Developing a Watershed Characteristics Database to Improve Low Streamflow Prediction

Chuck Kroll[1]; Joana Luz[2]; Brad Allen[3]; and Richard M. Vogel[4]

Abstract: Information regarding topographic, meteorologic, geologic, and geomorphic characteristics is increasingly available in spatially explicit digital formats. Of interest is whether enhanced spatial processing of newly available digital grids can lead to new estimators of watershed characteristics which may in turn, improve our ability to predict extreme hydrologic events. Regional hydrologic models of low-flow processes often produce estimators with unacceptably large errors. Using a continuous digital elevation model (DEM) of the conterminous United States, watershed boundaries were developed for the streamflow gauges of the USGS's Hydro-Climatic Data Network. Using these watershed boundaries, many watershed characteristics were developed from digital grids, including: the original DEM, the USDA's State Soil Geographic grids, and the Spatial Climate Analysis Service's orographically weighted precipitation and temperature grids of varying spatial and temporal resolution. Digital processing of grids leads to improvements in estimation and reproducibility of spatial statistics over traditional manual processing approaches. Low-flow regional regression models were developed for regions across the conterminous United States. Inclusion of the new watershed characteristics led to improvements in regional regression models for all regions. The inclusion of hydrogeologic indices, in particular a new smoothed baseflow recession constant estimator, led to dramatic improvements in low-flow prediction.

DOI: 10.1061/(ASCE)1084-0699(2004)9:2(116)

CE Database subject headings: Low flow; Predictions; Geographic information systems; Watersheds; Databases; Streamflow; Digital techniques.

Introduction

Understanding the frequency and duration of extreme hydrologic events is critical to the efficient management of water resources throughout the world. Floods and droughts are responsible for large monetary and human losses every year. While flood frequency analyses have received considerable attention in the research literature, the estimation of low streamflow statistics has received relatively little attention. Low streamflows are especially important for water quality management, where they provide critical dilution of nonpoint source and point source pollution discharges during dry periods, and water quantity management, where low streamflows greatly influence water use policy. For example, in every state, estimates of low streamflow statistics are needed for issuing and/or renewing National Pollution Discharge Elimination System permits, as required by provisions in the Clean Water Act of 1977. Low streamflow statistics are also used to plan water supply, hydropower, and irrigation systems, design cooling-plant facilities, site treatment plants and sanitary landfills, determine waste-load allocations, and make decisions regarding interbasin transfers of water and allowable basin withdrawals. In addition, low streamflow events are often critical periods for aquatic habitats due to potentially low dissolved oxygen concentrations and/or high pollutant concentration.

When a sufficient historic record is available at the river site of interest, low streamflow statistics may be obtained using a frequency analysis (Riggs 1965, 1968). When no historic streamflow record is available, a regional regression model may be developed. Regional regression techniques to estimate low streamflow statistics at ungauged river sites have been employed with varying degrees of success in a limited number of regions throughout the United States (Thomas and Benson 1970; Thomas and Cervione 1970; Parker 1977; Bingham 1986; Vogel and Kroll 1990, 1992; Dingman and Lawlor 1995) and elsewhere [see Smakhtin (2001), for recent review]. These techniques require a relationship between low streamflow statistics and topographic, meteorologic, geologic, and geomorphic characteristics of watersheds to be developed.

In most instances, the standard errors associated with low-flow regression models have been relatively high (Vogel and Kroll 1992; Smakhtin 2001). One reason may be low-flow processes are too complex to be described with a linear or log-linear model. Another reason may be that important explanatory variables have been excluded from these models, and/or the watershed characteristics employed as explanatory variables have not been of high quality. In this study a new database of watershed characteristics

[1]Associate Professor, Environmental Resources and Forest Engineering, SUNY College of Environmental Science and Forestry, Syracuse, NY 13210. E-mail: cnkroll@esf.edu

[2]Research Assistant, Environmental Resources and Forest Engineering, SUNY College of Environmental Science and Forestry, Syracuse, NY 13210;

Assistant Professor, Instituto de Geociências, Salvador, Brazil.

[3]Research Assistant, Environmental Resources and Forest Engineering, SUNY College of Environmental Science and Forestry, Syracuse, NY 13210.

[4]Professor, WaterSHED Center, Civil and Environmental Engineering, Tufts Univ., Medford, MA 02155. E-mail: richard.vogel@tufts.edu

Note. Discussion open until August 1, 2004. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on December 27, 2001; approved on June 6, 2003. This paper is part of the ***Journal of Hydrologic Engineering***, Vol. 9, No. 2, March 1, 2004. ©ASCE, ISSN 1084-0699/2004/2-116–125/$18.00.

is developed for stream gauges that are contained within the United States Geological Survey's Hydro-Climatic Data Network, or HCDN. The HCDN database contains high quality streamflow data at over 1,500 locations spatially distributed across the United States, and has been employed in numerous watershed studies [for examples see: Vogel et al. (1999); Kroll and Vogel (2002); Douglas et al. (2000), and citations therein].

A new database is developed using spatially explicit digital information regarding the topography, meteorology, geology, and geomorphology within each HCDN watershed. A 1 km digital elevation model (DEM) is employed to delineate watershed boundaries for each of the HCDN watersheds. Using these boundaries, summary statistics are estimated using a variety of digital grids including the United States Geological Survey's (USGS) 30 arc sec ($\sim$1 km) Hydro 1 K digital elevation model (DEM), the 1 km U.S. Department of Agriculture's (USDA) State Soil Geographic (STATSGO) grids, a 40-year monthly time series of the Spatial Climate Analysis Service's (PRISM) 0.5° ($\sim$49 km) orographically weighted precipitation and maximum and minimum temperature, and PRISM's 2.5 arc min ($\sim$4 km) average monthly and annual precipitation grids.

Our study addresses the following two related questions:

1. Can low-flow regional regression models be improved by the inclusion of digitally derived watershed characteristics?
2. What are the most important hydrogeologic characteristics to include in low-flow regional regression models?

Prior studies of low-flow regional regression models have met with only limited success, yet those studies are mostly limited to eastern regions of the United States. Using the newly derived database of watershed characteristics, low-flow regional regression models are developed for USGS water resource regions across the entire conterminous United States, allowing interregional comparisons between competing models. Of interest is whether new digital spatial information can improve the predictive capabilities of regional low-flow models. This is the initial study in a series that explores improvements in low-flow prediction across the United States.

In addition to gridded digital data, our database of watershed characteristics includes four hydrogeologic indices: three versions of the baseflow recession constant (Vogel and Kroll 1996) and the baseflow index (Institute of Hydrology 1980). Vogel and Kroll (1992) showed the importance of including hydrogeologic indices in low-flow regional regression equations in Massachusetts. Here we investigate whether this conclusion can be extended to other regions of the United States, and which indices are most important to include in regional low-flow models. Two new smoothed baseflow recession constant estimators are presented in an attempt to reconcile difficulties with the varying precision of the reported streamflow records.

