6	0.31	96.8%
10	−13.4592 (−25.1)	Area 0.92511 (35.9)	*P*-November 0.7742 (6.1)	Slope 0.50816 (7.5)	*P*-September 1.3205 (6.8)	—	6.6	0.45	91.7%
11	−38.4060 (−25.3)	Area 0.94032 (38.2)	Heat *D* 0.8481 (7.8)	μ_P 3.8781 (38.1)	—	—	1.8	0.29	96.4%
12	—	Area 0.92543 (25.8)	*P*-November 1.5788 (7.6)	*P*-October 1.6319 (3.4)	*T*-August −2.8288 (−12.0)	—	5.4	0.48	90.2%
13	−16.8060 (−14.9)	A/Perimeter 1.4481 (13.7)	*P*-March 1.2833 (10.6)	—	—	—	1.1	0.44	90.5%
14	−15.2070 (−12.6)	Area 0.94765 (26.4)	Snow 1.6706 (10.0)	—	—	—	1.3	0.38	94.5%
15	−39.1670 (−6.9)	Area 0.78736 (12.9)	*P*-October 6.9176 (9.2)	*P*-August −4.0588 (−7.1)	*T*-January 4.4457 (4.7)	—	2.3	0.55	96.8%
16	−32.1610 (−6.9)	μ_P 2.2588 (8.7)	Area 0.8861 (16.2)	Elevation 1.6641 (2.8)	—	—	1.6	0.49	90.2%
17	−16.1816 (−43.8)	Area 0.99076 (54.9)	μ_P 1.65398 (34.1)	Slope 0.33273 (6.2)	—	—	1.4	0.33	96.0%
18	—	Area 0.93609 (35.3)	May evaporation −1.4303 (−4.9)	*P*-January 1.5008 (13.5)	*P*-June 0.17065 (3.4)	μ_T −1.1532 (−5.5)	3.0	0.43	95.7%

$$\rho_y = \frac{\ln(1 + \rho_q C_v^2)}{\ln(1 + C_v^2)} \tag{10}$$

where $C_v = \sigma_q/\mu_q$. Appendix I uses a Taylor-series approximation to (7) and (8) to derive an approximation to the variance of an at-site maximum likelihood estimate (MLE) of μ_q and σ_q^2, respectively, leading to

$$\mathrm{Var}[\hat{\mu}_q] \cong \frac{\mu_q^2 \ln(1 + C_v^2)}{n}\left[\frac{1 + \rho_y}{1 - \rho_y} + \frac{\ln(1 + C_v^2)}{2}\left[\frac{1 + \rho_y^2}{1 - \rho_y^2}\right]\right] \tag{11}$$

and

$$\mathrm{Var}[\hat{\sigma}_q^2] = \frac{2\mu_q^4 \ln(1 + C_v^2)}{n} \left[2C_v^2 \left[\frac{1 + \rho_y}{1 - \rho_y} \right] + [1 + 2C_v^2]^2 \ln(1 + C_v^2) \left[\frac{1 + \rho_y^2}{1 - \rho_y^2} \right] \right] \tag{12}$$

In general, the variance of an at-site sample estimate of the mean or variance increases as C_v and ρ_q increase and as n decreases.

Hardison (1971, Table 3) and Moss and Karlinger (1974) reported equations that summarize the effective record length associated with a regression equation, when the dependent variable is an annual mean flow or the standard deviation of the annual flow. However, their equations assume that the annual flows are independent and normally distributed. In this study we account for the dependence and nonnormal structure of the annual flows that have been observed (Vogel and Wilson 1996; Vogel et al. 1998).

