WATER RESOURCES BULLETIN
VOL. 26, NO. 2 AMERICAN WATER RESOURCES ASSOCIATION APRIL 1990

GENERALIZED LOW-FLOW FREQUENCY RELATIONSHIPS FOR UNGAGED SITES IN MASSACHUSETTS[1]

Richard M. Vogel and Charles N. Kroll[2]

ABSTRACT: Regional hydrologic procedures such as generalized least squares regression and streamflow record augmentation have been advocated for obtaining estimates of both flood-flow and low-flow statistics at ungaged sites. While such procedures are extremely useful in regional flood-flow studies, no evaluation of their merit in regional low-flow estimation has been made using actual streamflow data. This study develops generalized regional regression equations for estimating the d-day, T-year low-flow discharge, $Q_{d,T}$, at ungaged sites in Massachusetts where d = 3, 7, 14, and 30 days. A two-parameter lognormal distribution is fit to sequences of annual minimum d-day low-flows and the estimated parameters of the lognormal distribution are then related to two drainage basin characteristics: drainage area and relief. The resulting models are general, simple to use, and about as precise as most previous models that only provide estimates of a single statistic such as $Q_{7,10}$. Comparisons are provided of the impact of using ordinary least squares regression, generalized least squares regression, and streamflow record augmentation procedures to fit regional low-flow frequency models in Massachusetts.
(KEY TERMS: low-flows; regional regression; low-flow frequency analysis; water quality management; Massachusetts; low-flow management; regional hydrology; generalized least squares regression; streamflow record augmentation; droughts.)

INTRODUCTION

The Massachusetts Department of Environmental Protection is currently making hundreds of estimates of low-flow statistics each year at ungaged sites for the purposes of determining waste-load allocations, issuing and/or renewing National Pollution Discharge Elimination System (NPDES) permits, siting treatment plants and sanitary landfills, and for decisions regarding interbasin transfers of water and allowable basin withdrawals. In addition, low-flow statistics are used widely in Massachusetts, and elsewhere, to determine minimum downstream release requirements from hydropower, irrigation, water-supply and cooling-plant facilities. Unfortunately, most locations where such low-flow statistics are required are not coincident with streamflow gaging stations; therefore, accurate estimation of low-flow statistics at ungaged sites is a fundamental and widespread problem.

The most widely used index of low-flow in the United States is the seven-day, ten-year low-flow ($Q_{7,10}$), defined as the annual minimum seven-day average daily discharge that is expected to be exceeded in nine out of every ten years (Riggs *et al.*, 1980). Since water resource applications often require knowledge of the magnitude of annual minimum low-flows for recurrence intervals (T) other than 10 years, and for durations (d) other than 7 days, this study takes a more general approach by specifying the general d-day, T-year event $Q_{d,T}$.

STUDY REGION

The regional regression equations developed in this study are based on data obtained from twenty-three U.S. Geological Survey streamflow-gaging stations in or near Massachusetts. Table 1 provides the U.S. Geological Survey streamflow-gaging station numbers, record lengths, and selected basin characteristics. In addition, site numbers used in this and other studies (Vogel and Kroll, 1989, 1990; Vogel *et al.*, 1989, Fennessey and Vogel, 1990) are provided in Table 1. The location of those stations is shown in Figure 1.

All of the rivers are perennial; that is, all streamflows are greater than zero. No significant withdrawals, diversions, or artificial recharge areas are within any of the basins; hence, the streamflows are considered to be unregulated (natural). The twenty-three gaging stations employed here are about the only long-term, unregulated, continuous flow records available in or near Massachusetts.

[1]Paper No. 89079 of the *Water Resources Bulletin*. **Discussions are open until December 1, 1990.**
[2]Respectively, Assistant Professor, Department of Civil Engineering, Tufts University, Medford, Massachusetts 02155; and Staff Engineer, Goldberg Zoino & Associates, Inc., 320 Needham St., Newton Upper Falls, Massachusetts 02164.