While some of the methodology employed in this paper is not new, the goal of this paper is to perform a series of low-flow regional regression analyses across the conterminous United States with a variety of new explanatory variables. The results of this study provide practitioners in each water resource region numerous watershed characteristics which one should consider as potential explanatory variables in a low-flow regional regression analysis.

This paper is broken into the following sections. The first and second sections are the Abstract and Introduction, respectively. The section, ''HCDN Watershed Characteristics and Streamflow Records'' describes the HCDN watersheds examined in this study, the watershed characteristics database included within the HCDN, and the estimation of at-site streamflow statistics. The section ''Watershed Boundary Delineation'' explains the method used to delineate watershed boundaries, while the section ''New Watershed Characteristics'' discusses our methodology for processing the digital grids and the development of new watershed characteristics. The section ''Regression Analysis'' describes the methodology employed in developing low-flow regional regression equations, and the section ''Results'' presents a comparison of various low-flow models both within and across geographic regions. The last section contains the paper's conclusions and future research directions.

HCDN Watershed Characteristics and Streamflow Records

The USGS's HCDN consists of streamflow records for sites throughout United States. The HCDN streamflow data meet certain measurement accuracy criteria as outlined by Slack and Landwehr (1993). In this study, only sites designated as having streamflow suitable on a daily time step were employed, resulting in 1,545 sites. The HCDN contains river flows from 1874 to 1988, with an average record length of 44 years.

The daily streamflow records were used to obtain estimates of the 7-day annual minimum streamflow which is on average exceeded nine out of every ten years, $Q_{7,10}$. The $Q_{7,10}$ is the most widely used low-flow statistic in the United States (Riggs 1980). In general, the USGS uses a log-Pearson type 3 (LP3) distribution to describe annual minimum streamflow series, as evidenced by its use in a variety of USGS studies (Barnes 1986; Wandle and Randall 1993; Rumenik and Grubbs 1996). It is important to note that there is no consensus as to the best methodology to perform a low-flow frequency analysis at a gauged river site. For instance, there appears to be no consensus as to the best probability distribution (Condie and Nix 1975; Tasker 1987; Vogel and Kroll 1989; Pearson 1995; Vogel and Wilson 1996; Önöz and Bayazit 1999; Kroll and Vogel 2002). In addition, if a poor distributional fit is present, other methods have been advocated, such as tail models (Durrans 1996; Durrans et al. 1999), nonparametric kernel-based methods (Tasker 1987), and graphical techniques (Riggs 1972).

For the sake of this study we assumed that 7-day annual minimum streamflows were adequately described by a LP3 distribution. We estimated the parameters of the LP3 distribution by the method of moments (Stedinger et al. 1993), and the $Q_{7,10}$ as the 10th percentile of the distribution. At sites with between 0 and 10% of 7-day annual minimum flows reported as zero, we used a conditional probability adjustment to estimate the $Q_{7,10}$ (Jennings and Benson 1969; Haan 1977). Any site with more than 10% of 7-day annual minimum flows reported as zero was assigned a $Q_{7,10}$ value of zero. Since we employed a simple log-linear regression model in this study, sites with a $Q_{7,10}$ estimated as zero were removed from the analysis, resulting in the removal of 243 sites. An alternative approach would be to employ a (Tobit) censored regression model in regions with at-site quantile estimates of zero (Kroll and Stedinger 1999).

The HCDN database contains a small collection of watershed characteristics for each of the gauged river sites, including drainage area, main channel slope, main channel length, mean basin elevation, mean annual precipitation, 2-year, 24 hour precipitation intensity, and mean January minimum temperature. Many of these watershed characteristics were developed using manual techniques, employing relatively old information, or were developed from a limited record (Slack et al. 1993). In many regions these watershed characteristics were not reported for all HCDN sites, which further complicated the analysis. We removed 180 watersheds with missing HCDN watershed characteristics because we considered it more important to keep watershed characteristics as opposed to sites in this analysis.

Watershed Boundary Delineation

We used a DEM of the United States to delineate the watershed boundaries for each of the remaining HCDN river sites. A relatively coarse DEM, GTOPO30, was employed in this analysis due to the computational challenge of delineating a large number of watersheds. GTOPO30 is a 1-km resolution raster grid of North America and the U.S. territories produced by the USGS. While this DEM may not provide an adequate topographic description to accurately delineate the boundaries of all watersheds, it does allow for the development of an initial database of watershed characteristics for these sites. It is important to note that DEMs contain errors as the result of blunders, systematic errors, and random errors (USGS 1995). These errors impact not only the DEM, but also estimators derived from the DEM. Numerous techniques, such as stochastic simulation (Goovaerts 1997), have been applied to assess the impact of DEM error on terrain modeling (Holmes et al. 2000). Such an analysis typically requires a semivariogram of the DEM error, which is difficult to obtain without estimates of the ''true'' elevation at numerous points across a watershed. With a coarser DEM, such as the GTOPO30, one would generally expect more accurate results for larger watersheds.

A watershed is defined as the upslope area that drains to a specific point on a river. The delineation of a watershed boundary is based on the assumption that water flows downhill. With a gridded DEM, there are a number of different approaches to determine flow pathways within a watershed. Each method, coupled with a flow accumulation algorithm, can be used to produce an estimator of the drainage area. We employed the USGS's Hydro 1 K flow directions, which are based on applying a single flow direction algorithm to the GTOPO30 DEM. Use of more complicated flow routing algorithms such as a multiple direction (Quinn et al. 1991) or steepest decent (Tarboton 1997) appear to impact watershed delineation in only relatively small watersheds. Tarboton (1997) provides a review of flow direction algorithms and their impact on flow paths and watershed delineation. Use of the Hydro 1 K flow direction grid avoided the need for ''filling'' the GTOPO30 DEM. Filling is often required to remove any depressions or flat areas within the DEM, which would adversely impact the flow accumulation algorithm. All digital grids were projected into an equal-area lambert projection. This projection maintains area across the raster grid, and thus should provide a good projection for watershed delineation.

We initially located stream gauges using the quoted latitude and longitude for the gauge reported by the USGS. Unfortunately, the latitude and longitude did not usually correspond to the optimal position of the gauge location within the DEM. This was most likely due to problems with the DEM not representing the true topography at the gauge location (insufficient resolution or errors in grid) or inaccuracies in the reported latitude and longitude. To determine the ''best'' location for each gauge, we used the following search algorithm. All grid positions within a 5 km radius from the original gauge location were searched to locate the grid position with a drainage area as close to the drainage area reported by the USGS as possible. If no position within 10% of the USGS value was found, all grid positions within radii of 7.5, 10, and 15 km were sequentially examined. If no location within a 15 km radius was found to have a drainage area within 10% of the USGS cited drainage area, the site was removed from the analysis. This procedure removed 192 sites from our analysis.

New Watershed Characteristics

After delineating the watershed boundaries, we used numerous digital grids to develop a database of watershed characteristics. In this analysis, we required grids with a continuous coverage of the United States. The digital grids used were the USGS's GTOPO30 DEM, the 1 km USDA's STATSGO soil maps, a 40-year monthly time series of PRISM's 0.5° orographically weighted precipitation and maximum and minimum temperature, and PRISM's 2.5 arc min average monthly and annual precipitation grids. We also employed new hydrogeologic characteristics derived from the streamflow records. Table 1 contains a summary of the watershed characteristics used in this study. The development of each watershed characteristic is discussed below.