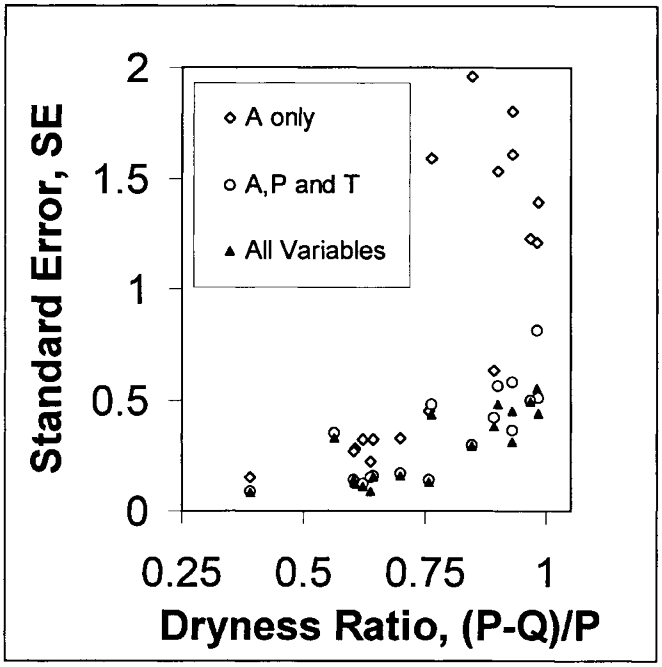


FIG. 2. Relationship between Standard Error of Regressions and Dryness Ratio

The coefficient of variation of a regression estimate of the mean or variance of annual streamflow can be determined from (5). Rearranging (5) one obtains the variance of a regression estimate of the mean of $\mathrm{Var}[\hat{\mu}_R] = \mu_q^2 SE^2$ and the variance of a regression estimate of the variance as $\mathrm{Var}[\hat{\sigma}_R^2] = \sigma_q^2 SE^2$, where SE^2 is defined in (5). Now equating these values with the variance of the sample estimates of μ_q and σ_q^2 given in (11) and (12) leads the following relations for the average effective record length associated with a regression estimate of μ_q and σ_q^2:

$$n_{\hat{\mu}_R} \cong \frac{\ln(1 + C_v^2)}{SE^2} \left[\frac{1 + \rho_y}{1 - \rho_y} + \frac{\ln(1 + C_v^2)}{2} \left[\frac{1 + \rho_y^2}{1 - \rho_y^2} \right] \right] \tag{13}$$

$$n_{\hat{\sigma}_R^2} = \frac{2 \ln(1 + C_v^2)}{C_v^4 SE^2} \left[2C_v^2 \left[\frac{1 + \rho_y}{1 - \rho_y} \right] + [1 + 2C_v^2]^2 \ln(1 + C_v^2) \left[\frac{1 + \rho_y^2}{1 - \rho_y^2} \right] \right] \tag{14}$$

where SE^2 is given (5).

Fig. 4 reports effective record lengths computed using (13) and (14) for each of the regression equations summarized in Tables 2–4. Regional mean values of C_v and ρ_y are obtained from Vogel et al. (1998) and values of SE correspond to each of the regression equations. The effective record length represents the length of record that would be equivalent to use of the regression equation at a (typical) ungauged site. Fig. 4 documents that use of scale alone (drainage area) to predict mean annual runoff is only equivalent to at most a few years of streamflow information, and usually much less, especially in the semiarid western regions. The figure also documents the enormous increases in the effective record lengths obtained by

