Regional Flow-Duration Curves for Ungauged Sites in Massachusetts

By Neil Fennessey[1] and Richard M. Vogel,[2] Associate Members, ASCE

Abstract: A regional hydrologic model is developed for estimating flow-duration curves at ungauged and unregulated basins in Massachusetts. Flow-duration curves often exhibit complex shapes, requiring probability density functions with three or more parameters. This study approximates the lower half of daily flow-duration curves using a two-parameter lognormal probability density function. A conjugate gradient algorithm is employed to fit lognormal density functions to the lower half of observed flow-duration curves at 23 basins. Regional regression equations are developed to describe the lognormal model parameters in terms of easily measured basin characteristics. The resulting regional flow-duration model only requires estimates of the watershed area, and a basin relief parameter, both of which are easily obtained from U.S. Geological Survey 7.5-min quadrangle maps. In addition, confidence intervals are derived for flow-duration curves estimated at ungauged sites. A validation experiment reveals that the resulting regional hydrologic model can provide remarkably precise estimates of a flow-duration curve at an ungauged site, considering the simplicity of the model and its ease of application.

Introduction

A flow-duration curve is simply the cumulative distribution function of daily streamflows at a site. Flow-duration curves were used widely during the first half of this century. Evidence of their widespread use is provided by Foster's (1934) description of flow-duration curves as one of the three most familiar graphical tools available to the hydrologist, the other two tools being the hydrograph and the mass curve. The first use of a flow-duration curve is attributed to Clemens Herschel in about 1880 (Foster 1934). Flow-duration curves have been advocated for use in hydrologic studies such as hydropower, water supply, and irrigation planning and design (Chow 1964; Warnick 1984). In perhaps the most complete manual on flow-duration curves ever written, Searcy (1959) describes additional applications to stream-pollution and water-quality management problems. Although most of the articles on flow-duration curves were written during the first half of this century, (Searcy 1959), current textbooks still contain discussions pertaining to this important tool (Linsley and Franzini 1979; Warnick 1984; Gupta 1989).

With increasing attention focused on surface water-quality management, many agencies routinely require estimates of low-flow statistics to assure the maintenance of water-quality standards. Other water resource interests that must be considered, in addition to the assimilation of point and nonpoint source discharges into streams, include recreation, maintenance of wetland habitats and endangered botanical species, and accommodation of stream discharges and withdrawals associated with water resource systems. Each

[1]Engr., Dept. of Envir. Protection Div. of Water Supply, 1 Winter St., Boston, MA 02108.

[2]Asst. Prof., Dept. of Civ. Engrg., Tufts Univ., Medford, MA 02155.

Note. Discussion open until December 1, 1990. 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 July 19, 1989. This paper is part of the ***Journal of Water Resources Planning and Management***, Vol. 116, No. 4, July/August, 1990. ©ASCE, ISSN 0733-9496/90/0004-0530/$1.00 + $.15 per page. Paper No. 24899.

interest places unique constraints on the low-flow regime in terms of the magnitude, frequency, and seasonality of both water-quality and -quantity requirements. Such competing interests create a formidable problem for agencies whose objective is to quantify acceptable (tolerable) low flows for the purposes of managing stream withdrawals and discharges in a river basin.

The management of the low-flow regime in a river basin requires evaluation of streamflow characteristics across a broad range of flow regimes. Flow-duration curves are ideally suited to such tasks because they can be modified to evaluate the impact of streamflow regulation (stream discharges and stream withdrawals) on the resulting magnitude and frequency of streamflows. For example, in the state of Massachusetts, flow-duration curves are currently employed for implementing the Water Management Act (WMA). The WMA regulates surface and ground-water withdrawals that exceed an average of 0.1 mgd for any three consecutive months of the year. Flow-duration curves are currently used in Massachusetts to evaluate the impact of proposed future withdrawals on the net basin yield and low-flow characteristics of a river basin.

Since most locations where flow-duration curves are required are not coincident with stream gages, this study focuses on the development of flow-duration curves for ungauged sites. Regional flow-duration curves for ungauged sites have been developed in Illinois, New Hampshire, the Philippines, and Greece by Singh (1971), Dingman (1979), Quimpo et al. (1983) and Mimikou and Kaemaki (1985), respectively.

Graphical Procedures for Constructing Flow-Duration Curves

Much of the literature on flow-duration curves concentrates upon graphical methods for constructing a flow-duration curve from sequences of daily streamflow data (Foster 1934; Beard 1942; Searcy 1959). Such procedures consist of ranking the observed streamflows q_i, $i = 1, \ldots, 365n$, to produce a set of ordered streamflows $q_{(i)}$, $i = 1, \ldots, 365n$, where n = the number of years of record, $q_{(1)}$ = the largest observation and $q_{(365n)}$ = the smallest observation. The flow-duration curve is constructed by plotting each ordered observation $q_{(i)}$ versus its corresponding plotting position p_i. A plotting position, p_i, is simply an estimate of the exceedance probability, p, associated with the ordered observation $q_{(i)}$.

In this study the Weibull plotting position

$$p_i = \frac{i}{365n + 1} \qquad (1)$$

is employed. Any plotting position (Cunnane 1978) would be reasonable, since they are all virtually indistinguishable for the large sample sizes encountered here. Note that daily streamflow records employed in this study were of length n = 12–72 years corresponding to daily streamflow record lengths, $365n$, equal to 4,352–26,328 daily observations, including leap year days (see Table 1).

A graphical flow-duration curve is essentially a nonparametric, empirical cumulative density function (cdf). Beard (1943) suggested constructing a flow-duration curve by fitting a two-parameter lognormal cdf. Beard sug-

TABLE 1. Basin Characteristics and Flow-Duration Curve Parameters for 23 Sites Used to Develop Regional Flow-Duration Curve Model

U.S. Geologic Survey gage number (1)	Site No. (2)	Record length (years) (3)	Drainage area A (sq mi) (4)	Basin relief H (ft) (5)	$\hat{\mu}$ (6)	$\hat{\sigma}$ (7)
01180500	1	54	52.70	1,765	3.854	1.449
01096000	2	35	63.69	1,161	4.145	1.276
01106000	3	38	8.01	227	2.445	2.569
01170100	4	17	41.39	1,873	3.871	1.205
01174000	5	35	3.39	531	1.353	1.827
01175670	6	24	8.68	417	2.248	1.690
01198000	7	20	51.00	1,317	3.778	1.524
01171800	8	12	5.46	530	1.619	1.202
01174900	9	23	2.85	585	1.060	1.688
01101000	10	39	21.30	277	3.189	2.119
01187400	11	32	7.35	877	2.041	1.748
01169000	12	44	89.00	1,667	4.469	1.217
01111300	13	20	16.02	393	2.900	1.928
01169900	14	18	24.09	1,298	3.340	1.158
01181000	15	48	94.00	1,739	4.542	1.388
01332000	16	53	40.90	2,068	3.881	1.135
01097300	17	21	12.31	248	2.500	1.835
01333000	18	35	42.60	2,658	3.857	1.152
01165500	19	66	12.10	797	2.351	1.427
01171500	20	46	54.00	1,476	3.961	1.138
01176000	21	72	150.00	801	5.086	1.133
01162500	22	66	19.30	718	2.831	1.521
01180000	23	29	1.73	643	0.326	1.650

gested plotting the $q_{(i)}$ versus p_i on lognormal probability paper and drawing a "best-fit" line through the data. Goodness-of-fit tests and hypothesis tests are now available for evaluating and selecting an appropriate probability density function using probability plots (Vogel and Kroll 1989).