Topography: GTOPO30

We used the GTOPO30 DEM coverage of the conterminous United States to derive a number of topographic estimators. The HCDN database contains an estimator of main channel slope. We developed two new watershed slope estimators: SLOPE2, the watershed slope from the highest elevation in the watershed to the watershed outlet; and SLOPE3, the watershed average of the individual planar slopes calculated on a cell-by-cell basis. The planar slope for an individual cell was calculated as

$$\mathrm{SLOPE3} = \left(\left[\frac{(E_4 + 2E_3 + E_2) - (E_6 + 2E_7 + E_8)}{(8 * \text{cell width})} \right]^2 + \left[\frac{(E_4 + 2E_5 + E_6) - (E_2 + 2E_1 + E_8)}{(8 * \text{cell width})} \right]^2 \right)^{1.2}$$

where E_i = elevation of the ith surrounding cell (a total of eight surrounding cells). This slope estimator is presented by Burroughs (1986) and is part of the ArcView Hydrologic Extension (ESRI 1998). In this analysis, the elevation of the gauge location on the GTOPO30 DEM was also used as an explanatory variable.

Soils: STATSGO (MUID)

The soil information was based on the USDA's STATSGO (MUID) digital grids. This is a 1 km resolution grid covering the entire United States. STATSGO, which was developed from the 1994 State Soil Geographic Database, was designed to support regional, multistate, state, and river basin resource planning, management, and monitoring (USGS 2001). For each watershed we computed the average of the high and low range values for the following soil parameters: permeability, organic matter content, available water capacity, high water table, total soil thickness, and bulk density. It should be noted that STATSGO was developed using information from individual states, and often there are discontinuities across state boundaries in these grids.

Climate: PRISM

Two sets of climate data from the Spatial Climate Analysis Service's (PRISM) project were used in this study. The first set of

Table 1. Description of Explanatory Variables

Symbol	Percent time entering model	Variable description	Source
Topographic and HCDN variables			
DA	93	Drainage area	USGS HCDN
SLOPE1	9	Main channel slope	USGS HCDN
SLOPE2	5	Slope of watershed from peak to outlet	GTOPO30 DEM
SLOPE3	14	Average of cell facet slopes	GTOPO30 DEM
ELEV1	17	Gauge elevation	USGS HCDN
ELEV2	3	Gauge elevation from DEM	GTOPO30 DEM
LENGTH	5	Channel length	USGS HCDN
PRECIP	29	Average annual precipitation	USGS HCDN
INTENS	9	Precipitation intensity	USGS HCDN
JANTMIN	9	January minimum temperature	USGS HCDN
Hydrogeologic variables			
K_{b-1}	3	Baseflow recession constant (K_b) based on daily streamflow	Computed
K_{b-2}	86	K_b based on three-days moving average	Computed
K_{b-3}	0	K_b based on decreasing flows in three-day moving average	Computed
BFI	31	Baseflow index	Computed
Geologic variables			
PL	0	Low value for the range of permeability	STATSGO
PH	0	High value for the range of permeability	STATSGO
OML	2	Low value for the range of organic matter content	STATSGO
OMH	0	High value for the range of organic matter content	STATSGO
AWCL	0	Low value for the range of available water capacity	STATSGO
AWCH	2	High value for the range of available water capacity	STATSGO
WDL	2	Low value for the range of depth to the high water table	STATSGO
WDH	3	High value for the range of depth to the high water table	STATSGO
RDL	0	Low value for the range of the total soil thickness	STATSGO
RDH	7	High value for the range of the total soil thickness	STATSGO
BDL	2	Low value for the range of bulk density	STATSGO
BDH	0	High value for the range of bulk density	STATSGO
Climatic variables			
ATMAX	0	90th percentile for maximum temperature for period Jun–Aug	PRISM
BTMAX	5	90th percentile for maximum temperature for period Sep–Nov	PRISM
CTMAX	2	90th percentile for maximum temperature for period Dec–Apr	PRISM
DTMAX	3	90th percentile for maximum temperature for period Apr–Mar	PRISM
ATMIN	7	90th percentile for minimum temperature for period Jun–Aug	PRISM
BTMIN	0	90th percentile for minimum temperature for period Sep–Nov	PRISM
CTMIN	3	90th percentile for minimum temperature for period Dec–Apr	PRISM
DTMIN	0	90th percentile for minimum temperature for period Apr–Mar	PRISM
APRCP	19	10th percentile for precipitation for period Jun–Aug	PRISM
BPRCP	2	10th percentile for precipitation for period Sep–Nov	PRISM
CPRCP	7	10th percentile for precipitation for period Dec–Apr	PRISM
DPRCP	3	10th percentile for precipitation for period Apr–Mar	PRISM
PANN	3	Average annual precipitation using 2.5 arc min grids	PRISM
PJAN	0	Average January precipitation using 2.5 arc min grids	PRISM
PFEB	7	Average February precipitation using 2.5 arc min grids	PRISM
PMAR	2	Average March precipitation using 2.5 arc min grids	PRISM
PAPR	7	Average April precipitation using 2.5 arc min grids	PRISM
PMAY	5	Average May precipitation using 2.5 arc min grids	PRISM
PJUN	10	Average June precipitation using 2.5 arc min grids	PRISM
PJUL	5	Average July precipitation using 2.5 arc min grids	PRISM
PAUG	7	Average August precipitation using 2.5 arc min grids	PRISM
PSEP	9	Average September precipitation using 2.5 arc min grids	PRISM
POCT	5	Average October precipitation using 2.5 arc min grids	PRISM
PNOV	3	Average November precipitation using 2.5 arc min grids	PRISM
PDEC	0	Average December precipitation using 2.5 arc min grids	PRISM

Note: HCDN=Hydro-Climatic Data Network; DEM=digital elevation model; STATSGO=State Soil Geographic Grids; and PRISM=Parameter-Elevation Regressions on Independent Slopes Model.

grids is interpolated raster data of monthly climate variables including precipitation, and minimum and maximum temperatures. A 40-year monthly time series of grids were employed, resulting in 480 grids for each climatic variable. These grids have a resolution of 0.5°, which roughly translates into a projected resolution of 49 km. While these grids have a poor spatial resolution, the temporal resolution produces a unique data set.

The average monthly precipitation, and maximum and minimum temperature were calculated for each of the HCDN watersheds for each of the 480 months, resulting in a time series of monthly averages at every site. Since low streamflows are caused by long-term extremes (as opposed to shorter duration events such as those impacting floods), four "seasonal windows" were selected: (1) June–August; (2) September–November; (3) December–March; and (4) April–May. These four windows capture either typical months when annual minimum streamflow occurs (late summer and early fall), or winter and spring conditions which can impact groundwater storage prior to the low-flow months.