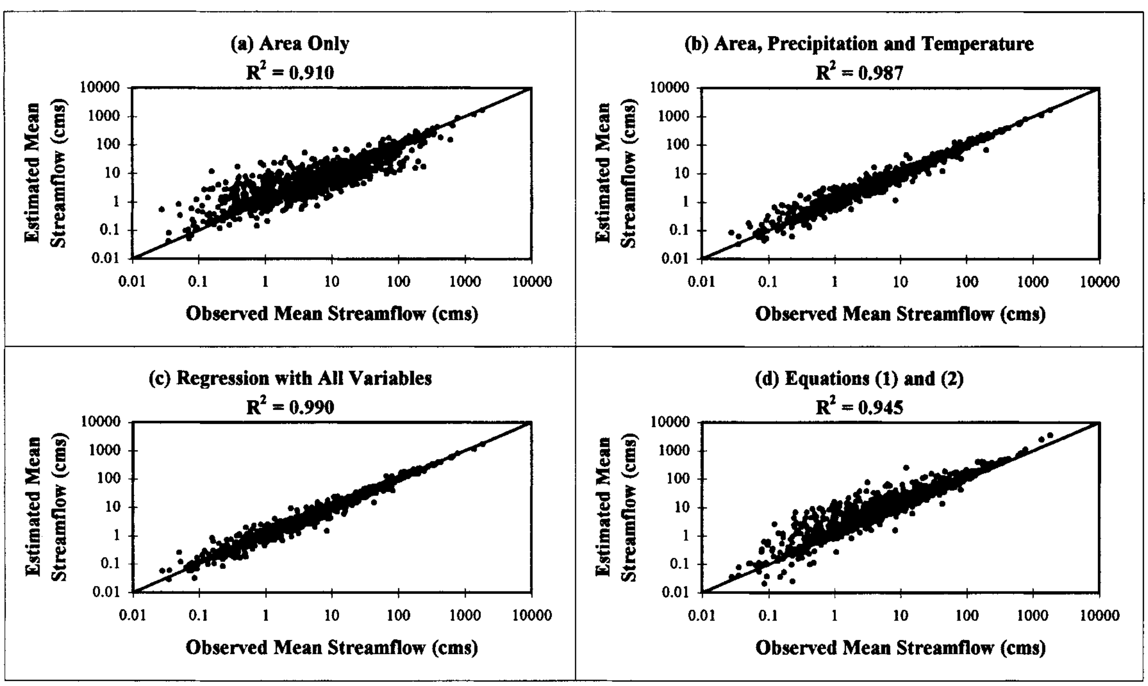


FIG. 3. Comparison of Estimated and Observed Mean Annual Streamflow Using Regression (a) with *A* Only; (b) *A*, *P*, and *T*; (c) All Variables; and (d) Pike Relationship in (1) and (2)

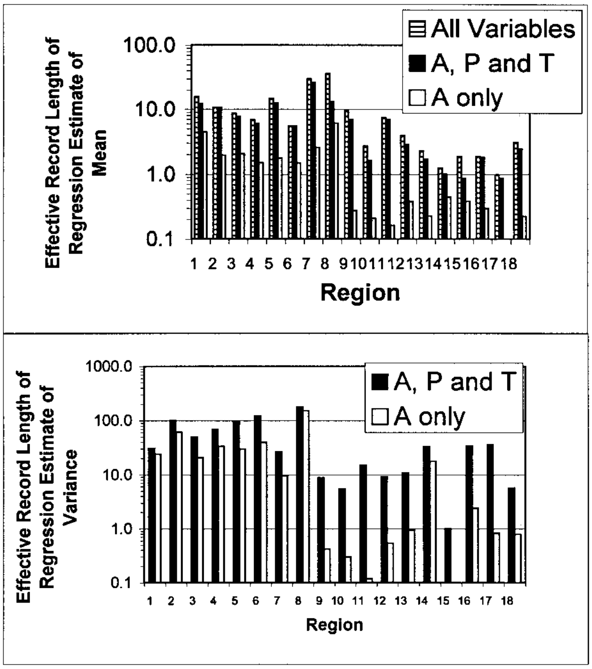


FIG. 4. Comparison of Effective Record Lengths Associated with Regression Estimates of Mean and Variance, Using Scale Only (*A* Only) versus Using Scale and Climate (*A*, *P*, and *T*) and All Variables

the inclusion of both precipitation and temperature into the regression equations. Inclusion of both scale and climate information produces regression equations that are equivalent to several years or sometimes even decades of streamflow information. The increases in information that result from inclusion of all variables in Table 1 is only marginal, in comparison with the gains resulting from inclusion of average annual temperature and precipitation.