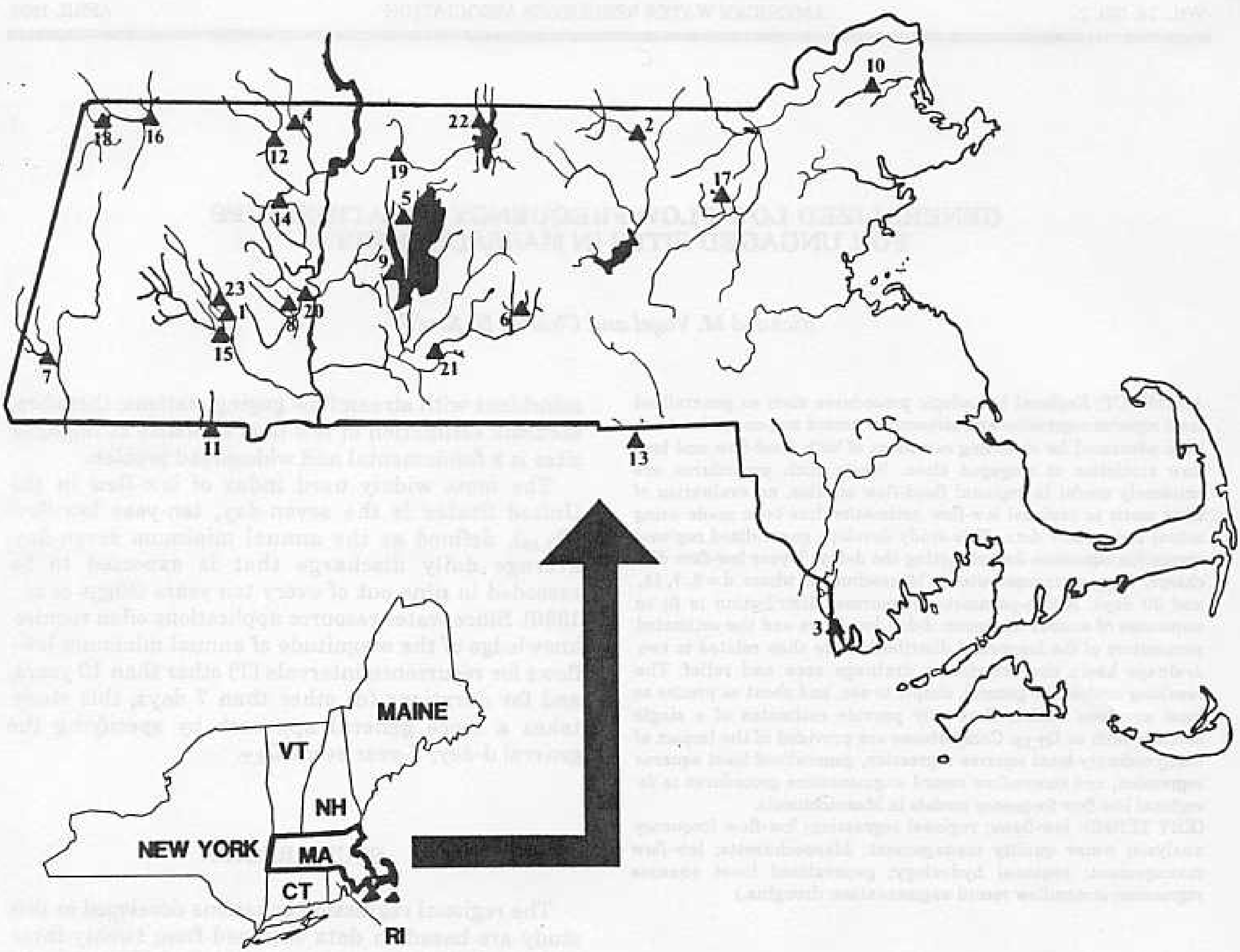

Figure 1. Location of the Twenty-Three U.S. Geological Survey Gaging Stations in or near Massachusetts Described in Table 1.

A REGIONAL PROBABILITY DISTRIBUTION FOR ANNUAL MINIMUM d-DAY AVERAGE DAILY STREAMFLOWS

The development of generalized regional regression equations for estimating $Q_{d,T}$ requires the selection of a probability distribution for modeling the frequency and magnitude of annual minimum d-day streamflow series in a region. One objective of this study is to relate easily measured drainage basin characteristics to parameters which summarize the probability distribution of the low-flow series. To obtain such relations we require that the regional probability density function (pdf) of annual minimum d-day low-flows be described by as few parameters as possible. Therefore a two-parameter pdf was employed.

Using the twenty-three sites listed in Table 1, Vogel and Kroll (1989) found that there is little evidence to contradict the hypothesis that annual minimum seven-day low-flows in Massachusetts arise from a two-parameter lognormal pdf. Vogel and Kroll (1990) generalize their results further by showing that sequences of 3- , 5- , 7- , 14- , and 30-day annual minimum low-flows in Massachusetts are extremely well approximated by a two-parameter lognormal distribution. Essentially these studies employed probability plot correlation coefficient test statistics to compare annual minimum flow series fit to 2- and 3-parameter lognormal, 2- and 3-parameter Weibull, and the log Pearson Type III distributions. On the basis of the previous studies, it was assumed in this study that all flow series considered here can be described by a two-parameter lognormal distribution.

TABLE 1. U.S. Geological Survey Gaging Station Numbers and Selected Basin Characteristics.

USGS Gage No.	Site No.*	Record Length (yrs)	Drainage Area A (mi²)	Basin Relief H (ft)
01180500	1	73	52.70	1765
01096000	2	34	63.69	1161
01106000	3	37	8.01	240
01170100	4	16	41.39	1873
01174000	5	34	3.39	531
01175670	6	23	8.68	417
01198000	7	19	51.00	1317
01171800	8	11	5.46	530
01174900	9	22	2.85	585
01101000	10	38	21.30	277
01187400	11	31	7.35	877
01169000	12	44	89.00	1667
01111300	13	20	16.02	393
01169900	14	17	24.09	1298
01181000	15	48	94.00	1739
01332000	16	52	40.90	2068
01097300	17	20	12.31	248
01333000	18	34	42.60	2658
01165500	19	65	12.10	797
01171500	20	45	54.00	1476
01176000	21	71	150.00	801
01162500	22	63	19.30	718
01180000	23	28	1.73	643

*See Figure 1 for the location of each site.

Maximum likelihood estimates of the low-flow statistics at each of the gaged sites are obtained from

$$Q_{d,T} = \exp(\hat{\mu}_y(d) + z_T\,\hat{\sigma}_y(d)) \qquad (1)$$

where

$$\hat{\mu}_y(d) = \frac{1}{n}\sum_{i=1}^{n} y_i(d)$$

$$\hat{\sigma}_y^2(d) = \frac{1}{n-1}\sum_{i=1}^{n}(y_i(d) - \hat{\mu}_y(d))^2$$

$$y_i(d) = \ln[q_i(d)]$$

$q_i(d)$ is the minimum d-day average discharge in year i, n is the number of years of record, and z_T is a standard normal random variable corresponding to the T-year event. In a comparative study of alternative estimators of quantiles of a two-parameter lognormal distribution, Stedinger (1980) advocated the use of the maximum likelihood estimators shown above when one's primary interest is in estimating the lower quantile of the distribution. An excellent approximation to z_T is obtained using a formula proposed by Tukey (1960)

$$z_T = 4.91\left[\left(\frac{1}{T}\right)^{0.14} - \left(1-\frac{1}{T}\right)^{0.14}\right] \qquad (2)$$

as long as $1.01 \le T \le 100$ years.

Figure 2 displays the fit of a two-parameter lognormal distribution to the twenty-three series of annual minimum 7-day low-flows shown in Table 1 using computer generated lognormal probability paper. The sites are ordered by probability plot correlation coefficient test statistics, i.e., site 1 displays the worst fit and site 23 displays the best fit to a two-parameter lognormal distribution. The empirical probability plot is shown using open circles and the fitted lognormal distribution is the solid line. The fitted two-parameter lognormal distribution provides an excellent representation of each empirical cumulative density function.

Each empirical probability plot is obtained by plotting the ordered observations $q_{(i)}$ versus Blom's plotting position (Blom, 1958)

$$P\{Q \le q_i\} = \frac{i - 3/8}{n + 1/4} \qquad (3)$$

on computer generated lognormal probability plotting paper. Blom's plotting position is unbiased when observations arise from a normal distribution, which is the case for the log transformed streamflows y_i. The subscript i corresponds to the rank order for each streamflow, with i=1 and i=n corresponding to the smallest and largest values, respectively. Vogel and Kroll (1989) described procedures for generating probability plots on a computer.