Since extreme climatic conditions such as low precipitation and/or high temperatures are generally responsible for low streamflow events, numerous extreme percentiles for each climatic time series were calculated and included as explanatory variables. Since the $Q_{7,10}$ is the 10th percentile of the distribution of 7-day annual minimums, in this analysis the 10th percentile of the distribution of precipitation and the 90th percentile of the distribution of temperature were estimated. This was accomplished by fitting each temperature series with a generalized extreme value (GEV) distribution with L-moment parameter estimators (Stedinger et al. 1993), and then estimating the 90th percentiles from the distribution. The GEV distribution has been used in practice to describe some climatic variables (Schaefer 1990). For the precipitation series a delta distribution with a point mass at zero and a two-parameter lognormal (LN2) distribution describing the nonzero observations was employed (Aitchison 1955). The parameters of the LN2 distribution were obtained by method of moments (Stedinger et al. 1993). The 10th percentile of the delta distribution was then estimated.

For the second climatic data set, PRISM's 2.5 arc min ($\sim$4 km) average monthly and annual precipitation grids were employed. This consists of 13 grids: 12 grids of monthly averages, and 1 grid of annual averages. While the temporal resolution of these grids is poor (averages over 40 years), the spatial resolution is much better than those of the first climatic data set.

Hydrogeology

None of the above watershed characteristics capture the hydrogeologic behavior of the watershed. Since low streamflow is generally the result of groundwater discharge to the stream during times of little or no precipitation, we expect hydrogeology to be an important parameter in explaining low streamflow processes and low streamflow statistics (Vogel and Kroll 1992). Most of the information in the MUID data set is near surface soil parameters and is not necessarily representative of a watershed's underlying aquifers. To address this issue, two hydrogeologic statistics were calculated: the baseflow recession constant (K_b) and the baseflow index (BFI). It should be noted that historic streamflow records were required to estimate K_b and BFI, and thus these parameters cannot be derived at ungauged sites using the techniques described below.

K_b is an estimator of the daily percentage decline in streamflow during times of no surface or shallow subsurface runoff. We estimated K_b based on Method 5 outlined in Vogel and Kroll (1996). This method is derived from the continuity equation when outflow from a watershed is linearly related to basin groundwater storage

$$\frac{dQ}{dt}=aQ^b=aQ^1 \tag{1}$$

resulting in the least squares estimator

$$K_b=\exp\left\{-\exp\left[\frac{1}{m}\sum_{t=1}^{m}\left\{\ln(Q_{t-1}-Q_t)-\ln\left(\frac{1}{2}(Q_t+Q_{t-1})\right)\right\}\right]\right\} \tag{2}$$

where Q_t = streamflow on Day t; and m = total number of streamflow pairs (Q_t and Q_{t-1}). Based on a linear solution to the Boussinesq equation, K_b is a function of hydraulic conductivity, porosity, drainage density, and groundwater slope (Vogel and Kroll 1992). Vogel and Kroll (1996) compared six K_b estimators' abilities to describe low streamflow statistics in Massachusetts. They found the K_b estimator in Eq. (2) to be the preferred estimator due to both its performance and simplicity.

To determine K_b, only streamflows that occur during a groundwater recession were employed. A streamflow recession was defined by at least a ten-day drop in a three-day moving average. The first 30% of the recession was removed to limit the impact of surface and shallow subsurface stormflows. The remaining days were considered the groundwater recession. This is also the procedure followed in Vogel and Kroll (1992, 1996).

In this study three K_b estimators were developed. The first, K_{b-1}, was developed using Eq. (2), employing any pairs of two consecutive decreasing daily streamflows during a groundwater recession period. The other two K_b estimators were developed to smooth errors due to the varying precision of USGS reported streamflow values. USGS streamflows less than 1 cfs are reported with two significant digits, all flows between 1 and 10 cfs are reported with one digit after the decimal point, flows between 10 and 1,000 cfs as integers, and flows greater than 1,000 cfs with three significant digits. Thus there are lower bounds on the value of dQ/dt in Eq. (1) over each of these ranges of streamflow. For instance, the minimum value of dQ/dt is 0.01 for flows less than 1 cfs, but is 0.1 for flows between 1 and 10 cfs. Kroll (1989) and Eng and Brutsaert (1999) examined the issue of varying streamflow measurements precision on baseflow recession analyses. In an attempt to smooth the data to reduce the impact of varying precision, we developed a new estimator K_{b-2}, which employs consecutive 3-day moving averages for the terms Q_{t-1} and Q_t in Eq. (2) (as opposed to daily streamflows in K_{b-1}). The other new estimator K_{b-3}, is similar to K_{b-2}, but has the requirement that over the four days which make up the two consecutive three-day moving averages, the streamflow must be decreasing. This restriction produced a large reduction in the sample size for the K_{b-3} estimator compared to the sample sizes for K_{b-1} and K_{b-2}.

BFI measures the long-term average fraction of annual streamflow that is contributed from groundwater. The BFI estimator employed was based on an Institute of Hydrology (1980) study. The BFI estimator uses a moving window approach to determine days in which the streamflow is comprised solely of groundwater and then linearly interpolates between these days to determine the baseflow contribution to streamflow. The entire streamflow record was used to estimate the BFI.

Regression Analysis

When no discharge record exists at a site of interest, a regional regression model can be used to estimate low streamflow statistics at the ungauged site. Using a region of gauged river sites, this method requires a relationship between low streamflow statistics and topographic, meteorologic, geologic, and geomorphic characteristics to be developed. These relationships most often have the following form:

$$Q_{d,T}=\alpha X_1^{\beta} X_2^{\gamma}\ldots \tag{3}$$

where $Q_{d,T}=d$-day; T-year low-flow statistic; $X_i=$drainage basin characteristics; and α, β, and $\gamma=$model parameters. Vogel and Kroll (1992) showed that Eq. (3) has a form consistent with the linear solution to the Boussinesq equation for groundwater discharge. The dependent variable in Eq. (3), $Q_{7,10}$, was obtained using at-site quantile estimates from gauged river sites. By taking the logarithm of both sides of Eq. (3), the model parameters can be estimated using ordinary (OLS), weighted (WLS), or generalized least squares (GLS) regression procedures (Stedinger and Tasker 1985; Kroll and Stedinger 1998). Once the model parameters have been estimated in a region, low-flow estimates at ungauged sites can be obtained using drainage basin characteristics for the ungauged site.

Sites with $Q_{7,10}$ estimated as zero were eliminated from the analysis. Kroll and Stedinger (1999) showed that when only a few sites have zero quantile estimates, dropping these sites did not adversely impact the regression model parameter estimators. For moderate to high censoring levels (10 to 50% of the sites in a region), a (Tobit) model should be employed. In this study, OLS regression estimators were employed to estimate the model parameters. Because model error variance is typically high in low-flow regional regression models, the model error variance tends to overwhelm the time sampling error in the models. Thus we expect only slight differences in the parameter estimators when WLS and GLS are compared to OLS (Kroll and Stedinger 1998).