Loaiciga (1989) evaluated the use of nonparametric empirical streamflow quantile estimation procedures in the context of floodflow frequency analysis. The expressions that Loaiciga (1989) derived for the variance of an empirical quantile using a Weibull (or other) plotting position are useful for estimating the variance of an empirical quantile of a flow-duration curve at a gauged site. Since daily streamflows exhibit significant serial correlation, however, the effective record length associated with a flow-duration curve will be much smaller than $365n$ (Tasker 1983), where n = the record length in years.

Regional Flow-Duration Curves

The sample sizes of the streamflow records used in this study are so large that there is little statistical advantage to using analytic (parametric) procedures instead of graphical (nonparametric) procedures to construct flow-duration curves at a gauged site. Analytic procedures are usually preferred over graphical procedures in flood studies and low-flow investigations, which use the much shorter sequences of annual maximum and annual minimum streamflows, respectively. Nevertheless, the construction of regional flow-duration curves does require fitting a cumulative probability density function to the observed flow series so that the derived parameters of the cdf may be related to topographic and geomorphic parameters of each watershed.

Analytic Procedures for Constructing Flow-Duration Curve

Our objective is to develop regional regression models that relate parameters of the daily flow-duration curve at a site to topographic and geomorphic basin characteristics. In order to obtain such regression models we require that the flow-duration curve be described by as few parameters as possible. Since flow-duration curves are known to exhibit rather complex shapes, (Searcy 1959; Dingman 1978) three or more parameters are probably necessary to describe the location, shape, and scale of the probability density function. A complex trade-off exists between the number of parameters required to describe the flow-duration curve and our ability to obtain regional regression models that relate those parameters to drainage basin characteristics. For the applications described earlier in Massachusetts, the WMA only requires knowledge of the flow-duration curve between the limits p = 0.5–0.99, where p = the exceedance probability defined by

$$p = P(Q > q_p) \quad (2a)$$

$$p = 1 - P(Q \leq q_p) \quad (2b)$$

In Eqs. 2*a* and *b*, q_p corresponds to that value of mean daily streamflow that is exceeded $p\%$ of the time. Beard (1943) suggested the use of a lognormal probability density function to approximate flow-duration curves. We show here that a two-parameter lognormal function provides a good approximation to the lower half of daily flow-duration curves in Massachusetts. Assuming that daily streamflows Q are distributed lognormal, Eqs. 2*a* and *b* may be rewritten as

$$p = 1 - (2\pi)^{-1/2} \int_{-\infty}^{z_p} \exp\left(-\frac{1}{2}t^2\right)dt \quad (3a)$$

$$p = g(q_p|\mu, \sigma) \quad (3b)$$

where $z_p = [\ln(q_p) - \mu]/\sigma$, which is the pth percentile of a zero mean, unit variance, normally distributed random variable; q_p is defined in Eqs. 2*a* and *b*; and μ and σ = the mean and variance, respectively, of the natural logarithms of the daily streamflows.

Normally, maximum likelihood estimators would be employed to obtain asymptotically unbiased and minimum variance estimates of μ and σ (Stedinger 1980). Here we assume that Eqs. 3*a* and *b* only hold for $0.5 \leq p \leq 0.99$, in which case one cannot use the standard maximum likelihood estimators. Instead, the optimum values of μ and σ, denoted $\hat{\mu}$ and $\hat{\sigma}$, are obtained using an unconstrained optimization algorithm known as the conjugate gradient method. This algorithm is available for use on a personal computer as a Fortran subroutine in the International Mathematical Subroutine Library ("Users" 1987). The conjugate gradient algorithm is used to solve the minimization problem over graphical procedures in flood studies and low-flow investigations, which use the much shorter sequences of annual maximum and annual minimum streamflows, respectively. Nevertheless, the construction of regional flow-duration curves does require fitting a cumulative probability density function to the observed flow series so that the derived parameters of the cdf may be related to topographic and geomorphic parameters of each watershed.

Analytic Procedures for Constructing Flow-Duration Curve

Our objective is to develop regional regression models that relate parameters of the daily flow-duration curve at a site to topographic and geomorphic basin characteristics. In order to obtain such regression models we require that the flow-duration curve be described by as few parameters as possible. Since flow-duration curves are known to exhibit rather complex shapes, (Searcy 1959; Dingman 1978) three or more parameters are probably necessary to describe the location, shape, and scale of the probability density function. A complex trade-off exists between the number of parameters required to describe the flow-duration curve and our ability to obtain regional regression models that relate those parameters to drainage basin characteristics. For the applications described earlier in Massachusetts, the WMA only requires knowledge of the flow-duration curve between the limits $p = 0.5$–0.99, where $p =$ the exceedance probability defined by

$$p = P(Q > q_p) \qquad (2a)$$

$$p = 1 - P(Q \leq q_p) \qquad (2b)$$

In Eqs. 2*a* and *b*, q_p corresponds to that value of mean daily streamflow that is exceeded p% of the time. Beard (1943) suggested the use of a lognormal probability density function to approximate flow-duration curves. We show here that a two-parameter lognormal function provides a good approximation to the lower half of daily flow-duration curves in Massachusetts. Assuming that daily streamflows Q are distributed lognormal, Eqs. 2*a* and *b* may be rewritten as

$$p = 1 - (2\pi)^{-1/2} \int_{-\infty}^{z_p} \exp\left(-\frac{1}{2}t^2\right) dt \qquad (3a)$$

$$p = g(q_p|\mu,\sigma) \qquad (3b)$$

where $z_p = [\ln(q_p) - \mu]/\sigma$, which is the pth percentile of a zero mean, unit variance, normally distributed random variable; q_p is defined in Eqs. 2*a* and *b*; and μ and σ = the mean and variance, respectively, of the natural logarithms of the daily streamflows.

Normally, maximum likelihood estimators would be employed to obtain asymptotically unbiased and minimum variance estimates of μ and σ (Stedinger 1980). Here we assume that Eqs. 3*a* and *b* only hold for $0.5 \leq p \leq 0.99$, in which case one cannot use the standard maximum likelihood estimators. Instead, the optimum values of μ and σ, denoted $\hat{\mu}$ and $\hat{\sigma}$, are obtained using an unconstrained optimization algorithm known as the conjugate gradient method. This algorithm is available for use on a personal computer as a Fortran subroutine in the International Mathematical Subroutine Library ("Users" 1987). The conjugate gradient algorithm is used to solve the minimization problem

$$\underset{\hat{\mu},\hat{\sigma}}{\text{minimize}}\; H(\mu,\sigma) = \sum_{p=0.5}^{0.99} \left[\frac{p - g(q_p|\hat{\mu},\hat{\sigma})}{p}\right]^2 \qquad (4)$$

where the summation is performed in steps of 0.005, hence there are 99 terms in the sum.

The optimization problem posed in Eq. 4 amounts to a search for the parameter estimates $\hat{\mu}$ and $\hat{\sigma}$, which minimize the sum of the squared residuals for the model in Eqs. 3*a* and *b*, using the empirical flow-duration curve described in Eq. 1 to obtain the necessary values of q_p. The residuals are normalized by dividing by p in Eq. 4 as suggested by Entekhabi et al. (1989). This normalization procedure provides a useful way to control the absolute magnitude, and hence the computational overflow, of the relative difference between the function and the observed data during the minimization/fitting process.