REGIONAL LOW-FLOW FREQUENCY MODELS IN MASSACHUSETTS

Many investigators have developed regional models for estimating low-flow statistics at ungaged sites from readily available geomorphic, geologic, climatic, and topographic characteristics (see, for example, Thomas and Benson, 1970). Such models have been developed for use at ungaged sites for many regions in the United States. Usually, these models take the form

$$Q_{d,T} = b_0 X_1^{b_1} X_2^{b_2} X_3^{b_3} \ldots \qquad (4)$$

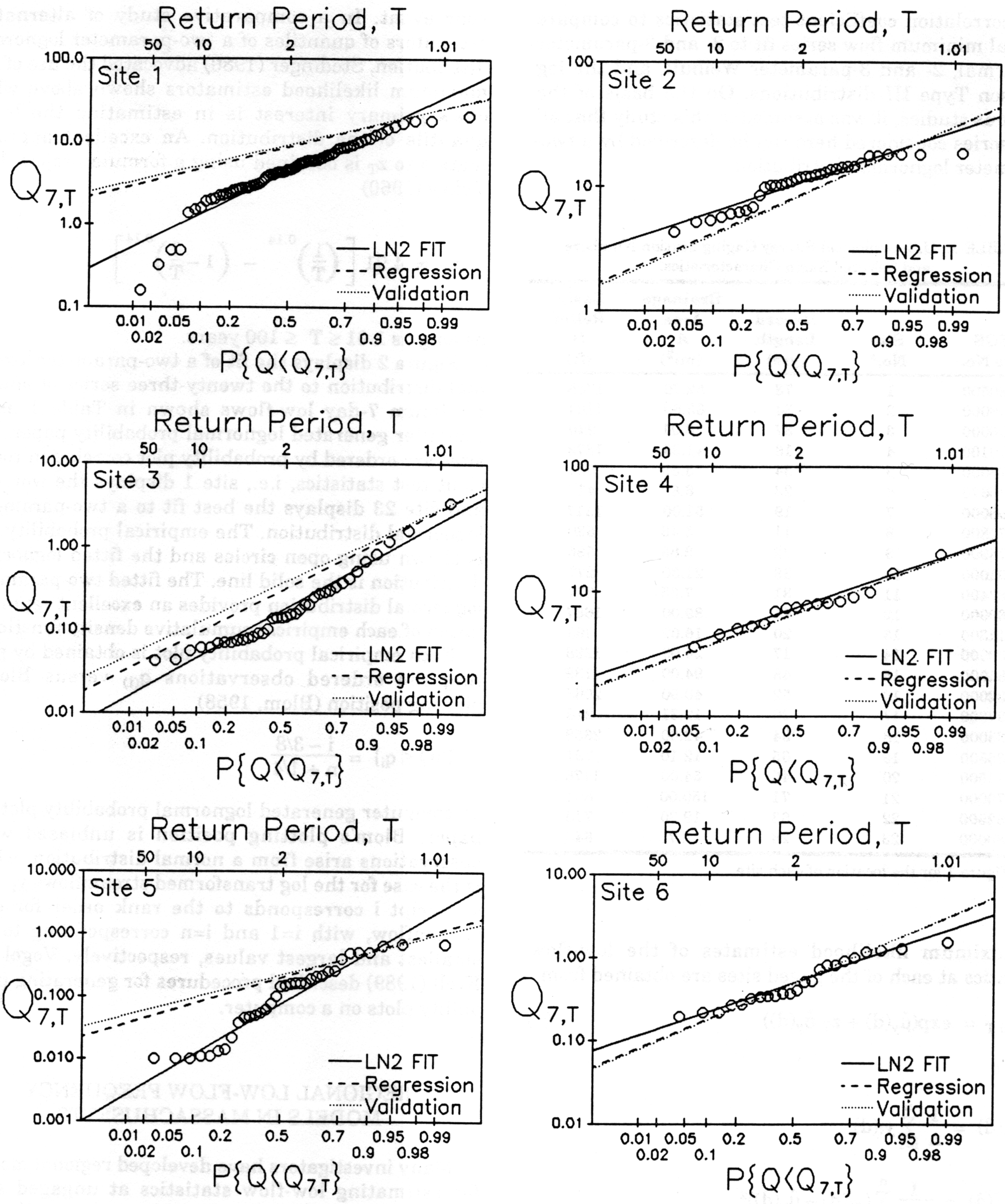

Figure 2. Comparison of the Observed Empirical Cumulative Distribution of Annual Minimum 7-Day Low-Flows (Open Circles) with the Fitted Lognormal Distribution (Solid Line), the Regression Estimate $Q'_{7,T}$, (Dashed Line) and the Validation Regression Estimate (Dotted Line).