After we removed sites due to $Q_{7,10}$ values estimated as zero, inaccurate watershed delineation, and missing HCDN watershed characteristics, 930 sites remained. Initially the 18 USGS regions in the conterminous United States were used to develop low-flow regression models, but this produced many extremely heterogeneous regions (such as Region 3 which spans from Florida to Virginia). Therefore, for this analysis state boundaries were used. If a state did not have at least 20 sites, it was combined with the adjacent state with the fewest sites. This approach resulted in 29 regions. While the decision to use state boundaries could be criticized due to the lack of hydrologic homogeneity in many states, many watershed characteristics are developed independently by states (such as the STATSCO data) and are stored in state-based geographic information system (GIS) clearinghouses. In fact, the last major national study of regional regression models for flood frequency reported results on a state-by-state basis (Jennings et al. 1994).

To develop a regional regression model in a specific region, a stepwise regression procedure was employed using a 5% significance level on the entering variables. One major issue is high correlation (multicollinearity) among the explanatory variables in the regression model. Multicollinearity can cause regression estimators to have inflated and correlated errors, which can produce inaccuracies in subsequent hypothesis tests regarding parameter significance (Johnston 1972). To handle this situation, a variance inflation factor (VIF), which indicates the possible presence of multicollinearity, was employed (Rawlings et al. 1998). The VIF is a function of the coefficient of determination (R^2) obtained by regressing each individual explanatory variable against all the other explanatory variables. A VIF ($\mathrm{VIF}=1/(1-R^2)$) greater than ten was used as a threshold to indicate possible multicollinearity problems (Rawlings et al. 1998). When this occurred, variables with high correlation (such as drainage area and stream length, or K_{b-1}, K_{b-2}, and K_{b-3}) were entered into the model individually. In an ongoing study, we are investigating the impact of using of principal components (Jolliffe 1986) as explanatory variables in our regression models.

Results

The results are broken into three sections. In the first section, models developed using the HCDN watershed characteristics were compared with models developed by including the new watershed characteristics. In the second section a comparison of model performance with competing variables (such as the three K_b estimators) is made. In the third section, a regional intercomparison of the low-flow models is presented.

Watershed Characteristics Comparison

This section provides an analysis of the impact of the new data sets on low-flow regional regression models. Regression models were developed in each of the 29 regions for six different sets of explanatory variables: the HCDN variables; HCDN plus PRISM's climatic grids; HCDN plus MUID and the new topographic estimators (TOPO); HCDN, PRISM and TOPO (ALL3); HCDN plus Hydrogeology; and HCDN, PRISM, TOPO, and Hydrogeology (ALL4). To compare the impact of each of these data sets, two performance metrics were calculated: the adjusted coefficient of determination (Adj-R^2) (Devore 1994) and the percent standard error of prediction (SE%). SE% was computed as $100[\exp(S_\varepsilon^2)-1]^{1/2}$, where S_ε^2 is an estimate of the variance of the residuals developed using Hardison's (1971) variance of the "space-sampling error." Similar conclusions were reached using mean square error of model estimators and a prediction error sum of squares statistic, which is a validation-type estimator of error (Helsel and Hirsch 1992). This section compares the overall performance of these data sets. The section "Variable Comparisons" examines the most important variables from each of these data sets.

Figs. 1 and 2 contain box plots of Adj-R^2 and SE%, respectively, for each of the data sets across the 29 regions. The solid line across each box represents the median, the ends of the box the 25th and 75th percentiles, the ends of the whiskers the 10th and 90th percentiles, and the circles values outside this range. With just the HCDN variables, the median Adj-R^2 was 67.6%, with a maximum of 92.7% and a minimum of 21.1%. The median, maximum, and minimum SE% were 124, 617, and 36.7%, respectively. With the addition of the MUID and new topographic variables (HCDN+TOPO), slight improvements in the models were made, with a median Adj-R^2 of 70.8, with a maximum of 94.0% and a minimum still of 21.1% (SE% median, maximum, and minimum were 107, 470, and 32.8%, respectively). PRISM's climatic grids produced a greater improvement than TOPO, with a median of 74.6%, a maximum of 92.7%, and a minimum of 34.4% (SE% median, maximum, and minimum were 97.0, 617, and 32.6%, respectively).

When three databases were included (ALL3), the median Adj-R^2 increased to 77.5%, with a maximum of 94.0% and a

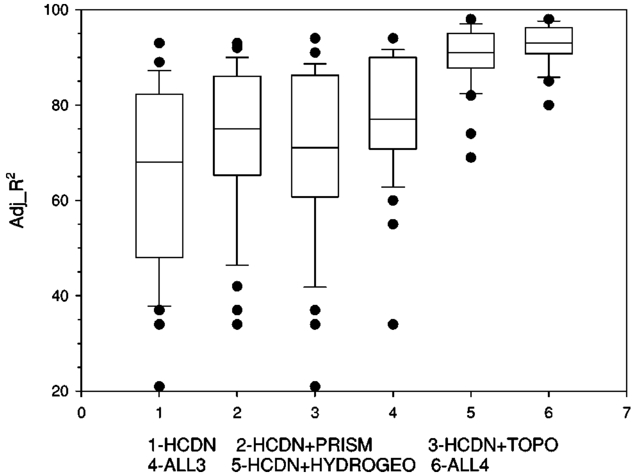

Fig. 1. Adjust coefficient of variation (Adj-R^2) for models from 29 regions across the conterminous United States using six data sets

minimum of 34.4%, while the median SE% was 90.1%, with a maximum of 470% and a minimum of 32.6%. With the inclusion of the new digitally derived spatial statistics, low-flow regional regression models were improved in every region of the United States. Unfortunately, in many regions low-flow regional regression models are still inadequate since they produce estimators with large variances, and thus are not suitable for design purposes.

In this analysis, we also considered the inclusion of three estimators of the baseflow recession constant (K_b) and one estimator of the baseflow index (BFI). These hydrogeologic indices were estimated from the historic records at each of the sites, and thus we are currently unable to estimate these indices accurately at ungauged sites. Of interest is whether these indices improve low-flow regional regression models. With the inclusion of these indices with only the HCDN variables (HCDN +HYDROGEO), the median Adj-R^2 rose to 90.8%, with a maximum of 98.4% and a minimum of 69.2%, while the SE% median was 55.9%, with a maximum of 166% and a minimum of 13.1%. When all four data sources (ALL4) were included, the median Adj-R^2 was 93.1%, the maximum was 98.5%, and the minimum was 79.5%, while for the SE% the median was 43.3%, the maximum 112%, and the minimum 10.4%. These results indicate that the inclusion of hydrogeologic indices in low-flow models produces dramatic increases in model performance.