Application of the conjugate gradient algorithm to the problem posed in Eqs. 3*a* and *b* and 4 requires analytic estimates of the partial derivatives $\partial H(\mu,\sigma)/\partial\mu$, and $\partial H(\mu,\sigma)/\partial\sigma$, which may be obtained via the chain rule as

$$\frac{\partial H}{\partial \mu} = \frac{\partial H}{\partial g} \cdot \frac{\partial g}{\partial z_p} \cdot \frac{\partial z_p}{\partial \mu} \qquad (5a)$$

$$\frac{\partial H}{\partial \sigma} = \frac{\partial H}{\partial g} \cdot \frac{\partial g}{\partial z_p} \cdot \frac{\partial z_p}{\partial \sigma} \qquad (5b)$$

The terms $\partial H/\partial g$, $\partial z_p/\partial\mu$, and $\partial z_p/\partial\sigma$ are easily obtained from Eqs. 3*a* and *b* and 4; however, analytic values of $\partial g/\partial z_p$ are difficult to obtain from Eqs. 3*a* and *b*, therefore a series solution to Eqs. 3*a* and *b* is employed so that $\partial g/\partial z_p$ can be derived. Abramowitz and Stegun (1964) provide the series representation of Eqs. 3*a* and *b*

$$g(q_p|\mu,\sigma) = \frac{1}{2} - (2\pi)^{-1/2} \sum_{n=0}^{j} \left[\frac{(-1)^n z_p^{(2n+1)}}{n!2^n(2n+1)}\right] + \epsilon \qquad (6)$$

where again $z_p = [\ln(q_p) - \mu]/\sigma$. The series solution of Eqs. 3*a* and *b* given in Eq. 6, is exact when $j = \infty$, that is, when $j = \infty$, $\epsilon = 0$. In this study we use an approximation to $g(q_p|\mu,\sigma)$ obtained by using enough terms in Eq. 6 to assure that $\epsilon < 10^{-9}$.

Analytic Flow-Duration Curves in Massachusetts

The aforementioned procedures were applied to historic mean daily flows at 23 of the U.S. Geological Survey's streamflow-gauging stations in or near Massachusetts. Stations with the following attributes were selected:

1. All of the rivers are perennial and all streamflows are greater than zero.
2. No significant withdrawals, diversions, or artificial recharge areas are contained in the basins; hence, we consider the streamflows to be essentially unregulated.

Table 1 contains the U.S. Geological Survey's streamflow-gauging station numbers, site numbers for this and other studies (Vogel and Kroll 1989; and

Vogel et al. 1989), and record length in years. The location of each station is shown in Fig. 1. The period of record corresponding to each flow record is documented in Figs. 2(*a*), (*b*), (*c*), and (*d*).

Empirical flow-duration curves were developed for each site by plotting the ordered streamflows, $q_{(i)}$, versus their corresponding plotting position p_i given by Eq. 1. For each value of p over the interval (0.5, 0.99), in steps of 0.005, corresponding values of q_p were obtained from the empirical flow-duration curve using cubic spline interpolation. In many instances interpolation was unnecessary, because the sample sizes are so large that typically an observed discharge $q_{(i)}$ was available for most values of p considered. The resulting empirical flow-duration curves are displayed in Figs. 2(*a*), (*b*), (*c*), and (*d*) using open circles.

The conjugate gradient algorithm was applied to solve the optimization problem in Eq. 4 at each of the stations described in Table 1. The resulting optimal estimates $\hat{\mu}$ and $\hat{\sigma}$ are summarized in Table 1. Eqs. 3*a* and *b* may be inverted to obtain a direct estimate of the pth quantile of average daily streamflow q_p using

$$q_p = \exp(\hat{\mu} + z_p\hat{\sigma}) \quad (7)$$

where a reasonable approximation to z_p, the pth quantile from a standard normal distribution, proposed by Tukey (1960) is

$$z_p = 4.91[(1-p)^{0.14} - p^{0.14}] \quad (8)$$

The approximation in Eq. 8 is excellent as long as $0.01 \le p \le 0.99$, which is the case here. Figs. 2(*a*), (*b*), (*c*), and (*d*) depicts the fitted values of q_p using dotted lines. The fitted values of q_p are obtained from Eqs. 7 and 8 using the optimal parameters $\hat{\mu}$ and $\hat{\sigma}$ given in Table 1. Overall, the fitted relationships described by the dotted lines are in very good agreement with the observed flow-duration curves represented by the open circles in Figs. 2(*a*), (*b*), (*c*), and (*d*). In most instances the agreement is so good that the two curves are virtually indistinguishable! Nevertheless, at about half the gauging stations, the fitted relations (dotted lines) are downward biased for exceedance probabilities in excess of about 0.9. These results suggest that future research should examine the ability of alternative probability density functions to fit the lower tail of flow-duration curves.

Regional Flow Duration Curves in Massachusetts

In this section the analytic flow-duration model represented by Eqs. 7 and 8 is extended for use at an ungauged site by estimating regional regression equations for $\hat{\mu}$ and $\hat{\sigma}$ in Eq. 7. Previous studies in Massachusetts (Male and Ogawa 1982; Tasker 1972; Vogel and Kroll 1990) and elsewhere have developed multiple linear regression equations for estimating low-flow statistics using models of the form

$$\Theta = aX_1^bX_2^cX_3^d \quad (9)$$

where the independent variables X_1, X_2, and X_3 are physiographic, geologic, climatic and/or geomorphic parameters that are easily measured at an ungauged site; a, b, c, and d are fitted model parameters; and Θ is some statistic of the streamflows to be estimated at an ungauged site. Recently, Vogel et al. (1989) derived a model of the form given in Eq. 9 from a conceptual