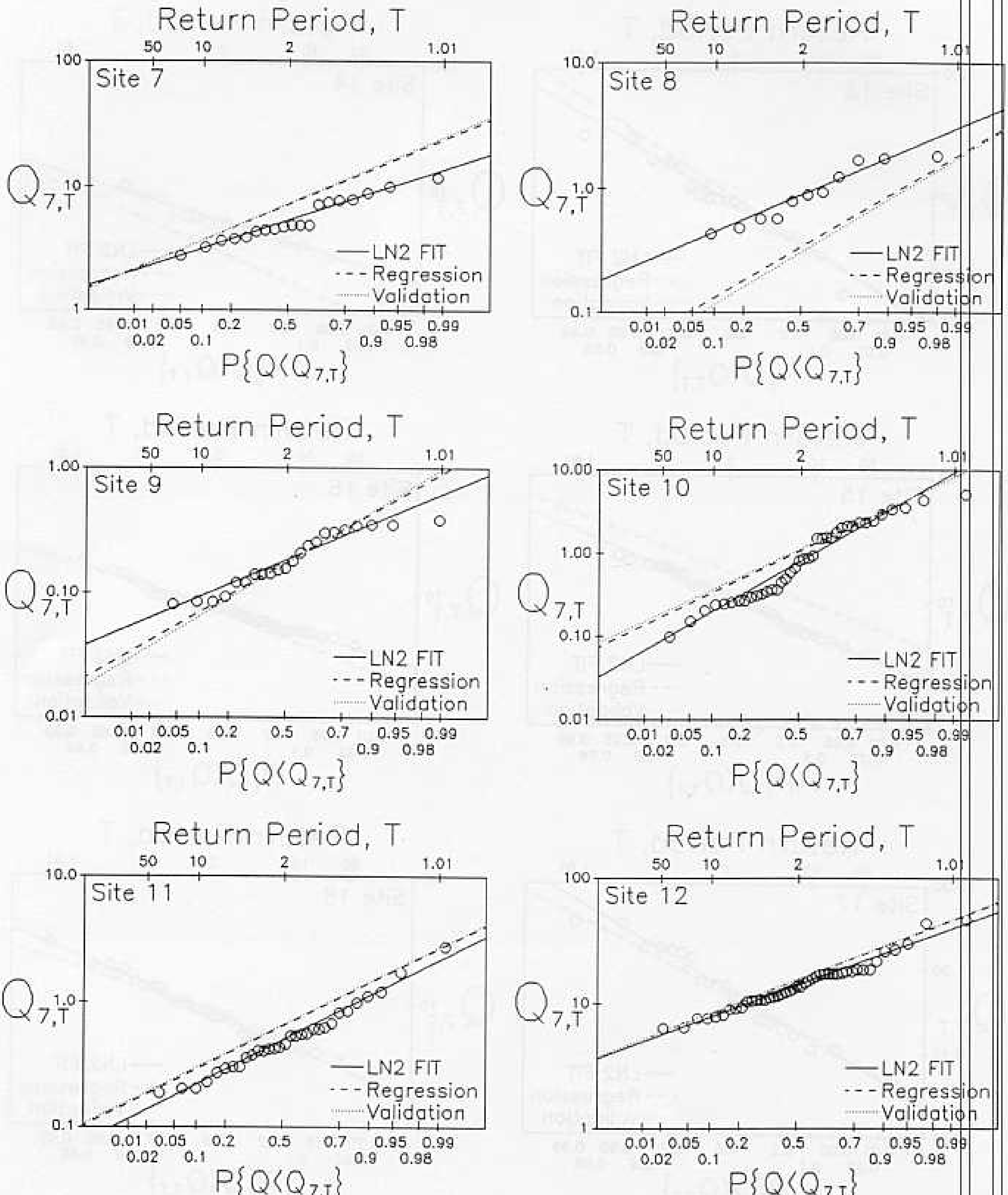

Figure 2. Comparison of the Observed Empirical Cumulative Distribution of Annual Minimum 7-Day Low-Flows (Open Circles) with the Fitted Lognormal Distribution (Solid Line), the Regression Estimate $Q'_{7,T}$, (Dashed Line) and the Validation Regression Estimate (Dotted Line). (cont'd.)

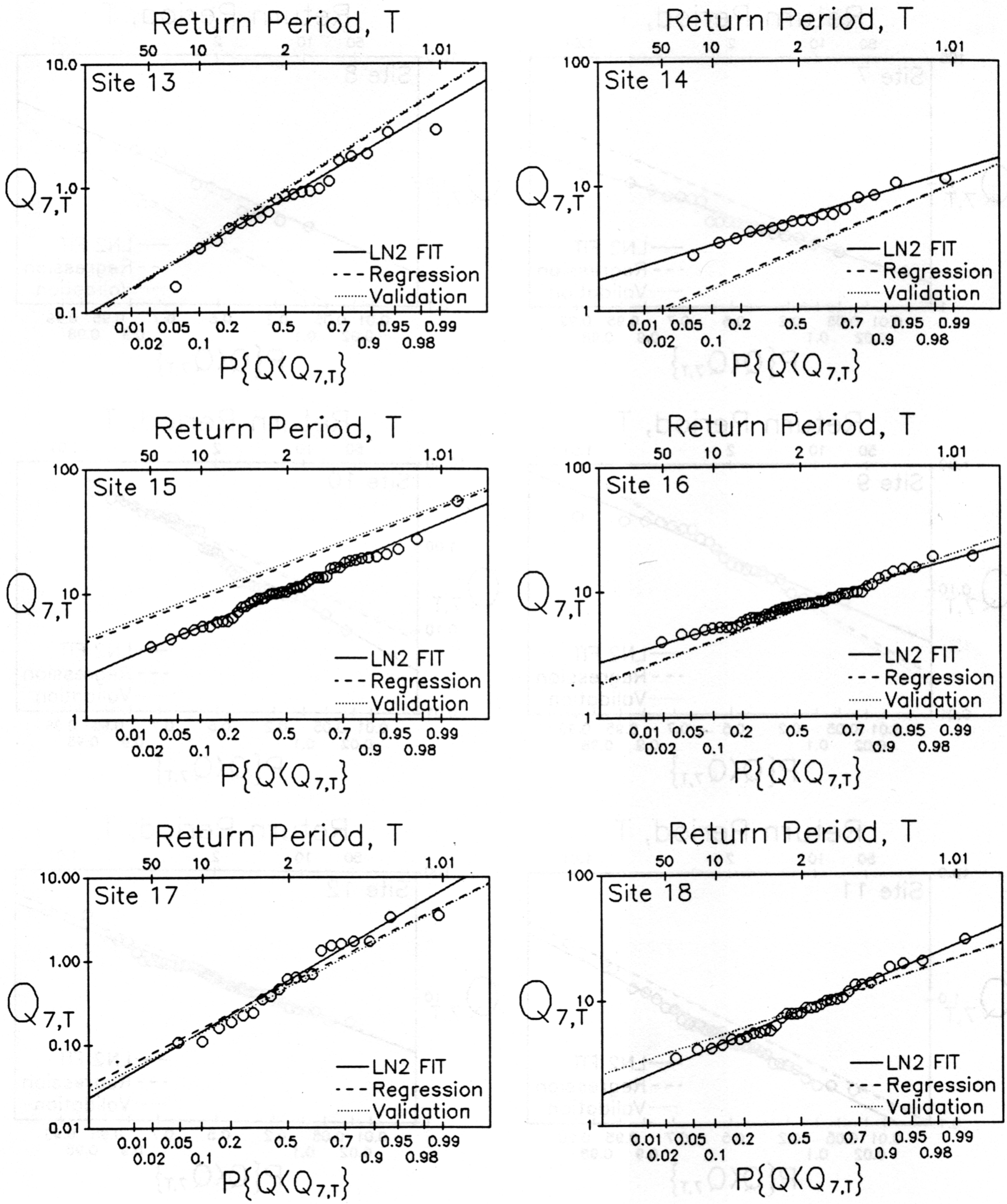

Figure 2. Comparison of the Observed Empirical Cumulative Distribution of Annual Minimum 7-Day Low-Flows (Open Circles) with the Fitted Lognormal Distribution (Solid Line), the Regression Estimate $Q'_{7,T}$, (Dashed Line) and the Validation Regression Estimate (Dotted Line).
(cont'd.)