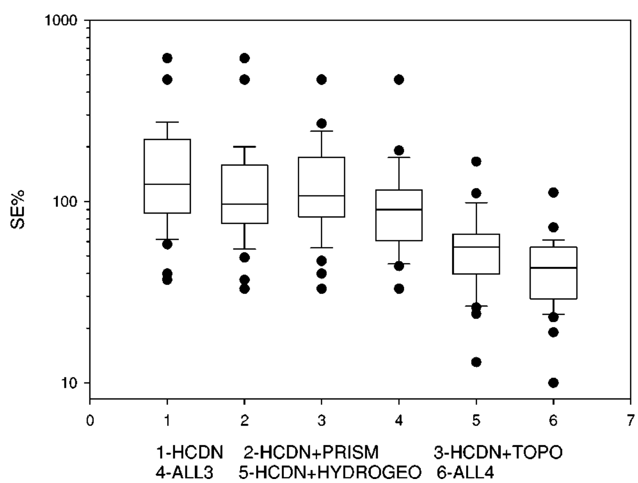

Fig. 2. Percent standard error of prediction (SE%) for models from 29 regions across conterminous United States using six data sets

Variable Comparisons

The derived watershed characteristics allow for an interesting comparison between similar watershed characteristics and their impact on low streamflow models. In this section, a number of these comparisons are made. To facilitate the discussion, Table 1 contains a column representing the percentage of the time an explanatory variable entered a final model. For all variables except hydrogeology, this includes the 29 models where HCDN, PRISM, and TOPO (ALL3) were included, as well as when Hydrogeology (ALL4) was included, for a total of 58 models. For the hydrogeologic variables, only the 29 models from ALL4 are considered.

Of the topographic parameters, drainage area entered the regression models most frequently. When drainage area did not enter a model, main channel length did. Eng and Brutsaert (1999) showed that channel length and drainage area are typically highly correlated within a region. To avoid multicollinearity programs, in most regions drainage area and main channel length were never included in the same model (the VIF was greater than 10).

All three slope estimators entered some of the final models, with the average watershed slope (SLOPE3) entering slightly more often than the other two slope estimators. Gauge elevation from the USGS HCDN entered the model much more often than the gauge elevation from the DEM, indicating potential inaccuracies within the DEM or errors associated with siting the gauge within the DEM.

None of the PRISM precipitation statistics entered the model more often than the HCDN value of mean annual precipitation. Of the PRISM precipitation statistics, the 10th percentile of average precipitation over the period from June to August entered most often. This statistic was derived from the 40-years of low resolution monthly precipitation time series grids. Of the higher resolution monthly average precipitation grids, statistics for the summer and early fall months generally entered the models more often than those for the winter months. Results potentially indicate that spring recharge of groundwater resources is less important to low streamflow processes than summer precipitation quantities.

In general, the PRISM temperature grids did not enter the models frequently. The exception is for the 90th percentile of the maximum and minimum temperature for the period from September to November. This result indicates the importance of evapotranspiration to low-flow processes in some regions during the-later summer and early fall months, which is when low streamflows primarily occur. The statistics derived from the STATSGO (MUID) soils database generally performed poorly, with no variable entering more than 5% of the models.

For every region examined, either the baseflow recession constant estimators or the baseflow index entered the final model. Of the baseflow recession constant estimators K_{b-2}, which was derived using three-day moving averages, was almost always preferred to the other two baseflow recession constant estimators. K_{b-2} was developed to smooth lower bounds on the change in streamflow over time (dQ/dt), which occurs due to varying precision of reported streamflow values. K_{b-3} was probably not preferred due to the reduction in sample size for this estimator. BFI, which attempts to capture groundwater storage characteristics, entered 31% of the final models. Results again emphasize the importance of hydrogeology in low streamflow prediction.

Table 2. Final Regression Models

States	Number of sites	ALL4: All variables — Variables in the model	Adj-R^2	SE%	ALL3: All variables except Kb and baseflow index — Variables in the model	Adj-R^2	SE%
Maine, New Hampshire, and Vermont	31	DA KB2 PJUN	95.4	25.0	DA PAN ELEV	92.3	32.6
Connecticut, Rhode Island, and Massachusetts	20	DA KB2	94.5	53.1	DA PRECIP POCT OML	93.5	54.4
New York	34	DA ELEV JANMIN KB2 CPRCP	97.5	19.5	DA PRECIP JANMIN CPRCP BTMIN	93.2	32.8
New Jersey	21	DA KB1 APRCP PFEB SLOPE3 RDH	96.9	27.8	DA ELEV CPRCP PJUN PAUG SLOPE3 BDL WDH WDH	91.3	51.4
Pennsylvania	30	DA KB2 BFI CPRECP	94.4	28.9	DA PMAY SLOPE2 WDL	79.2	61.8
Delaware and Maryland	24	DA KB2 PFEB	93.1	35.6	DA RDH WDH	87.3	46.8
Virginia	32	DA LENGTH KB2 BFI BTMAX AWCH	95.2	25.1	DA PAPR PAUG RDH	70.8	80.7
Georgia	58	DA ELEV KB2 BFI POUT	96.2	44.6	DA ELEV1 CPRPC PMAY PJUN PJUL SLOPE3 OML	90.4	130
North Carolina and South Carolina	46	DA KB2 BFI SLOPE2	90.1	51.3	DA BPRCP PJUN PJUL SLOPE3	79.9	134
Florida	21	KB2 PSEP	85.1	112	LENGTH DPRECP PAPR	75.2	174
Alabama, Tennessee, and Kentucky	29	DA SLOPE1 PRECIP KB2 BFI PAUG ELEV	97.6	22.6	DA PRECIP	55.4	175
Ohio and West Virginia	22	DA KB2 BTMAX PJUL	95.8	28.7	DA LENGTH BTMAX BPRECIP SLOPEBR	84.5	44.0
Indiana and Michigan	27	DA KB2 BTMAX	97.2	31.7	DA INTENS	81.0	99.0
Wisconsin and Minnesota	43	DA LENGTH, PRECIP, ATMIN	70.7	59.9	DA PRECIP ATMIN	67.4	93.3
Iowa	27	DA SLOPE1 BFI APRPC PSEP PFEB	93.1	34.7	DA INTENS DPRCP PAPR PSEP	90.9	68.5
Illinois	23	DA KB2 APRPC PSEP PNOV	96.6	43.3	DA APRCP PJUN	90.2	74.2
Missouri	20	DA KB2 PAN	97.4	39.9	DA ELEV INTENS POUT SLOPE2	71.2	107
Mississippi	24	DA KB2 PAPR PAUG PSEP PAN LCV	98.5	10.4	LENGTH PRECIP	72.2	77.3
Arkansas and Louisiana	26	DA PRECIP BFI	90.8	56.2	JANMIN ATMIN SLOPE3 WDH	60.5	158
Texas	26	DA KB2 APRCP DTMAX	88.2	72.4	DA ELEV INTENS CTMAX	80.5	111
New Mexico and Arizona	26	DA PRECIP KB2	87.2	56.2	DA ELEV PNOV RDH	71.7	90.3
Colorado, Nevada, and Utah	44	DA PRECIP KB2	87.6	54.7	DA PRECIP	77.5	78.9
Kansas, Nebraska, and Oklahoma	25	DA KB2 BFI APRCP DTMAX	91.8	37.2	DA	37.7	470
Wyoming and South Dakota	36	DA PRECIP KB2 SLOPE3	83.7	60.0	DA PRECIP	70.2	100
Montana and North Dakota	27	DA PRECIP BFI	90.8	58.6	DA ELEV PFEB PJUN	82.2	90.1
Idaho	54	DA ELEV PRECIP KB2 APRCP CTMIN	95.2	45.4	DA ELEV PRECIP CTMIN PJUL	89.8	74.5
Washington	30	DA SLOPE1 PRECIP KB2 APRCP	95.6	34.2	DA INTENS SLOPE2	89.8	56.1
Oregon	43	LENGTH KB2 PAUG	91.8	44.2	LENGTH ELEV INTENS JANMIN ATMIN SLOPEBR	67.6	109
California	61	DA JANMIN KB2 APRCP ATMIN	92.2	61.8	DA APRCP	66.6	191