CONCLUSIONS

This study has sought to develop general relationships between climate, geomorphology, and annual average streamflow for the conterminous United States. Regression relationships for the mean and variance of annual streamflow as a function of drainage area, A, mean basin precipitation μ_P, mean basin temperature μ_T, and other basin characteristics are developed for each of the 18 water resource regions depicted in Fig. 1. The resulting regional regression relationships exhibit adjusted R^2 values ranging from 90.2 to 99.8%, with an average value of 96.2%. Tables 2 and 3 and Fig. 2 documented significant increases in the precision of regional regression equations when climate is added to scale to predict mean annual streamflow. The use of drainage area alone often led to regressions with equivalent record lengths of only a few years or less, whereas when precipitation, temperature, and other variables are included in the models their precision dramatically improves.

The regional relations introduced here provide significant improvements over the use of runoff maps, because these relationships account for variations in climate that cannot be accounted for when using a runoff map. Furthermore, the relationships introduced here have the added advantage that their information content can be quantified using the concept of effective record length. Future research seeks to develop improvements in the structure of regional models of annual streamflow through the use of models that are more physically based. Although physical models may lead to more reliable and precise predictions, it may not be possible to quantify the equivalent record length associated with such models, as was the case for the statistical models introduced here.

The goodness-of-fit of regional hydrologic models of annual streamflow seems to depend significantly upon the dryness ratio. These results indicate that improvements in our ability to describe basin aridity should lead to improvements in our ability to estimate mean annual streamflow at ungauged sites. The regional regression relationships introduced here are also shown to be an improvement over previous empirical annual models developed by Turc (1954), Pike (1964), and others based on thousands of river basins worldwide.

APPENDIX I. DERIVATION OF (11) AND (12)

The following section documents first-order Taylor-series approximations to the variance of MLEs for an AR (1) log-normal model using the standard method of first-order variance estimation (Tung 1996).

Derivation of Var[$\hat{\mu}_q$] and Var[$\hat{\sigma}_q^2$]. Approximating (7) and (8) with a first-order Taylor series about μ_y and σ_y^2, and applying the variance operator, leads to

$$\mathrm{Var}[\hat{\mu}_q] \cong \left[\frac{\partial\hat{\mu}_q}{\partial\mu_y}\right]^2 \mathrm{Var}[\hat{\mu}_y] + \left[\frac{\partial\hat{\mu}_q}{\partial\sigma_y^2}\right]^2 \mathrm{Var}[\hat{\sigma}_y^2] + 2\left[\frac{\partial\hat{\mu}_q}{\partial\mu_y}\right]\cdot\left[\frac{\partial\hat{\mu}_q}{\partial\sigma_y^2}\right]\mathrm{Cov}[\hat{\mu}_y, \hat{\sigma}_y^2] \tag{15}$$

and

$$\mathrm{Var}[\hat{\sigma}_q^2] \cong \left[\frac{\partial\hat{\sigma}_q^2}{\partial\mu_y}\right]^2 \mathrm{Var}[\hat{\mu}_y] + \left[\frac{\partial\hat{\sigma}_q^2}{\partial\sigma_y^2}\right]^2 \mathrm{Var}[\hat{\sigma}_y^2] + 2\left[\frac{\partial\hat{\sigma}_q^2}{\partial\mu_y}\right]\cdot\left[\frac{\partial\hat{\sigma}_q^2}{\partial\sigma_y^2}\right]\mathrm{Cov}[\hat{\mu}_y, \hat{\sigma}_y^2] \tag{16}$$

where the estimators $\hat{\mu}_q$ and $\hat{\sigma}_q^2$ are obtained from (7) and (8), respectively, replacing μ_y and σ_y^2 with the estimators $\hat{\mu}_y$ and $\hat{\sigma}_y^2$ in those functions. The necessary partial derivatives in (15) and (16), evaluated at μ_y and σ_y^2, are

$$\frac{\partial\hat{\mu}_q}{\partial\mu_y} = \exp\left[\mu_y + \frac{\sigma_y^2}{2}\right] \tag{17}$$

$$\frac{\partial\hat{\mu}_q}{\partial\sigma_y^2} = \frac{1}{2}\exp\left[\mu_y + \frac{\sigma_y^2}{2}\right] \tag{18}$$

$$\frac{\partial\hat{\sigma}_q^2}{\partial\mu_y} = 2C_v^2\exp(2\mu_y + \sigma_y^2) \tag{19}$$

$$\frac{\partial\hat{\sigma}_q^2}{\partial\sigma_y^2} = [2C_v^2 + 1]\exp(2\mu_y + \sigma_y^2) \tag{20}$$