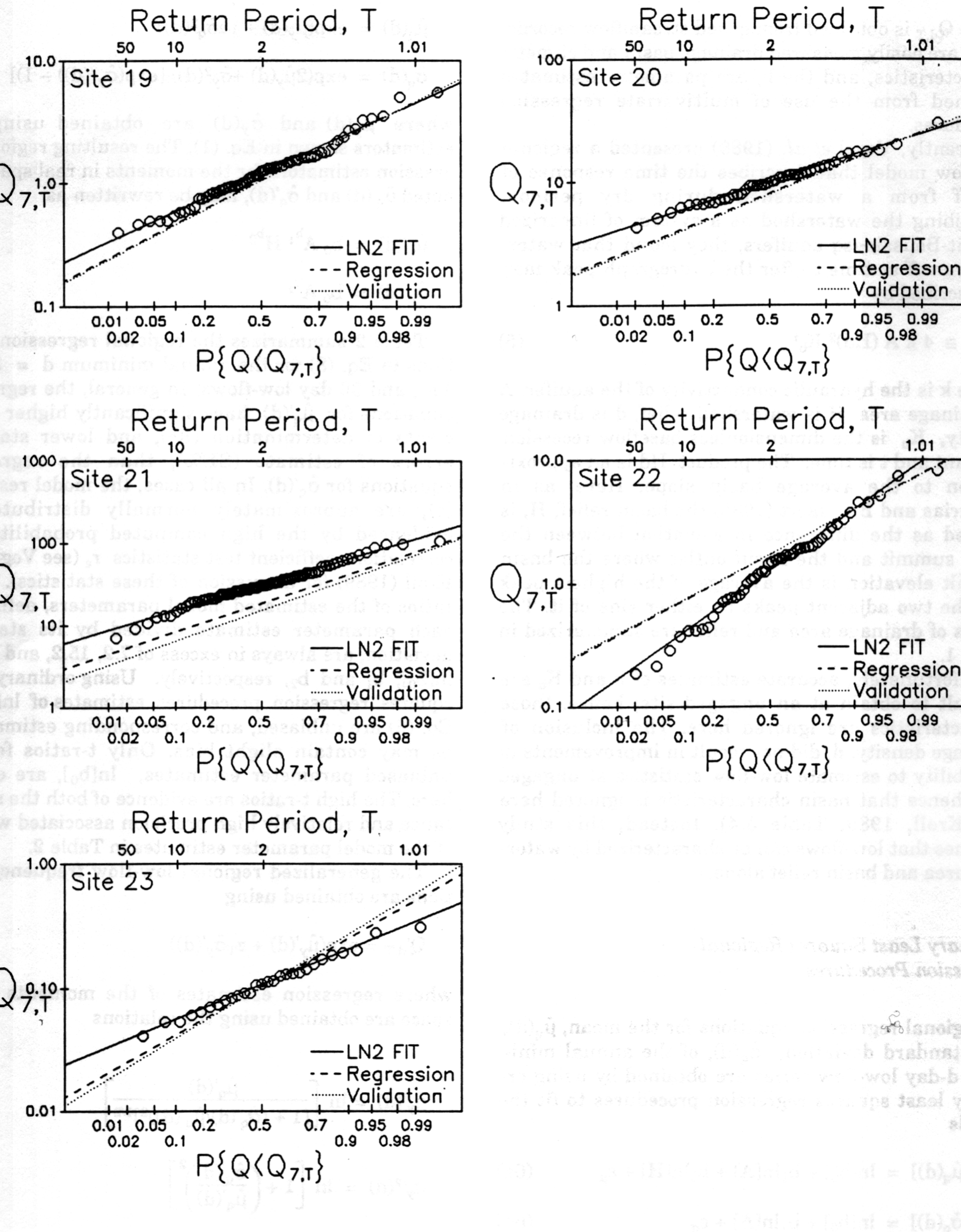

Figure 2. Comparison of the Observed Empirical Cumulative Distribution of Annual Minimum 7-Day Low-Flows (Open Circles) with the Fitted Lognormal Distribution (Solid Line), the Regression Estimate $Q'_{7,T}$, (Dashed Line) and the Validation Regression Estimate (Dotted Line).
(cont'd.)

where $Q_{d,T}$ is obtained from gaged streamflow records, the X_i are easily measured drainage basin and climatic characteristics, and the b_i are parameter estimates obtained from the use of multivariate regression procedures.

Recently, Vogel *et al.* (1989) presented a regional low-flow model that describes the time response of runoff from a watershed during dry periods. Describing the watershed as a system of linearized Dupuit-Boussinesq aquifers, they found that watershed runoff at time t after the hydrograph peak may be described by

$$Q_t \cong 4\,k\,A\,(Hd)^2\,K_b^{\,t} \tag{5}$$

where k is the hydraulic conductivity of the aquifer, A is drainage area, H is watershed relief, d is drainage density, K_b is the dimensionless baseflow recession constant and t is time. The product Hd is an approximation to the average basin slope. Here, as in Zecharias and Brutsaert (1985) the basin relief, H, is defined as the difference in elevation between the basin summit and the basin outlet where the basin summit elevation is the average of the highest peak and the two adjacent peaks on either side of it. The values of drainage area and relief are summarized in Table 1.

Unfortunately, accurate estimates of k and K_b are difficult to obtain at an ungaged site; hence, those characteristics are ignored here. The inclusion of drainage density, d, did not result in improvements in our ability to estimate low-flow statistics at ungaged sites hence that basin characteristic is ignored here (see Kroll, 1989, Table 5.4). Instead, this study assumes that low-flows can be characterized by watershed area and basin relief alone.

Ordinary Least Squares Regional Regression Procedures

Regional regression equations for the mean, $\hat{\mu}_q(d)$, and standard deviation, $\hat{\sigma}_q(d)$, of the annual minimum d-day low-flow series are obtained by using ordinary least squares regression procedures to fit the models

$$\ln[\hat{\mu}_q(d)] = \ln[b_0] + b_1\ln[A] + b_2\ln[H] + \varepsilon_\mu \tag{6a}$$

$$\ln[\hat{\sigma}_q(d)] = \ln[b_0] + b_1\ln[A] + \varepsilon_\sigma \tag{6b}$$

Estimates of the moments of the annual minimum d-day streamflows in real space are obtained from the moments in log space using the relations

$$\hat{\mu}_q(d) = \exp(\hat{\mu}_y(d) + 1/2\,\hat{\sigma}_q^2(d)) \tag{7a}$$

$$\hat{\sigma}_q(d) = \exp(2\hat{\mu}_y(d) + \hat{\sigma}_y^2(d))\,[\exp(\hat{\sigma}_y^2(d)) - 1)] \tag{7b}$$

where $\hat{\mu}_y(d)$ and $\hat{\sigma}_y(d)$ are obtained using the estimators shown in Eq. (1). The resulting regional regression estimators for the moments in real space, denoted $\hat{\mu}_q'(d)$ and $\hat{\sigma}_q'(d)$, may be rewritten as

$$\hat{\mu}_q'(d) = b_0\,A^{b_1}\,H^{b_2} \tag{8a}$$

$$\hat{\sigma}_q'(d) = b_0\,A^{b_1} \tag{8b}$$

Table 2 summarizes the regional regression equations in Eq. (8) for the annual minimum d = 3-, 7-, 14-, and 30-day low-flows. In general, the regression equations for $\hat{\mu}_q'(d)$ have significantly higher coefficients of determination (R^2), and lower standard errors of estimate (SE%), than the regression equations for $\hat{\sigma}_q'(d)$. In all cases, the model residuals (ε), are approximately normally distributed as evidenced by the high computed probability plot correlation coefficient test statistics r_ε (see Vogel and Kroll (1989) for discussion of these statistics). The t-ratios of the estimated model parameters, defined as each parameter estimate divided by its standard deviation, are always in excess of 7.2, 15.2, and 3.5 for $\ln[b_0]$, b_1 and b_2, respectively. Using ordinary least squares regression procedures, estimates of $\ln[b_0]$ in Eq. (6) are unbiased, and corresponding estimates of b_0 may contain slight bias. Only t-ratios for the unbiased parameter estimates, $\ln[b_0]$, are quoted here. The high t-ratios are evidence of both the significance and relatively high precision associated with all of the model parameter estimates in Table 2.