Regional Performance

The models developed for each of the regions using the ALL3 and ALL4 variable databases are presented in Table 2. ALL3 contains all variables except hydrogeology, and ALL4 contains all variables. In general, low-flow regional regression models perform well in the northern United States. Wisconsin and Minnesota are an exception. Snow accumulation and snow melt processes are not captured in the databases we developed, and may impact low-flow processes in Wisconsin and Minnesota. Our results agree with the findings of Vogel and Kroll (1990) and Dingman and Lawlor (1995) that low-flow regional regression models perform well in the northeastern United States, even without the inclusion of hydrogeologic variables. Models in the southern regions generally perform worse than those in the northern regions. Additionally, model performance along the southeastern coastal regions is generally better than the southwestern coastal regions.

The regions that produce the worst low-flow models are: Kansas, Nebraska, and Oklahoma; Alabama, Tennessee, and Kentucky; and Arkansas and Louisiana. One reason for this may be these regions are relatively large, and thus some of the processes that impact the regional differences in low-flow processes are not captured by the databases. Interestingly, when hydrogeologic variables are included in the models, all of these regions produce Adj-R^2 values in excess of 90%, and SE% less than 40%.

Conclusions and Future Directions

Spatial processing of newly available gridded topographic, meteorologic, geologic, and geomorphic data using a geographic information system (GIS) can generate spatially representative watershed based information useful for the development of regional hydrologic models. This initial study has generated a new set of watershed characteristics for the HCDN watersheds that were then used to develop regional relationships for estimating low-flow statistics. This database is publicly available on the internet at http://www.esf.edu/erfeg/cnkroll/research. Our results indicate the following:

1. In all regions of the conterminous United States, low-flow regional regression models were improved with the inclusion of watershed characteristics from the newly developed spatially processed digital databases, when compared with watershed characteristics derived from more traditional manual approaches.
2. The performance of low-flow regional regression models varies widely across the United States. In general, the best models were obtained in northern regions of the United States.
3. The inclusion of hydrogeologic variables greatly improved low-flow regional regression models. This result emphasizes the importance of developing new approaches for estimating hydrogeologic variables at ungauged watersheds.
4. The inclusion of climatic variables generally had only a small impact on the models. This could be due to a lack of resolution of the digital climatic grids employed. Further investigation into the impact of climatic grid resolution on the accuracy of regional low-flow models is warranted.
5. The delineation of watershed boundaries using a 1-km DEM was not adequate for all of the sites examined. Often either the incorrect watershed was chosen, or a watershed with an incorrect drainage area was delineated. Future research should examine the impact of using finer resolution DEMs for delineating watershed boundaries.
6. Many other watershed characteristics that may be employed to model low streamflow statistics are either not available in digital form, or their coverage does not include the entire conterminous United States. Including information such as soil surveys, subsurface geology and land use, should lead to improvements in modeling low streamflow in some regions of the United States.
7. While low-flow regional regression models of adequate precision may be formulated in some regions of the United States, initial results indicate that this technique does not perform adequately throughout the entire United States. This further indicates that other techniques, such as baseflow correlation, which requires some discharge measurements at ungauged sites (Stedinger and Thomas 1985), may be needed in some regions.
8. Our initial results document the enormous challenge associated with the estimation of low-flow statistics at ungauged sites for most regions of the United States. Our hope is that this study will inspire future research to explore additional digitally gridded information that can better characterize the hydrogeology of watersheds, because it is this information that should yield the greatest improvements in our ability to develop regional low-flow models.
9. The results of this paper provide regional water resource planners an indication of which watershed characteristics may aid in describing the low streamflow processes. This information should aid in the modeling and management of low streamflows.

References

Aitchison, J. (1955). ''On the distribution of a positive random variable having a discrete probability mass at the origin.'' *J. Am. Stat. Assoc.*, 50, 901–908.

Barnes, C. R. (1986). ''Methods of estimating low-flow statistics for ungaged streams in the lower Hudson River Basin, NY.'' *U.S. Geological Survey Water Resources Investigations Rep. 85–4070*, U.S. Geological Survey, Reston, Va.

Bingham, R. H. (1986). ''Regionalization of low-flow characteristics of Tennessee streams.'' *U.S. Geological Survey Water Resources Investigations, Rep. 85–4191*, U.S. Geological Survey, Reston, Va.

Burroughs, P. A. (1986). *Principles of geographical information systems for land resources assessment*, Oxford University Press, New York.

Condie, R., and Nix, G. A. (1975). ''Modeling of low-flow frequency distributions and parameter estimation.'' *Proc., Int. Water Resource Symposium*, Water for Arid Lands, Teheran, Iran.

Devore, J. L. (1994). *Probability and statistics for engineering and the sciences*, 4th Ed., Duxbury Press, Belmont, Mass.

Dingman, S. L., and Lawlor, S. C. (1995). ''Estimating low-flow quantiles from drainage-basin characteristics in New Hampshire and Vermont.'' *Water Resour. Bull.*, 31(2), 243–256.

Douglas, E. M., Vogel, R. M., and Kroll, C. N. (2000). ''Trends in flood and low flows in the United States: Impact of spatial correlation.'' *J. Hydrol.*, 240(1–2), 90–105.

Durrans, S. R. (1996). ''Low-flow analysis with a conditional Weibull tail model.'' *Water Resour. Res.*, 32(6), 1749–1760.

Durrans, S. R., Ouarda, T. B. M. J., Rasmussen, P. F., and Bobée, B. (1999). ''Treatment of zeroes in tail modeling of low flows.'' *J. Hydrologic Eng.*, 4(1), 19–27.

Eng, K., and Brutsaert, W. (1999). ''Generality of drought flow characteristics within the Arkansas River Basin.'' *J. Geophys. Res., [Atmos.]*, 104(19), 435–441.

Environmental Systems Research Institute (ESRI). (1998). *ArcView spatial analyst online user guide*, Redlands, Calif.

Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*, Oxford Univ. Press, New York.

Haan, C. T. (1977). *Statistical methods in hydrology*, Iowa State Univ. Press, Ames, Iowa.

Hardison, C. H. (1971). ''Prediction error of regression estimates of streamflow characteristics at ungaged sites.'' *U.S. Geological Survey Professional Paper, 750–C, C228–C236*, U.S. Geological Survey, Reston, Va.

Helsel, D. R., and Hirsch, R. M. (1992). *Statistical methods in water resources*, Elsevier, New York.