The logarithms of streamflow y follow an AR (1) normal process, in which case Loucks et al. (1981 Appendix 3C) document that

$$\mathrm{Var}[\hat{\mu}_y] = \frac{\sigma_y^2}{n}\left[1 + \frac{2\rho_y}{n}\left[\frac{n}{(1-\rho_y)} - \frac{(1-\rho_y^n)}{(1-\rho_y)^2}\right]\right] \tag{21}$$

$$\mathrm{Var}[\hat{\sigma}_y^2] \cong \frac{2\sigma_y^4}{n}\left[\frac{1+\rho_y^2}{1-\rho_y^2}\right] \tag{22}$$

and

$$\mathrm{Cov}[\hat{\mu}_y, \hat{\sigma}_y^2] = 0 \tag{23}$$

Eq. (21) can be simplified by noting that for most situations encountered in this study, the term $(1-\rho_y^n)/(1-\rho_y)^2$ is negligible in comparison to $n/(1-\rho_y)$, so that

$$\mathrm{Var}[\hat{\mu}_y] \cong \frac{\sigma_y^2}{n}\left[1 + \frac{2\rho_y}{1-\rho_y}\right] = \frac{\sigma_y^2}{n}\left[\frac{1+\rho_y}{1-\rho_y}\right] \tag{24}$$

Substitution of (17)–(24) into (15) and (16) leads to the results in (11) and (12).

ACKNOWLEDGMENTS

Although the research described in this article has been funded in part by the United States Environmental Protection Agency through grant number R 824992-01-0 to Tufts University, it has not been subjected to the Agency's required peer and policy review and therefore does not necessarily reflect the views of the Agency, and no endorsement should be inferred. The writers are indebted to Thomas A. Adams, of the National Weather Service, for providing us with a digital grid of mean annual pan evaporation data for the United States. We are also indebted to the two anonymous reviewers whose comments led to significant improvements in our manuscript.

APPENDIX II. REFERENCES

Arcview GIS 3.1 and Arcview Spatial Analyst 3.1. (1998). Environmental Systems Research Institute, Inc., Redlands, Calif.

Arnell, N. W. (1995). ''Grid mapping of river discharge.'' *J. Hydro.*, Amsterdam, 167, 39–56.

Bishop, G. D., and Church, M. R. (1992). ''Automated approaches for regional runoff mapping in the Northeastern United States.'' *J. Hydro.*, Amsterdam, 138, 361–383.

Busby, M. W. (1963). ''Yearly variations in runoff for the conterminous United States, 1931–60.'' *U.S. Geological Survey Water-Supply Paper 1669-S*, U.S. Government Printing Office, Washington, D.C.

Cook, R. D., and Weisberg, S. (1982). *Residuals and influence in regression*. Chapman and Hall, New York.

Daly, C., Neilson, R. P., and Phillips, D. L. (1994). ''A statistical-topographic model for mapping climatological precipitation over mountainous terrain.'' *J. Appl. Meteorology*, 33(2), 140–158.

Daly, C., Taylor, G., and Gibson, W. (1997). ''The PRISM approach to mapping precipitation and temperature.'' *Proc., 10th AMS Conf. on Appl. Climatology*, American Meteorological Society, Boston, Mass.

Dooge, J. C. I. (1992). ''Sensitivity of runoff to climate change: A Hortonian approach.'' *Bull. Am. Meteorological Soc.*, 73(12), 2013–2024.

Duell, L. F. W. Jr. (1994). ''The sensitivity of Northern Sierra Nevada streamflow to climate change.'' *Water Resour. Bull.*, 30(5), 841–859.