The generalized regional low-flow frequency relations are obtained using

$$Q'_{d,T} = \exp(\hat{\mu}_y'(d) + z_T\hat{\sigma}_y'(d)) \tag{9}$$

where regression estimates of the moments in log space are obtained using the relations

$$\hat{\mu}_y'(d) = \ln\left[\frac{\hat{\mu}_q'(d)}{[1 + (\hat{\sigma}_q'(d)/\hat{\mu}_q'(d))^2]^{1/2}}\right] \tag{10a}$$

$$\hat{\sigma}_y'^2(d) = \ln\left[1 + \left(\frac{\hat{\sigma}_q'(d)}{\hat{\mu}_q'(d)}\right)^2\right] \tag{10b}$$

TABLE 2. Summary of Ordinary Least Squares Regression Equations for Moments of the Anual Minimum d-day Low-Flows in Real Space for the 23 Study Sites.

$$\hat{\mu}_q'(d) = b_0 A^{b_1} H^{b_2}$$
$$\hat{\sigma}_q'(d) = b_0 A^{b_1}$$

MODEL	b_0	b_1	b_2	SE%§	R^2‡	r_ε†
$\mu_q'(3)$	0.0031 (−7.7)¤	1.152 (15.2)	0.454 (3.5)	35.1	95.7	0.967
$\mu_q'(7)$	0.00444 (−7.5)	1.132 (15.6)	0.430 (3.5)	33.6	95.8	0.968
$\mu_q'(14)$	0.0065 (−7.4)	1.100 (16.0)	0.411 (3.5)	31.8	96.0	0.978
$\mu_q'(30)$	0.0136 (−7.2)	1.065 (17.8)	0.369 (3.6)	27.5	96.7	0.988
$\sigma_q'(3)$	0.0429 (−14.1)	1.168 (16.9)		40.8	92.8	0.983
$\sigma_q'(7)$	0.0558 (−13.0)	1.122 (16.3)		40.6	92.3	0.976
$\sigma_q'(14)$	0.0759 (−11.2)	1.082 (15.2)		42.2	91.2	0.978
$\sigma_q'(30)$	0.1226 (−10.7)	1.058 (17.4)		35.5	93.2	0.979

† Critical value of r_ε is 0.963 using 5% significance level. r_ε is the probability plot correlation coefficient test statistic (see Vogel and Kroll, 1989).

‡ The values of R^2 are adjusted for the number of degrees of freedom which remain after parameter estimation.

§ Computed using $SE\% = 100[\exp(\sigma_\varepsilon^2)-1]^{1/2}$, where σ_ε^2 is the variance of the residuals ε, defined in (6).

¤ Values in parentheses are t-ratios of estimated model parameters $\ln[b_0]$, b_1, and b_2.

where $\hat{\mu}_q'(d)$ and $\hat{\sigma}_q'(d)$ are given by Eq. (8) and z_T is obtained from Eq. (2). Figure 2 summarizes the regression estimates of the distribution of annual minimum 7-day low-flows at the twenty-three sites using dashed lines. Overall, the results are quite good except for five sites (site numbers 1, 5, 8, 14, and 15) where the regression estimates are only fair approximations to both the fitted lognormal distribution (LN2) and the empirical distribution shown using open circles. In general, the lack-of-fit is probably due to model error that results from not including additional basin characteristics such as the hydraulic conductivity and baseflow recession constant, both of which describe the geohydrologic response of each watershed. Riggs (1961), Vogel *et al.* (1989), and Kroll (1989) describe the improvement in regional low-flow frequency models that result from the inclusion of the baseflow recession constant.

Detection of Influential Observations

An experiment was performed to detect whether the deletion of any one of the sites affects the fit of the regional regression model. If one or more sites are particularly influential, the resulting model would be questionable and more sites would be required to bolster our confidence in the regression models. Draper and Smith (1981, Section 3.12 and Section 6.8) summarize a number of standard statistics that can be employed to detect unusually influential observations.

As an alternative to these standard influence statistics, we use a graphical approach that is easier to explain. Each of the regression equations in Eq. (6) was fit to all sites except the one being validated, and the resulting estimates of $Q_{7,T}$ are shown as dotted lines in Figure 2. (Each dotted line represents the regional estimate of $Q_{7,T}$ when that site is dropped from the original sample of 23 sites.) In all cases, the subsample (validation) estimates are almost indistinguishable from the regression estimates based on all 23 sites. We conclude that none of the twenty-three sites employed in this study has an unusually large impact on the estimation of any of the parameters of the models in Eq. (6).

The Impact of Generalized Least Squares Regression Procedures

In the previous section ordinary least squares (OLS) regression procedures were employed to estimate the parameters of the regional model (eq. 6). Stedinger and Tasker (1985) observed that most regionalization problems, where the dependent variables are both cross-correlated and based on different record lengths, violate the commonly made assumptions that the model residuals are homoscedastic and independently distributed. They developed generalized least squares (GLS) regression procedures that account for the fact that the dependent variables are both cross-correlated and are based on samples with widely varying record lengths. A computer program which implements the GLS procedures is available (Tasker *et al.*, 1987) and has already been employed in regional hydrologic studies for flood frequency analysis (Tasker *et al.*, 1986; Potter and Faulkner, 1987) and urban runoff quality (Tasker and Driver, 1988). The authors are unaware of any previous applications of these procedures to regional low-flow frequency analysis.

GLS regression procedures were used to fit the following regional models for annual minimum 7-day average streamflows

$$\ln(\hat{Q}_{7,T}) = \ln[b_0] + b_1 \ln[A] + b_2 \ln[H] + \nu \qquad (12)$$

where

$$\hat{Q}_{7,T} = \exp(\hat{\mu}_y(7) + z_T \hat{\sigma}_y(7))$$

The residual error ν is made up of model error ε and sampling error η. The sampling error is due to the finite and varying length of records used to estimate $Q_{7,T}$. In addition, sampling error in the dependent variable occurs because the sequences of annual minimum 7-day low-flows are cross-correlated in space. The average cross-correlation of the sequences was 0.35 and this value was used to characterize the covariance matrix of the residuals. An alternative would have been to model the cross-correlation of flow-series as a function of distance between sites as suggested by Tasker *et al.* (1987); however, we could not find any significant relations between cross-correlation and distance.

Table 3 compares the parameter estimates and residual errors in Eq. (12) using both OLS and GLS regression procedures. The total prediction error ν is equal to the sum of the model error ε and the sampling error η. Stedinger and Tasker (1985) and Tasker *et al.* (1986) describe procedures for estimating these error components using both GLS and OLS procedures. The model parameter estimates and their associated t-ratios obtained using OLS and GLS procedures are almost indistinguishable. However, because the GLS procedures account for the heteroscedastic structure of the errors and the cross-correlation of the flow-series used to estimate the dependent variables, they result in smaller model errors than the OLS models. Tasker and Driver (1988) reached similar conclusions for regional urban runoff water quality models.