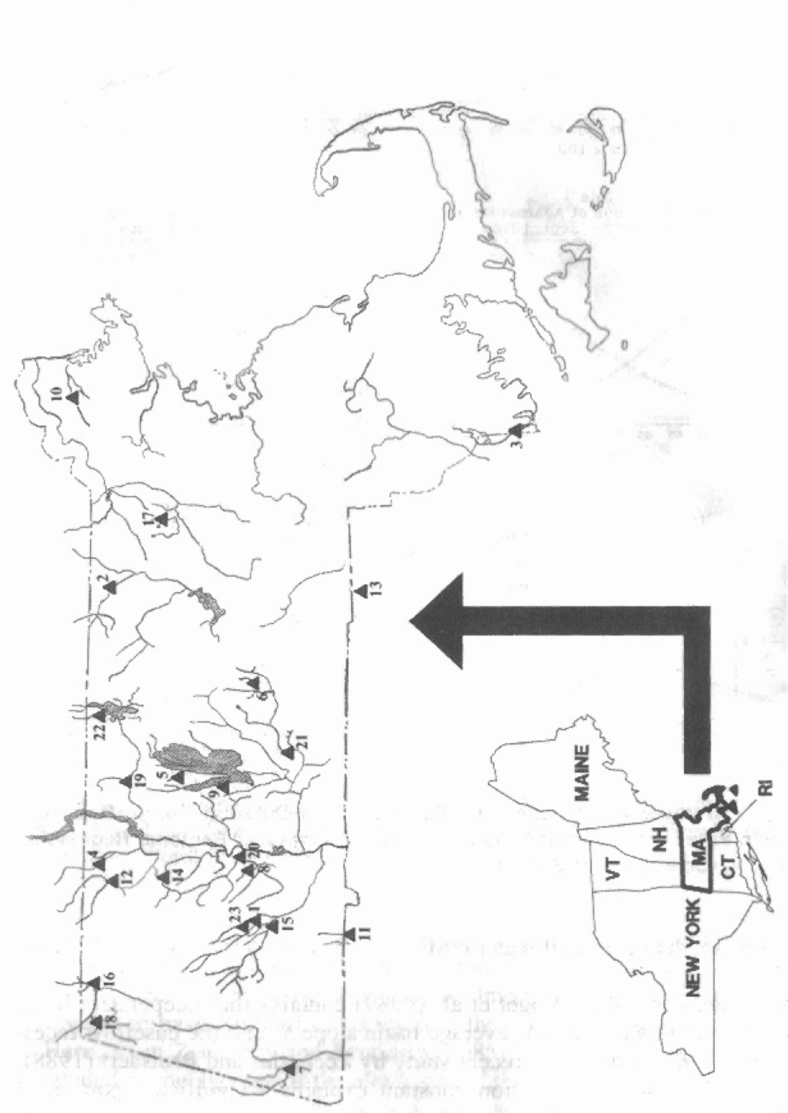

FIG. 1. Location of 23 U.S. Geological Survey Stations Described in Table 1

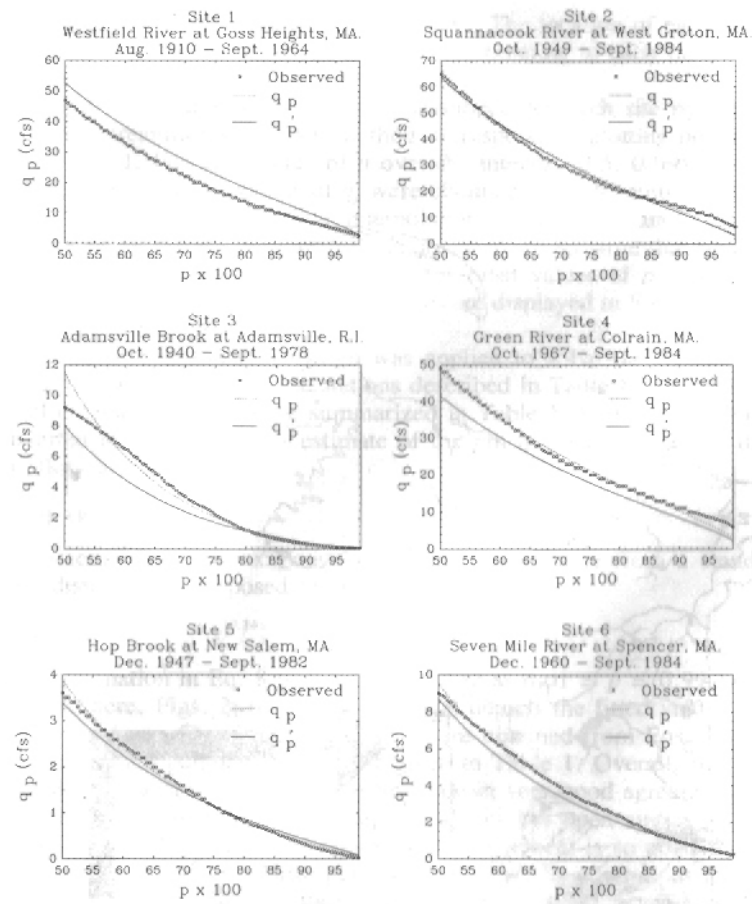

FIG. 2(*a*). Comparison of Observed Empirical Flow-Duration Curve (Open Circles) with Fitted Flow-Duration Curve q_p (Dashed Line) and Regional Regression Estimate q_p' (Solid Line) at Sites 1–6

watershed model of ground-water outflow described by Brutsaert and Nieber (1977).

The model derived by Vogel et al. (1989) contains the independent basin parameters: watershed area A, average basin slope S, and the baseflow recession constant K_b, similar to a recent study by Zecharias and Brutsaert (1988). Although the baseflow recession constant explains a significant portion of the low-flow component of streamflow in Massachusetts (see Vogel et al. 1989) it is difficult to obtain estimates of K_b at an ungauged site, hence we ignore that parameter here. Instead, we attempt to describe the parameters $\hat{\mu}$ and $\hat{\sigma}$ in Eq. 7 using the basin parameters: watershed area A and average basin slope S.

Zecharias and Brutsaert (1985) showed that S is highly correlated with Strahlers' (1950) estimate of average basin slope, which is $S = Hd$, where

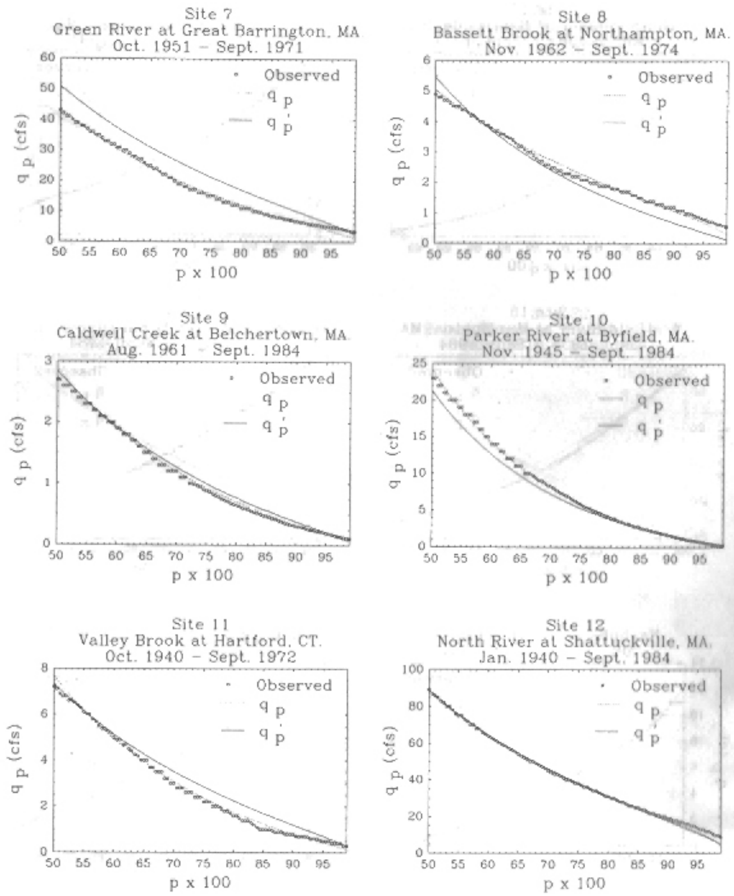

FIG. 2(*b*). Comparison of Observed Empirical Flow-Duration Curve (Open Circles) with Fitted Flow-Duration Curve q_p (Dashed Line) and Regional Regression Estimate q_p' (Solid Line) at Sites 7–12

H = the basin relief and d = the drainage density. Drainage density, d, is the ratio of the total length of stream channels in the basin, L, divided by the watershed area A. Basin relief, H, is simply a measure of the difference between the basin summit elevation and the channel outlet elevation.

Here, as in Zecharias and Brutsaert (1985), the basin summit elevation is estimated as the average of the elevations of the highest peak along the drainage divide and the two adjacent peaks on each side of it. In this study, we found that the basin parameters A and H alone, could be used to obtain regional regression equations for $\hat{\mu}$ and $\hat{\sigma}$. The drainage areas and relief values for the 23 basins are summarized in Table 1.