Eagleson, P. S. (1978). ''Climate, soil and vegetation 7. A derived distribution of annual water yield.'' *Water Resour. Res.*, 14(5), 765–776.

Fiering, M. B. (1967). *Streamflow synthesis*. Harvard Press, Cambridge, Mass.

Finlayson, B. L., and McMahon, T. A. (1992). ''Global Runoff.'' *Encyclopedia of Earth System Science*, Vol. 2, Academic, San Diego.

Frind, E. O. (1969). ''Rainfall-runoff relationships expressed by distribution parameters.'' *J. Hydro.*, Amsterdam, 9, 405–426.

Gebert, W. A., Graczyk, D. J., and Krug, W. R. (1987). ''Average annual runoff in the United States, 1951–80.'' *Hydrologic Investigations Atlas*, HA-70, U.S. Geological Survey, Reston, Va.

Hardison, C. H. (1971). ''Prediction error of regression estimates of streamflow characteristics at ungaged sites.'' *U.S. Geological Survey Profl. Paper 750-C*, U.S. Government Printing Office, Washington, D.C., C228–C236.

Hawley, M. E., and McCuen, R. H. (1982). ''Water yield estimation in western United States.'' *J. Irrig. and Drain. Div.*, ASCE, 108(1), 25–35.

Jennings, M. E., Thomas, W. O. Jr., and Riggs, H. C. (1994). ''Nationwide summary of U.S. Geological Survey regional regression equations for estimating magnitude and frequency of floods for ungaged sites.'' *Water-Resources Investigations Rep. 94-4002*, U.S. Geological Survey, Reston, Va.

Johnson, C. G. (1970). ''A proposed streamflow data program for central New England.'' *Open-File Rep.*, U.S. Geological Survey, Boston, Mass.

Kalinin, G. P. (1971). *Global hydrology*. National Technical Information Service, Springfield, Va.

Kuhnel, V., Dooge, J. C. I., O'Kane, P. J. and Romanowicz, R. J. (1991). ''Partial analysis applied to scale problems in surface moisture fluxes.'' *Surveys in Geophysics*, Vol. 12, 221–247.

Langbein, W. B. et al. (1949). ''Annual runoff in the United States.'' *Geological Survey Circular 5*, U.S. Dept. of the Interior, Government Printing Office, Washington, D.C.

Langbein, W. B., and Slack, J. R. (1979). ''Yearly variations in runoff and frequency of dry years for the conterminous United States,'' 1911–79, *Open-File Rep. 82-751*, U.S. Geological Survey, Reston, Va.

Lettenmaier, D. P., Wood, E. F., and Wallis, J. R. (1994). ''Hydro-climatological trends in the continental United States,'' 1948–88, *J. Climate*, 7, 586–607.

Lins, H. F. (1985). ''Streamflow variability in the United States.'' *J. Climate Appl. Meteor.*, 24, 463–471.

Lins, H. F. (1997). ''Regional streamflow regimes and hydroclimatology of the United States.'' *Water Resour. Res.*, 33(7), 1655–1667.

Loucks, D. P., Stedinger, J. R., and Haith, D. A. (1981). *Water resources systems planning and analysis*. Prentice-hall, Englewood Cliffs, N.J.

Lull, H. W., and Sopper, W. E. (1966). ''Factors that influence streamflow in the northeast.'' *Water Resour. Res.*, 2(3), 371–379.

Majtenyi, S. I. (1972). ''A model to predict annual watershed discharge.'' *J. Hydr. Div.*, ASCE, 93(7), 1171–1186.

Markovic, R. D. (1965). ''Probability of best fit to distributions of annual precipitation and runoff.'' *Hydro. Paper No. 8*, Colorado State University, Fort Collins, Colo.

McMahon, T. A., Finlayson, B. L., Haines, A. T., and Srikanthan, R. (1992). *Global runoff—continental comparisons of annual flows and peak discharges*. Catena Paperback.