Holmes, K. W., Chadwick, O. A., and Kyriakidis, P. C. (2000). ''Error in a USGS 30-meter digital elevation model and its impact on terrain modeling.'' *J. Hydrol.*, 233, 154–173.

Institute of Hydrology. (1980). ''Low-flow studies.'' *Rep. No 1*, Wallingford, Oxon, U.K.

Jennings, M. E., and Benson, M. A. (1969). ''Frequency curves for annual flood series with some zero events or incomplete data.'' *Water Resour. Res.*, 5(1), 276–280.

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, 1993.'' *U.S. Geological Survey Water-Resources Investigations Rep., 94–4002*, Reston, Va. Va.

Johnston, J. (1972). *Econometric methods*, 2nd Ed., McGraw-Hill, New York.

Jolliffe, I. T. (1986). *Principal component analysis*, Springer, New York.

Kroll, C. N. (1989). *The estimation and usage of baseflow recession constants*, Master's thesis, Tufts Univ., Medford, Mass.

Kroll, C. N., and Stedinger, J. R. (1998). ''Generalized least squares regression procedures revisited.'' *Water Resour. Res.*, 34(1), 121–128.

Kroll, C. N., and Stedinger, J. R. (1999). ''Development of regional regression relationships with censored data.'' *Water Resour. Res.*, 35(3), 775–784.

Kroll, C. N., and Vogel, R. M. (2002). ''The probability distribution of low streamflow series in the United States.'' *J. Hydrologic Eng.*, 7(2), 137–146.

Önöz, B., and Bayazit, M. (1999). ''GEV-PWM model for distribution of minimum flows.'' *J. Hydrologic Eng.*, 4(3), 289–292.

Parker, G. W. (1977). ''Methods for determining selected flow characteristics for streams in Maine.'' *U.S. Geological Survey Open-File, Rep. 78–871*, U.S. Geological Survey, Reston, Va.

Pearson, C. P. (1995). ''Regional frequency analysis of low flows in New Zealand rivers.'' *J. Hydrol.*, 30(2), 53–64.

Quinn, P. E., Beven, K. J., Chevallier, P., and Planchon, O. (1991). ''The prediction of hillslope flow paths for distributed hydrologic modeling using digital terrain models.'' *Hydrolog. Process.*, 5, 59–79.

Rawlings, J. O., Pantula, S. G., and Dickey, D. A. (1998). *Applied regression analysis: A research tool*, 2nd Ed., Springer, New York.

Riggs, H. C. (1965). ''Estimating probability distributions of drought flows.'' *Water Sewage Works*, 112(5), 153–157.

Riggs, H. C. (1968). ''Frequency curves.'' *U.S. Geological Survey techniques of water resource investigations*, Book 4, Chap. A2, U.S. Geological Survey, Reston, Va.

Riggs, H. C. (1972). ''Low-flow investigations.'' *U.S. Geological Survey techniques of water resource investigations*, Book 4, Chap. B1, U.S. Geological Survey, Reston, Va.

Riggs, H. C. (1980). ''Characteristics of low flows.'' *J. Hydraul. Eng.*, 106(5), 717–731.

Rumenick, R. P., and Grubbs, J. W. (1996). ''Methods for estimating low-flow characteristics for ungaged streams in selected areas, Northern Florida.'' *U.S. Geological Survey Water Resources Investigation Rep. 96–4124*, U.S. Geological Survey, Reston, Va.

Schaefer, M. G. (1990). ''Regional analyses of precipitation annual maxima in Washington State.'' *Water Resour. Res.*, 26(1), 119–131.

Slack, J. R., Lumb, A. M., and Landwehr, J. M. (1993). ''Hydro-climatic data network (HCDN): A U.S. Geological Survey streamflow data set for the United States for the study of climate variations, 1874–1988.'' *U.S. Geological Survey Water Resources Investigation Rep., 93–4076*, U.S. Geological Survey, Reston, Va.

Smakhtin, V. U. (2001). ''Low-flow hydrology: A review.'' *J. Hydrol.*, 240(3–4), 147–186.

Stedinger, J. R., and Tasker, G. D. (1985). ''Regional hydrologic analysis 1. Ordinary, weighted, and generalized least squares compared.'' *Water Resour. Res.*, 21(9), 1421–1432.

Stedinger, J. R., and Thomas, W. O., Jr. (1985). ''Low-flow frequency estimation using base-flow measurements.'' *U.S. Geological Survey Open-File Rep., 85–95*, U.S. Geological Survey, Reston, Va.

Stedinger, J. R., Vogel, R. M., and Foufoula-Georgiou, E. (1993). ''Frequency analysis of extreme events.'' *Handbook of hydrology*, D. R. Maidment, ed., Chap. 18, McGraw-Hill, New York.

Tarboton, D. G. (1997). ''A new method for the determination of flow direction and upslope areas in grid digital elevation models.'' *Water Resour. Res.*, 33, 309–319.

Tasker, G. D. (1987). ''A comparison of methods for estimating low-flow characteristics of streams.'' *Water Resour. Bull.*, 23(6), 1077–1083.

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, Reston, Va.

Thomas, M. P., and Cervione, M. A. (1970). ''A proposed steamflow data program for Connecticut.'' *Connecticut Water Resources Bulletin*, No. 23, U.S. Geological Survey, Reston, Va.

USGS. (1995). *National mapping program technical instructions Part 2 specifications*, Dept. of the Interior, Washington, D.C.

USGS. (2001). ''State soil geographic (STATSGO) data base for the conterminous United States.'' ⟨http://water.usgs.gov/GIS/metadata/usgswrd/ussoils.html⟩ (July 6, 2001).

Vogel, R. M., and Kroll, C. N. (1989). ''Low-flow frequency analysis using probability-plot correlation coefficients.'' *J. Water Resour. Plan. Manage.*, 115(3), 338–357.

Vogel, R. M., and Kroll, C. N. (1990). ''Generalized low-flow frequency relationships for ungauged sites in Massachusetts.'' *Water Resour. Bull.*, 26(2), 241–253.

Vogel, R. M., and Kroll, C. N. (1992). ''Regional geohydrologic-geomorphic relationships for the estimation of low-flow statistics.'' *Water Resour. Res.*, 28(9), 2451–2458.

Vogel, R. M., and Kroll, C. N. (1996). ''Estimation of baseflow recession constants.'' *Water Resour. Manage.*, 10, 303–320.

Vogel, R. M., and Wilson, I. (1996). ''Probability distribution of annual maximum, mean, and minimum streamflows in the United States.'' *J. Hydrologic Eng.*, 1(2), 69–76.

Vogel, R. M., Wilson, I., and Daly, C. (1999). ''Regional regression models of annual streamflow for the United States.'' *J. Irrig. Drain. Eng.*, 125(3), 148–157.

Wandle, S. W., Jr., and Randall, A. D. (1993). ''Effects of surficial geology, lakes and swamps, and annual water availability on low flows of streams in central New England, and their use in low-flow estimation.'' *U.S. Geological Survey Water Resources Investigations Rep., 93–4092*, U.S. Geological Survey, Reston, Va.