The regional regression equations for $\hat{\mu}$ and $\hat{\sigma}$ were obtained by applying ordinary least-squares regression procedures to the data in Table 1. For $\hat{\mu}$ the resulting regional regression is

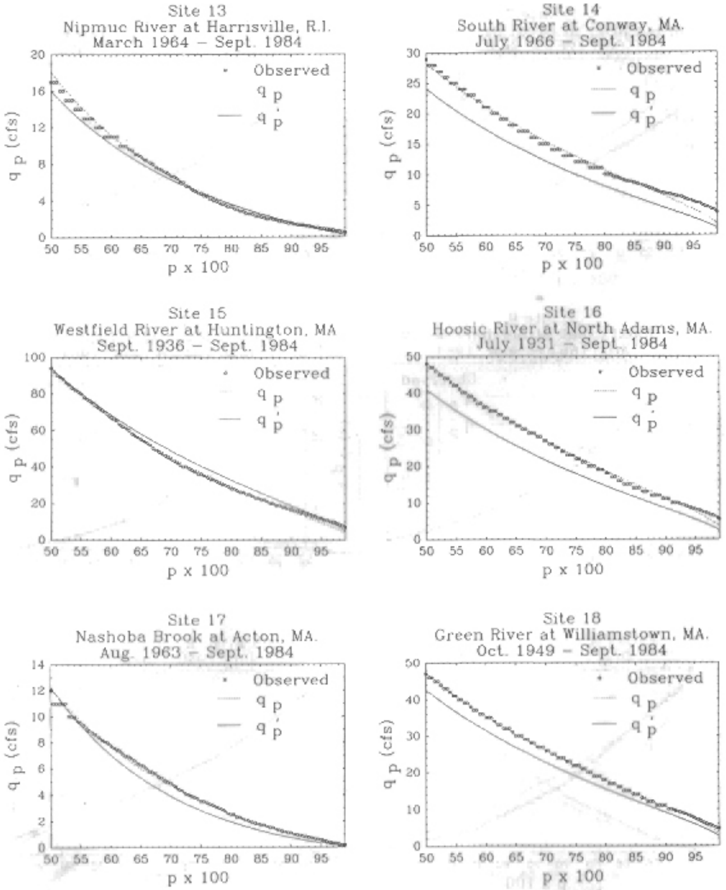

FIG. 2(*c*). Comparison of Observed Empirical Flow-Duration Curve (Open Circles) with Fitted Flow-Duration Curve q_p (Dashed Line) and Regional Regression Estimate q_p' (Solid Line) at Sites 13–18

$$\mu' = 1.0088 \ln (A) + \epsilon \quad (10)$$

Here the coefficient of determination $R^2 = 0.998$, $\sigma_\epsilon = 0.1343$, and the *t*-ratio of the slope term is 116.2. Since the slope term is nearly unity, we set $\mu' = \ln (A)$ in this study. Note that $\exp (\hat{\mu}) = q_{0.5}$, hence the median daily discharge in cubic feet per second (cfs) is approximately equal to 1 cfs/sq mi of drainage area in Massachusetts. Similarly, for $\hat{\sigma}$ the resulting regional regression is

$$\sigma' = 1.10 + \frac{271}{H} + \eta \quad (11)$$

Here the coefficient of determination $R^2 = 0.720$, $\sigma_\eta = 0.2013$, and the *t*-ratio of the intercept and slope terms are 15.67 and 7.35, respectively. The large *t*-ratios in Eqs. 10 and 11 are evidence of the high precision associated

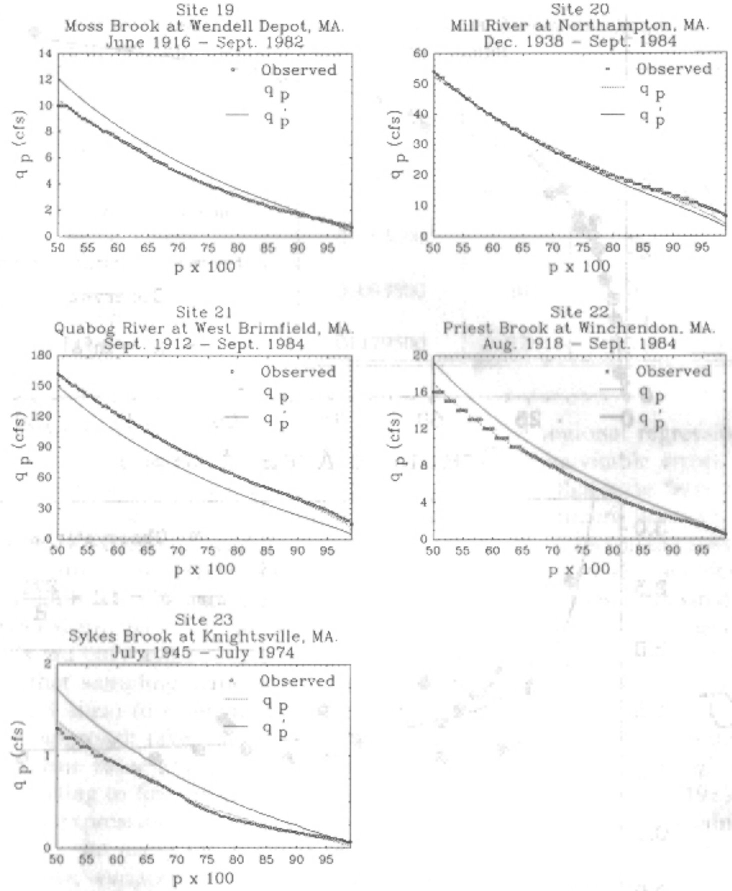

FIG. 2(*d*). Comparison of Observed Empirical Flow-Duration Curve (Open Circles) with Fitted Flow-Duration Curve q_p (Dashed Line) and Regional Regression Estimate q_p' (Solid Line) at Sites 19–23

with each of the model parameter estimates. Using probability plot correlation coefficient tests (Vogel and Kroll 1989) we could not reject the hypotheses that the residuals ϵ in Eq. 10, and η in Eq. 11, are normally distributed, at the 5% significance level. Perhaps the best summary of Eqs. 10 and 11 is provided in Fig. 3, where the regression estimates μ' and σ' (using solid lines) are compared to the data upon which those regressions are based (using darkened circles). Here the darkened circles are the optimal parameter estimates $\hat{\mu}$ and $\hat{\sigma}$. Fig. 3 indicates that drainage area is an excellent predictor of $\hat{\mu}$, yet relief is not nearly as good a predictor of $\hat{\sigma}$. Multivariate regression procedures were employed to test a variety of combinations of model forms and independent variable combinations, yet the simple linear regression equations in Eqs. 10 and 11 produced the best results. Using μ' and σ' in Eqs. 10 and 11, one may obtain a regional regression estimate of q_p at an ungauged site using