Milly, P. C. D. (1994a). ''Climate, interseasonal storage of soil water, and the annual water balance.'' *Advances in Water Resources*, Vol. 17, Elsevier Science, Amsterdam, 19–24.

Milly, P. C. D. (1994b). ''Climate, soil water storage, and the average annual water balance.'' *Water Resour. Res.*, 30(7), 2143–2156.

Moss, M. E., and Karlinger, M. R. (1974). ''Surface water network design by regression analysis simulation.'' *Water Resour. Res.*, 10(3), 427–433.

Orsborn, J. F. (1974). ''Determining streamflows from geomorphic parameters.'' *J. Irrig. and Drain. Div.*, 100(4), 455–475.

Pike, J. G. (1964). ''The estimation of annual runoff from meteorological data in a tropical climate.'' *J. Hydro.*, Amsterdam, 2, 116–123.

Revelle, R. R., and Waggoner, P. E. (1983). ''Effects of carbon dioxide induced climate change on water supplies in the western United States.'' *Changing Climate*, Carbon Dioxide Assessment Com., National Academy Press, Washington, D.C., 419–432.

Risbey, J. S., and Entekhabi, D. (1996). ''Observed Sacramento basin streamflow response to precipitation and temperature changes and its relevance to climate impact studies.'' *Journal of Hydro.*, 184, 209–223.

Slack, J. R., Lumb, A. M., and Landwehr, J. M. (1993). ''Hydroclimatic data network (HCDN): A U.S. Geological Survey streamflow data set for the United States for the study of climate variation, 1874–1988.'' *Water-Resource Investigations Rep. 93-4076*, U.S. Geological Survey, Washington, D.C. ⟨http://wwwrvares/er/usgs.gov/hcdn_cdrom/1st_page.html⟩.

Snow survey and water supply products reference. (1988). USDA Soil Conservation Service, Snow Survey Program, West National Technical Center, Portland, Ore.

Stedinger, J. R., and Tasker, G. D. (1986). ''Regional hydrologic analysis 1. Ordinary, weighted and generalized least squares compared.'' *Water Resour. Res.*, 21(9), 1421–1432.

Tasker, G. D. (1980). ''Hydrologic regression with weighted least squares.'' *Water Resour. Res.*, 16(6), 1107–1113.

Taylor, G. H., Daly, C., Gibson, W. P., and Sibul-Weisberg, J. (1997). ''Digital and map products produced using PRISM.'' *Proc., 10th AMS Conf. on Appl. Climatology*, American Meteorological Society, Boston, Mass.

Thomas, D. M., and Benson, M. A. (1970). ''Generalization of streamflow characteristics from drainage-basin characteristics.'' *U.S. Geological Survey Water-Supply Paper 1975*, U.S. Geological Survey, Washington, D.C.

Tung, Y.-K. (1996). ''Chapter 7: Uncertainty and reliability analysis.'' *Water resources handbook*, L. W. Mays, ed. in chf., Mc-Graw Hill, New York, 7.3.2.1.

Turc, L. (1954). ''Le Bilan D'eau des sols. Relation entre les precipitation, l'evaporation et l'ecoulement.'' *Annales Agronomique*, 5, 491–595 (in French).

Vogel, R. M., Bell, C., and Fennessey, N. M. (1997). ''Climate, streamflow and water supply in the northeastern United States.'' *J. Hydro.*, Amsterdam, 198, 42–68.

Vogel, R. M., Tsai, Y., and Limbrunner, J. F. (1998). ''The regional persistence and variability of annual streamflow in the United States.'' *Water Resour. Res.*, 34(12), 3445–3459.

Vogel, R. M., and Wilson, I. (1996). ''The probability distribution of annual maximum, mean and minimum streamflows in the United States.'' *J. Hydrol. Engrg.*, ASCE, 1(2), 69–76.

Yates, D. N. (1997). ''Approaches to continental scale runoff for integrated assessment models.'' *J. Hydro.*, Amsterdam, 201, 289–310.