The total prediction error is the sum of the model error, sampling error, and measurement error. The latter is ignored here. GLS procedures document that the total prediction errors are larger than OLS procedures because GLS procedures account for the varying sampling errors associated with the dependent variables, but even GLS estimates of total error are low because they ignore measurement errors. The OLS and GLS model results in Table 3 show that the total prediction errors are largely due to model error; sampling error remains insignificant by comparison. If measurement errors are ignored, these results indicate that future research in regional low-flow frequency analysis should concentrate primarily upon reducing model error.

The Impact of Streamflow Record Augmentation Procedures

Streamflow record augmentation procedures developed by Fiering (1963) and Vogel and Stedinger (1985), exploit the cross-correlation between concurrent flow records to obtain improved estimates of the moments of the flows at a short-record site. Using the procedures outlined by Vogel and Stedinger (1985) and Vogel and Kroll (1990), record augmentation estimates of the mean and variance of the logarithms of the annual minimum 7-day low-flow series were obtained.

Streamflow record augmentation requires the selection of a nearby hydrologically similar basin with a longer record than the site in question. For each of the 23 basins described in Table 1, the basin that led to the largest increase in effective record length (see Vogel and Kroll, 1990) was considered the long-record neighbor to be exploited in the record-augmentation procedure. This amounted to choosing a neighbor with simultaneously a high cross-correlation and long record length. The record-augmentation estimates of the moments in log space, denoted $\hat{\mu}_y^*(7)$ and $\hat{\sigma}_y^*(7)$, were transformed into real space using the standard moment transformations for the lognormal distribution given in eq. 7, the resulting estimates are denoted $\hat{\mu}_q^*(7)$ and $\hat{\sigma}_q^*(7)$. The record-augmentation estimators $\hat{\mu}_q^*(7)$ and $\hat{\sigma}_q^*(7)$

were used to obtain regional regression equations for $\mu_q(7)$ and $\sigma_q(7)$.

Table 4 compares the OLS regional regression equations based upon the at-site sample estimates with the OLS regional regression equations based upon the record augmentation estimates of $\mu_q(7)$ and $\sigma_q(7)$. The model parameter estimates are quite similar in all cases except for the constant term b_0, which differs substantially in the regression equations for $\hat{\mu}_q{}^{*\prime}(7)$ and $\hat{\mu}_q{}'(7)$. For the mean, record augmentation also led to a slight reduction in the model error from 33.6% down to 32.0%, whereas for the standard deviation the reverse was true. We found that the small differences in the parameter estimates of b_0, b_1, and b_2 with and without the use of record augmentation, documented in Table 4, do not have a significant impact on resulting estimates of $\mu_q(7)$ and $\sigma_q(7)$, and hence $Q_{7,T}$, at the twenty-three sites used to develop the model.

TABLE 3. Comparison of OLS and GLS Regional Regression Equations for Estimating $Q_{7,T}$.

$$\hat{Q}_{7,T} = b_0 A^{b_1} H^{b_2} e^{\varepsilon} e^{\eta}$$

RETURN PERIOD (yrs.)	METHOD	ln[b0]	b_1	b_2	MODEL ERROR σ_ε	AVERAGE SAMPLING ERROR* σ_η	TOTAL ERROR σ_v
2	OLS	–6.96 (–7.4)†	1.13 (11.9)†	0.62 (3.8)†	0.428 (44.8)‡	0.154 (15.5)‡	0.455 (47.9)‡
	GLS	–6.99 (–7.3)†	1.14 (12.3)†	0.62 (3.9)†	0.405 (42.2)‡	0.162 (16.3)‡	0.436 (45.8)‡
10	OLS	–10.04 (–6.9)	1.12 (7.6)	0.96 (3.8)	0.665 (74.6)	0.240 (24.3)	0.698 (79.2)
	GLS	–10.05 (–6.8)	1.13 (7.8)	0.95 (3.8)	0.641 (71.3)	0.247 (25.1)	0.687 (77.7)

† Values in parentheses are t-ratios of estimated parameters.
‡ Values in parentheses are standard errors of estimate in real space (in percent) computed using $100[\exp(\sigma^2) - 1]^{1/2}$. The total error is computed using $\sigma_v{}^2 = \sigma_\varepsilon{}^2 + \sigma_\eta{}^2$.
* The average sampling error corresponding to the OLS procedures is computed as $p\sigma_\varepsilon{}^2/n$ where p is the number of model parameter estimates and n is the number of sites (n=23).

TABLE 4. Comparison of Ordinary Least Squares Regression Equations for Moments of the Annual Minimum 7-Day Low-Flows in Real Space with and without the Use of Streamflow Record Augmentation Procedures.

$$\hat{\mu}_q'(d) = b_0 A^{b_1} H^{b_2}$$

$$\hat{\sigma}_q'(d) = b_0 A^{b_1}$$

MODEL	b_0	b_1	b_2	SE%§	R^2‡	r_ε†
$\mu_q'(7)$	0.00444 (–7.5)¤	1.132 (15.6)¤	0.430 (3.5)¤	33.6	95.8	0.968
$\mu_q{}^{*\prime}(7)$	0.0078 (–7.0)	1.122 (16.2)	0.363 (3.1)	32.0	96.0	0.979
$\sigma_q'(7)$	0.0558 (–13.0)	1.122 (16.3)		40.6	92.3	0.976
$\sigma_q{}^{*\prime}(7)$	0.0696 (–11.7)	1.086 (15.4)		41.7	91.5	0.956

† Critical value of r_ε is 0.963 using 5% significance level.
‡ The values of R^2 are adjusted for the number of degrees of freedom which remain after parameter estimation.
§ Computed using $SE\% = 100[\exp(\sigma_\varepsilon{}^2)-1]^{1/2}$, where $\sigma_\varepsilon{}^2$ is the variance of the residuals ε, defined in (6).
¤ Values in parentheses are t-ratios of estimated model parameters ln[b0], b_1, and b_2.

CONCLUSIONS

This study describes a procedure for developing generalized regional regression equations for estimating the d-day, T-year low-flow statistic ($Q_{d,T}$) at ungaged sites in Massachusetts for d = 3, 7, 14, and 30 days. The resulting models (equations 8, 9, and 10) only require estimates of the drainage area and basin relief of the ungaged site. The procedure for estimating $Q_{d,T}$ at an ungaged site consists of substitution of the drainage area A, and basin relief H into Eq. (8) to obtain $\hat{\mu}_q'(d)$ and $\hat{\sigma}_q'(d)$, which are then substituted into Eq. (10) to obtain $\hat{\mu}_y'(d)$ and $\hat{\sigma}_y'(d)$, which in turn are substituted into Eq. (9) to obtain the final regression estimate $Q'_{d,T}$. The low-flow predictions from this model are comparable with previous regional low-flow investigations in terms of average prediction errors. The models developed here are useful in practice due to their ease of application and their ability to generate generalized frequencies and magnitudes of low-flows at ungaged sites.