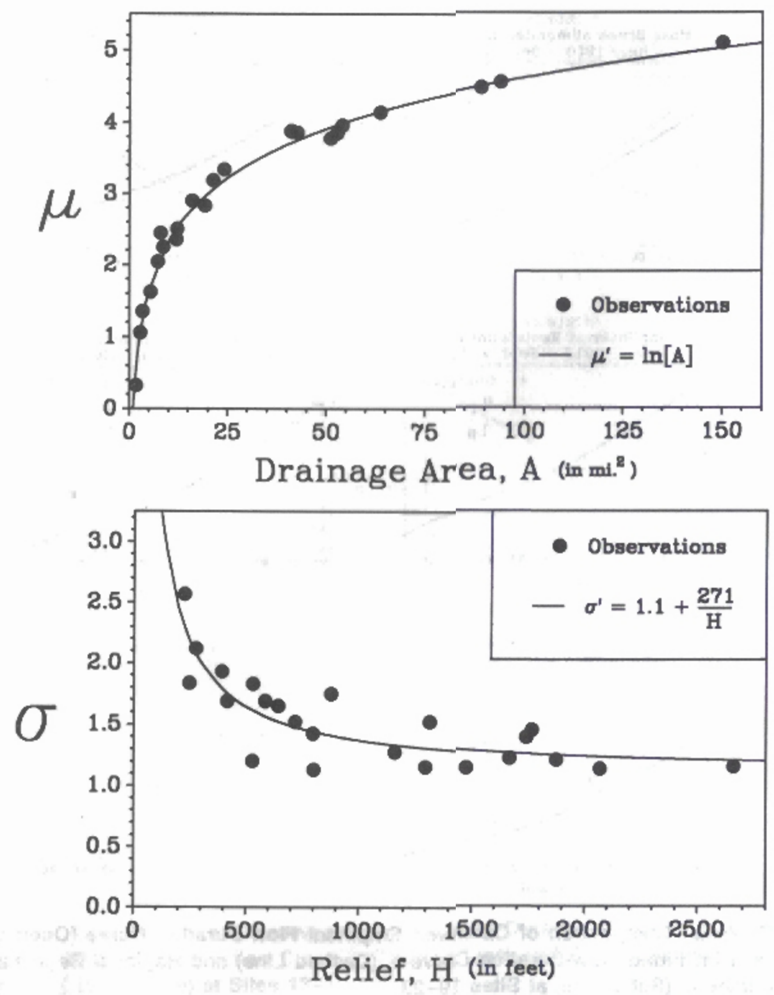

FIG. 3. Comparison of Fitted Observations $\hat{\mu}$ and $\hat{\sigma}$, with Regression Estimates μ' and σ'

$$q_p' = \exp(\mu' + z_p\sigma') \quad (12)$$

where z_p is given in Eq. 8. Figs. 2(*a*), (*b*), (*c*), and (*d*) compare the regression estimates q_p' (using solid lines), with the fitted quantiles q_p (using dotted lines) described earlier. In general, the agreement is quite good. At four sites q_p' is consistently less than q_p, and at five sites q_p' is consistently greater than q_p; otherwise, at the remaining fourteen sites, the agreement is about as good as one might hope for.

Confidence Intervals for Predictions at Ungauged Sites

The previous sections describe the agreement between an empirical flow-duration curve developed from a streamflow record, q_p, and the flow-duration curve estimated at an ungauged site, q_p'. The regional regression estimator q_p' contains substantial variability due to the inevitable errors associated with the regression models; the sampling errors that arise from fitting the models to short, cross-correlated (in space) and autocorrelated (in time) streamflow sequences; and finally, to the unavoidable errors associated with all streamflow measurements. In this section and the appendix we develop a procedure for estimating approximate confidence intervals associated with individual predictions q_p' obtained at an ungauged site. These confidence intervals are only approximate because they ignore measurement error and assume that sampling error is only attributed to the limited number of sites used (23 sites) to estimate the regional regressions.

TABLE 2. Basin Characteristics and U.S. Geological Survey Site Numbers of Three Sites Used to Validate Regional Flow-Duration Model

Site location (1)	U.S. Geologic Survey gage number (2)	Record length (years) (3)	Drainage area A (sq mi) (4)	Relief H (ft) (5)
East Branch Tully River near Athol, Massachusetts	01165000	32	50.5	627
Stony Brook near Temple, New Hampshire	01093800	20	3.6	996
Westfield River at Knightsville, Massachusetts	01179500	32	161.0	1,867

The approach taken here was to derive approximate confidence intervals for the true value of q_p, when one employs the regression estimator q_p', without resorting to first-order Taylor series approximations. Loaiciga (1989) derives an expression for the variance of an empirical quantile q_p, obtained at a gauged site using the Weibull plotting position given in Eq. 1. Here we require the variance of the regional regression quantile estimator q_p' to be employed at an ungauged site. We found that approximations to var (q_p') using a Taylor-series approximation to q_p' led to inconsistencies because the Taylor-series approximation did not capture the full nonlinear behavior of the combination of Eqs. 10, 11, and 12. Instead we recommend the following approximate $100(1-2\alpha)\%$ confidence interval for the true value of q_p

$$\exp\left[y_p' \mp t_{\nu,\alpha}\sqrt{\text{var}(y_p')}\right] \quad (13)$$

where $y_p' = \ln(q_p')$, $t_{\nu,\alpha}$ = a students' t random variate with ν degrees of freedom that is exceeded $100\alpha\%$ of the time and the variance of a predicted quantile, var (y_p'), is derived in the Appendix. Here $\nu = m - p$ where m = the number of sites used to estimate the regional regression equations in Eqs. 10 and 11 and p = the number of model parameter estimates in Eqs. 10 and 11, hence $\nu = 23 - 3 = 20$. For example, to construct 95% confidence intervals $t_{\nu,\alpha} = t_{20,0.975} = 2.086$.

The confidence intervals in Eq. 13 correspond to random intervals about the true value q_p, when one employs the estimator q_p' at an ungauged site. A test was performed to determine whether or not such randomly constructed 95% confidence intervals enclose the assumed true values of q_p in Figs. 2(*a*), (*b*), (*c*), and (*d*), denoted by the open circles. Those confidence intervals

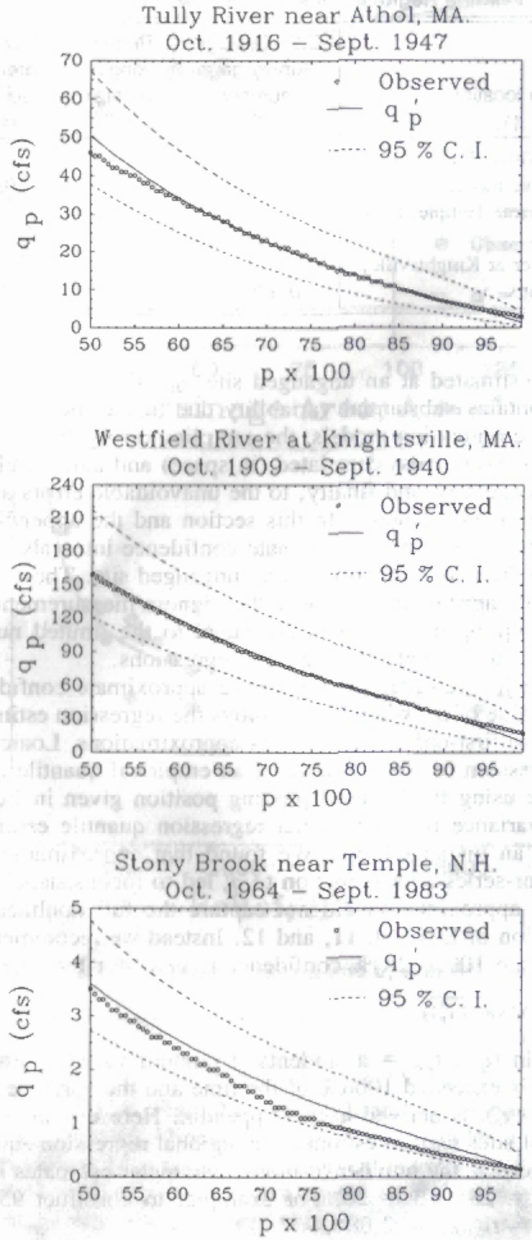

FIG. 4. Comparison of Observed Empirical Flow-Duration Curve (Open Circles) with Regression Estimate q_p' (Solid Line) at Three Validation Sites Described in Table 2

are not included in Figs. 2(*a*), (*b*), (*c*), and (*d*) because the figures would become too complex. As anticipated, only one site (site no. 23) contained values of q_p that lay just outside the 95% confidence intervals. Since such intervals only contain the true value q_p 95% of the time, one would expect to obtain one failure when examining 23 sites. This experiment serves as a check for us to assure ourselves that the derived confidence intervals are reasonable. In the next section, we display confidence intervals in the context of a validation study.