Regional hydrologic models similar to those developed here, known as "state equations," are in widespread use for estimating peak floodflows at ungaged sites. For example, Newton and Herrin (1983) recommend such statistically based regional regression equations over the use of commonly used deterministic models for estimating floodflows at ungaged sites. Their recommendations are based upon a large nationwide study using competing methods developed by a number of federal agencies. Unfortunately, most studies, such as this one, which have attempted the same approach to estimate low-flow statistics at ungaged sites have met with only limited success. For example, Thomas and Benson (1970) show that average prediction errors for low-flow regional regression models in the Potomac river basin are at least twice as large as for analagous floodflow regional regression models in the same basin. Riggs (1961), Vogel *et al.* (1989), and Kroll (1989) show that further reductions in the prediction errors associated with such regional low-flow models will likely result from the inclusion of additional watershed characteristics such as basin hydraulic conductivity, average basin slope, and the baseflow recession constant.

This study has also compared the use of generalized least squares (GLS) and ordinary least squares (OLS) regression procedures. GLS regression procedures led to almost identical regional regression model parameter estimates when compared with the OLS procedures. Although GLS procedures led to only marginal gains in the prediction errors associated with low-flow regional regression equations, that result only reflects the fact that all sites had at least eleven years of data, and most had more than twenty years of data. In addition, the large model error component of the total prediction errors implies that the sampling errors had only a marginal impact on the analysis. In general, GLS procedures will have significant advantages over OLS procedures in studies which seek to include very short-records such as at partial-record sites. In such instances, GLS procedures can lead to significant improvements because the number of sites included in the analysis can be increased considerably.

In this study, record lengths ranged from 11 to 73 years and the average cross-correlation among concurrent traces of annual minimum 7-day low-flows was only 0.35. If the average cross-correlation among concurrent low-flow series were higher and if sites with record lengths smaller than, say, 10 years were used, then GLS procedures would likely result in significantly different and more precise model parameter estimates than OLS procedures.

Both OLS and GLS procedures document that only a small fraction of the total model error was due to sampling error; thus, efforts to reduce sampling error, such as the application of streamflow record augmentation procedures, led to only modest improvements in the models described here. In the future, attention should focus on the remaining sources of model error: measurement error, model form, and the inclusion of additional explanatory basin characteristics. After all, as Wallis (1965) so clearly showed, one cannot hope to uncover basic physical relationships using multivariate statistical procedures without prior knowledge of the physical relationships.

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 to Katherine Driscoll Martel for providing the estimates of relief in Table 1.

LITERATURE CITED

Blom, G., 1958. Statistical Estimates and Transformed Beta Variables. John Wiley and Sons, New York, N.Y., pp. 68-75 and pp. 143-146.

Draper, N. R. and H. Smith, 1981. Applied Regression Analysis, John Wiley & Sons, Second Edition, New York.

Fiering, M. B., 1963. Use of Correlation to Improve Estimates of the Mean and Variance. U.S. Geological Survey Professional Paper 434-C.

Fennessey, N. and R. M. Vogel, 1990. Regional Flow-Duration Curves at Ungaged Sites in Massachusetts. Journal of Water Resource Planning and Management, ASCE, 116(4).

Kroll, C. N., 1989. Estimation and Usage of Baseflow Recession Constants, M.S. Thesis, Tufts University, Medford, Massachusetts.

Newton, D. W. and J. C. Herrin, 1983. Estimating Flood Frequencies at Ungaged Locations. Proceedings of the Conference in Hydraulic Engineering, ASCE, pp. 528-533.

Potter, K. W. and E. B. Faulkner, 1987. Catchment Response Time as a Predictor of Flood Quantiles. Water Resources Bulletin 23(5):857-861.

Riggs, H. C., 1961. Regional Low Flow Frequency Analysis. U.S. Geological Survey Professional Paper 424-B.

Riggs, H. C., *et al.*, 1980. Characteristics of Low Flows. Journal of Hydraulic Engineering, ASCE, 106(5):717-731.

Stedinger, J. R., 1980. Fitting Log Normal Distributions. Water Resources Research 16(3):481-490.

Stedinger, J. R. and G. D. Tasker, 1985. Regional Hydrologic Analysis. 1. Ordinary, Weighted, and Generalized Least Squares Compared. Water Resources Research 21(9):1421-1432.

Tasker G. D., J. H. Eychaner, and J. R. Stedinger, 1986. Application of Generalized Least Squares in Regional Hydrologic Regression Analysis. U.S. Geological Survey Water-Supply Paper 2310, pp. 107-115.

Tasker, G. D., A. M. Lumb, W. O. Thomas, Jr., and K. M. Flynn, 1987. Computer Procedures for Hydrologic Regression and Network Analysis Using Generalized Least Squares. U.S. Geological Survey Open-File Report, Reston, Virginia.

Tasker, G. D. and N. E. Driver, 1988. Nationwide Regression Models for Predicting Urban Runoff Quality at Unmonitored Sites, Water Resources Bulletin 24(5):1091-1101.

Thomas, D. M. and M. A., Benson, 1970. Generalization of Streamflow Characteristics from Drainage-Basin Characteristics. U.S. Geological Survey Water-Supply Paper 1975.

Tukey, J. W., 1960. The Practical Relationship Between the Common Transformation of Percentages or Fractions and of Amounts. Technical Report No. 36, Statistical Research Group, Princeton, New Jersey.

Vogel, R. M. and J. R. Stedinger, 1985. Minimum Variance Streamflow Record Augmentation Procedures. Water Resources Research 21(5):715-723.

Vogel, R. M. and C. N. Kroll, 1989. Low-Flow Frequency Analysis Using Probability-Plot Correlation Coefficients. Journal of Water Resources Planning and Management, ASCE, 115(3):338-357.

Vogel, R. M., C. N. Kroll, and K. Driscoll, 1989. Regional Geohydrologic-Geomorphic Relationships for the Estimation of Low-Flows. Proceedings of the International Conference on Channel Flow and Catchment Runoff, University of Virginia, Charlottesville, Virginia, pp. 267-277.

Vogel, R. M. and C. N. Kroll, 1990. The Value of Streamflow Record Augmentation Procedures in Low-Flow and Flood-Flow Frequency Analysis. Journal of Hydrology (in press).

Wallis, J. R., 1965. Multivariate Statistical Methods in Hydrology—A Comparison Using Data of Known Functional Relationship. Water Resources Research 1(4).

Zecharias, Y. B. and Brutsaert, W., 1985. Ground Surface Slope as a Basin Scale Parameter. Water Resources Research 21(12):1895-1902.