Validation Experiment

Three additional unregulated U.S. Geological Survey streamflow gaging stations were selected to validate the regional flow-duration model. The basin characteristics, record length, and U.S. Geological Survey site numbers for the validation stations are summarized in Table 2. These three sites have drainage areas that range from 3.6 to 161 sq mi and values of basin relief that range from 627 to 1,867 ft. These basins encompass the wide variability of drainage basin characteristics observed in the state of Massachusetts. Furthermore, each of the validation sites are located in different river basins, hence they capture a wide geographic region.

Fig. 4 depicts the agreement between the observed empirical flow-duration curve (using open circles) and the regional regression estimate q_p' (using solid lines). Overall, the agreement between the observed empirical flow-duration curve and the regional regression estimate q_p' is excellent. The constructed 95% confidence intervals, displayed using dashed lines, enclose both the observed quantile q_p and the regional regression estimate q_p' at all three sites.

Interpretation of Flow-Duration Curves and Low-Flow Statistics

Presently, the most widely used index of low flow in the United States is the seven-day, 10-year low flow, $Q_{7,10}$ (Riggs et al. 1980). The $Q_{7,10}$ is an annual-event based statistic where the event is defined as the annual minimum seven-day average daily discharge. The $Q_{7,10}$ is a discharge that is expected to be exceeded in nine out of 10 years. Studies that concentrate solely upon the impact of discharges into rivers during low-flow periods may benefit from focusing on a singular event such as the annual minimum seven-day low flow. For example, instream flow requirements for the maintenance of fish populations and other aquatic life are often defined in terms of a minimum tolerable streamflow. Similarly, wasteload allocations are usually made on the basis of a maximum allowable constituent concentration that corresponds to a minimum tolerable streamflow. The minimum tolerable streamflow corresponding to wasteload allocations, however, is usually different from, and typically lower than, the minimum tolerable streamflow corresponding to the maintenance of fish populations.

Unfortunately, no clear scientific basis exists for any of the commonly used annual-event based low-flow statistics (Male and Ogawa 1984). There is no well-defined annual low-flow event that has a precise scientific basis analogous to flood events in which the *T*-year annual maximum flood-flow corresponds to a flood event with a $1/T$ probability of occurrence in any given year. Nevertheless annual-event based low-flow statistics such as $Q_{7,10}$ are widely used and accepted, and perhaps as a result, hydrologists and water to solve a diverse class of water resource problems ranging from water supply, hydropower, and irrigation, to stream pollution studies. The use of flow-duration curves should be limited to problems in which the sequential nature of streamflows is unimportant. There exists a wide class of reservoir operations problems associated with flood control, water supply, irrigation, and hydropower for which the sequential nature of streamflows must be accounted for and hence, the use of flow-duration curves is no longer appropriate. Instead, computer simulation approaches (sequential routing procedures) have replaced the use of the flow-duration curve for problems where the sequential nature of streamflows is important. Nevertheless, with increasing attention focused on surface-water-quality management, there is a growing need for the development of methods that describe streamflow characteristics across a broad range of flow regimes. Flow-duration curves are ideally suited to such tasks because they describe the frequency and magnitude of streamflows over a broad range and they can be modified to evaluate the impact of streamflow regulation (stream withdrawals and stream discharges). For example, flow-duration curves are currently used in Massachusetts to evaluate the impact of proposed future withdrawals on the net basin yield and low-flow characteristics of a river basin.

Since most locations where flow-duration curves are required, are not coincident with stream gages, this study focuses on the development of a method for estimating a flow-duration curve at an ungauged site. Twenty-three unregulated gauged river basins are used to develop a regional flow-duration model in Massachusetts. The resulting model is easily implemented on a hand calculator and only requires estimates of the watershed area and basin relief associated with the ungauged site, both of which may be obtained from U.S. Geological Survey 7.5-min topographic quadrangle maps. A validation experiment, using three additional unregulated gauging stations, reveals that the derived model produces good estimates of observed flow-duration curves, especially considering its simplicity and ease of application. Finally, a method is derived for obtaining approximate confidence intervals associated with a flow-duration curve estimated at an ungauged site.

The regional flow-duration model developed here should only be used in Massachusetts for ungauged drainage basins with watershed areas in the approximate range of 1.73 to 150 sq mi, and reliefs in the approximate range of 227–2,658 ft. Hopefully future studies will extend the regional flow-duration curves developed here to other regions so that the model's range of applicability can be defined more precisely.

Acknowledgments

This research was supported by a cooperative agreement between Tufts University and the U.S. Geological Survey with matching funds from the Massachusetts Division of Water Pollution Control in the Department of Environmental Protection. We are grateful for the assistance provided by S. William Wandle, Jr., and Katherine D. Martel. In addition, we appreciate the helpful comments of Dara Entekhabi, Monique Villars, Kernell G. Ries, Richard Fontaine, and two anonymous reviewers.

Appendix I. Derivation of Confidence Intervals for Prediction q_p'

In order to construct confidence intervals for a predicted value q_p', Eq. 13

requires an estimate of the variance of the regression estimator y_p', where $y_p' = \ln(q_p')$, hence Eq. 12 may be rewritten as

$$y_p' = \mu' + z_p\sigma' \quad (14)$$

If the estimators μ' and σ' are considered to be independent then

$$\text{var}(y_p') = \text{var}(\mu') + z_p^2 \text{ var}(\sigma') \quad (15)$$

where the variance associated with a prediction, μ', at an ungauged site, using Eq. 10, is

$$\text{var}(\mu') = \text{var}[b_1 \ln(A) + \epsilon] \quad (16a)$$

$$\text{var}(\mu') = \ln^2(A) \text{ var}(b_1) + \sigma_\epsilon^2 \quad (16b)$$

$$\text{var}(\mu') = [0.00868 \ln(A)]^2 + 0.018 \quad (16c)$$

Apparently, the variance associated with μ' increases with drainage area; however, even for the largest site considered in this study (A = 150 sq mi), the contribution of the first term in Eqs. 16*a*, *b*, and *c* is an order of magnitude lower than the contribution due to the model error term σ_ϵ^2. This is largely due to the high precision associated with the single model parameter b_1 as evidenced by its t-ratio of 116.2.

Similarly, the variance associated with a prediction, σ', at an ungauged site, using Eq. 11, is

$$\text{var}(\sigma') = \text{var}(a_2 + b_2X_2 + \eta) \quad (17)$$

with $a_2 = 1.1$, $b_2 = 271$, and $X_2 = 1/H$. The ordinary least-squares estimator for a_2 is

$$a_2 = \bar{\sigma} - b_2\bar{X}_2 \quad (18)$$

which can be combined with Eq. 17 to yield

$$\text{var}(\sigma') = \text{var}(\bar{\sigma} + b_2(X_2 - \bar{X}_2) + \eta)$$

$$\text{var}(\sigma') = \frac{\sigma_\eta^2}{m} + (X_2 - \bar{X}_2)^2 \text{ var}(b_2) + \sigma_\eta^2 \quad (19)$$

Since $\sigma_\eta = 0.2013$, $m = 23$, $\bar{X}_2 = 0.00153$, and $\text{var}(b_2) = 36.93^2$, Eq. 19 can be simplified to

$$\text{var}(\sigma') = 0.0423 + \left[\left(\frac{36.93}{H}\right) - 0.0565\right]^2 \quad (20)$$

From Eq. 20 we observe that var (σ') is equal to its minimum value of 0.0423 at H = 654 ft. Similarly var (σ') reaches its maximum values of 0.044 and 0.0536 for the largest (H = 2,658 ft) and smallest (H = 227 ft) values of relief considered here, respectively. As with the variance of μ', the variance of σ' is primarily due to the model error variance σ_η^2. Substitution of Eqs. 15, 16*a*, *b*, and *c*, and 20 into Eq. 13 yields the required confidence interval for a predicted value of q_p at an ungauged site.

Appendix II. References

Abramowitz, M., and Stegun, I. A. (1964). *The handbook of mathematical functions*. Dover Press, New York, N.Y.

Beard, L. R. (1943). "Statistical analysis in hydrology. *Trans.*, ASCE, 108, 1110–1160.
Brutsaert, W., and Nieber, J. L. (1977). "Regionalized drought flow hydrographs from a mature glaciated plateau." *Water Resour. Res.*, 13(3), 637–643.
Chow, V. T. (1964). *Handbook of applied hydrology*. McGraw Hill Book Co., New York, N.Y.
Cunnane, C. (1978). "Unbiased plotting positions—a review." *J. Hydrol.*, 37, 205–222.
Dingman, S. L. (1978). "Synthesis of flow-duration curves for unregulated streams in New Hampshire." *Water Resour. Bull.*, 14(6), 1481–1502.
Entekhabi, D., Rodriguez-Iturbe, I., and Eagleson, P. S. (1989). "Probabilistic representation of the temporal rainfall process by a modified Neyman-Scott rectangular pulses model: Parameter estimation and validation." *Water Resour. Res.*, 25(2), 295–302.
Foster, H. A. (1934). "Duration curves." *Trans.*, ASCE, 99, 1213–1267.
Gupta, R. S. (1989). *Hydrology and hydraulic systems*. Prentice-Hall, Inc., N.J., 363–367.
Johnson, C. G. (1970). *A proposed streamflow data program for central New England*. U.S. Geological Survey, Boston, Mass.
Linsley, R. K., and Franzini, J. B. (1979). *Water-resources engineering*. McGraw Hill Book Co., New York, N.Y., 127–129.
Loaiciga, H. A. (1989). "Variability of empirical flow quantiles." *J. Hydr. Engrg.*, ASCE, 115(1), 82–100.
Male, J. W., and Ogawa, H. (1982). "Low flows of Massachusetts streams." *Publication No. 125*, Water Resour. Res. Center, Univ. of Massachusetts, Amherst, Mass.
Male, J. W., and Ogawa, H. (1984). "Tradeoffs in water quality management." *J. Water Resour. Plng. and Mgmt.*, ASCE, 110(4), 434–444.
Mimikou, M., and Kaemaki, S. (1985). "Regionalization of flow duration characteristics." *J. Hydrol.*, 82, 77–91.
Quimpo, R. G., Alejandrino, A. A., and McNally, T. A. (1983). "Regionalized flow duration curves for Philippines." *J. Water Res. Plng. and Mgmt.*, ASCE, 109(4), 320–330.
Riggs, H. C. (1972). "Low-flow investigations." *U.S. Geological Survey Techniques of Water Resources Investigations*, Book 4, U.S. Geological Survey, Washington, D.C.
Riggs, H. C., et al. (1980). "Characteristics of low flows." *J. Hydr. Engrg.*, ASCE, 106(5), 717–731.
Searcy, J. K. (1959). "Flow-duration curves." *Paper 1542-A*, U.S. Geological Survey Water-Supply, U.S. Government Printing Office, Washington, D.C.
Singh, K. P. (1971). "Model flow duration and streamflow variability." *Water Resour. Res.*, 7(4), 1031–1036.
Stedinger, J. R. (1980). "Fitting log normal distributions." *Water Resour. Res.*, 16(3), 481–490.
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.
Strahler, A. N. (1950). "Equilibrium theory of erosional slopes approached by frequency distribution analysis." *Am. J. Sci.*, 248, 673–696, 800–814.
Tasker, G. D. (1972). "Estimating low-flow characteristics of streams in southeastern Massachusetts from maps of groundwater availability." *Professional Paper 800-D*, U.S. Geological Survey, Washington, D.C., D217–D220.
Tasker, G. D. (1983). "Effective record length for the *T*-year event." *J. Hydrol.*, 64, 39–47.
Tukey, J. W. (1960). "The practical relationship between the common transformation of percentages or fractions and of amounts." *Technical Report No. 36*, Statistical Res. Group, Princeton, N.J.
"Users manual—math/library." (1987). Version 1.0, Int. Mathematical Subroutine Library, Houston, Tex.

Vogel, R. M., and Kroll, C. N. (1989). "Low-flow frequency analysis using probability-plot correlation coefficients." *J. Water Res. Plng. and Mgmt.*, ASCE, 115(3), 338–357.
Vogel, R. M., and Kroll, C. N. (1990). "Generalized low-flow frequency relationships for ungaged sites." *Water Resour. Bull.*
Vogel, R. M., Kroll, C. N., and Driscoll, K. (1989). "Regional geohydrologic-geomorphic relationships for the estimation of low-flows." *Proc. Int. Conf. on Channel Flow and Catchment Runoff*, Univ. of Virginia, 267–277.
Warnick, C. C. (1984). *Hydropower engineering*. Prentice-Hall, Inc., N.J., 59–73.
Zecharias, Y. B., and Brutsaert, W. (1985). "Ground surface slope as a basin scale parameter." *Water Resour. Res.*, 21(12), 1895–1902.
Zecharias, Y. B., and Brutsaert, W. (1988). "Recession characteristics of groundwater outflow and baseflow from mountainous watersheds." *Water Resour. Res.*, 24(10), 1651–1658.

APPENDIX III. NOTATION

The following symbols are used in this paper:

A = drainage area (sq mi);
H = basin relief (ft);
m = number of sites used to estimate regional regression equations (m = 23);
mgd = million gallons per day;
p = exceedance probability for average daily streamflow;
p_i = plotting position estimate of exceedance probability associated with ordered observation $q_{(i)}$;
q_i = observed mean streamflow on day i (cfs);
$q_{(i)}$ = ith smallest observed streamflow (cfs);
q_p = average daily streamflow with exceedance probability p (cfs);
q_p' = regression estimate of q_p at ungauged site (cfs);
R^2 = coefficient of determination;
$t_{\nu,\alpha}$ = students' t random variate with ν degrees of freedom, exceeded 100α% of time;
z_p = standard normal variate exceeded 100p% of time;
ϵ = residual error in regression model;
η = residual error in regression model;
μ = true value of mean of logarithms of average daily streamflow;
$\hat{\mu}$ = optimal estimator of μ using streamflow data;
μ' = regional regression estimate of μ;
σ = true value of standard deviation of logarithms of average daily streamflow;
$\hat{\sigma}$ = optimal estimator of σ using streamflow data; and
σ' = regional regression estimate of σ.