International Seminar on Hydrology of Extremes

1-3 December, 1988
Roorkee (India)

MPS METHOD FOR FREQUENCY ANALYSIS OF LOW FLOWS

A. Ramachandra Rao   Chang-Hsin Hsieh

School of Civil Engineering, Purdue University
West Lafayette, Indiana 47907, USA

SYNOPSIS

The maximum product spacing (MPS) method of parameter estimation has been recently proposed as a method with superior characteristics. The MPS method is tested in this study by using synthetic and observed low flow data. The results from the MPS method are compared to those from the maximum likelihood (ML) method. The results from the MPS method are comparable to those from the ML method.

I. INTRODUCTION

Determining appropriate probability distributions to characterize low flow data is not an easy problem. Gumbel's type III distribution has been used for low flow analysis because of theoretical underpinnings related to this distribution. Log-normal and Gamma distributions have also been used in low flow frequency analysis.

Whichever distribution is used for low flow frequency analysis, the parameter estimation problem is difficult, especially for small number of observations. The maximum likelihood (ML) method may not converge and the moment estimates usually have large standard errors. In order to handle cases when the ML method may not converge, the maximum product of spacings (MPS) method has recently been proposed by Cheng and Amin (1979, 1982, 1983). The objective of the present study is to test the performance of the ML and MPS methods for estimation of parameters of Weibull distribution. The Weibull distribution is another form of Gumbel's type III distribution. Both generated and real data are used in this study.

The paper is organized as follows. The data used in the study are discussed in Section II. The probability distributions are discussed in Section III. The results obtained by generated data are presented in Section IV and those obtained by using real data are given in Section V. A set of conclusions are given in Section VI.

The major conclusion of the study is that both ML and MPS methods must be used to estimate parameters of these distributions, especially when sample sizes are small. This conclusion is based on the fact that the criterion functions for the ML or the MPS method may not have maxima in many cases, and hence the algorithms to estimate the parameters may not converge.

II. DATA USED IN THE STUDY

Three types of data are used in this study. The first set of data are generated by using the Weibull probability distribution. The Weibull distribution has three parameters α, β and γ (eq. 1). The parameters α, β and γ were selected so that three different distribution types are generated. For each case, five sets of data containing 20, 40, 60 80 and 100 observations are generated.

The second set of data are real hydrological data of low flows. These are the 1, 7, and 30 day low flow data from 12 stations in Indiana, U.S.A. The details of location of these stations, their areas, and the number of observations used in the study are given in Table 1. The locations of these stations are shown in Figure 1. The mean, $(\bar{x})$ standard deviation $(\hat{s}_x)$ and skewness (g_x) of these data indicate considerable variation in the characteristics of these data. Consequently the parameter estimation methods will be tested by using a variety of data.

Table 1. Sources of Low Flow Data

Station No. (USGS)	Name	Lat. (o)	Long. (o)	Watershed Area (mile^2)	Size N
03275500	East Fork Whitewater River at Richmond, IN	39.8067	84.9072	121	28
03333700	Wildcat Creek at Kokomo, IN	40.4733	86.1572	242	28
03326500	Mississinewa River at Marion, IN	40.5761	85.6594	682	58
03335500	Wabash River at Lafayette, IN	40.4219	86.8969	7267	60
03353000	White River at Indianapolis, IN	39.7514	86.1750	1635	54
04101000	St. Joseph River at Elkhart, IN	41.6917	85.9750	3370	36
05518000	Kankakee River at Shelby, IN	41.1828	87.3425	1779	61
03341500	Wabash River at Terre Haute, IN	39.4669	87.4189	12263	56
03361500	Big Blue River at Shelbyville, IN	39.5292	85.7819	421	40
03374000	White River at Petersburg, IN	38.5108	87.2894	11125	56
04182000	St. Marys River near Fort Wayne, IN	40.9878	85.1008	762	49
03363500	Flatrock River at St. Paul, IN	39.4175	85.6342	303	53

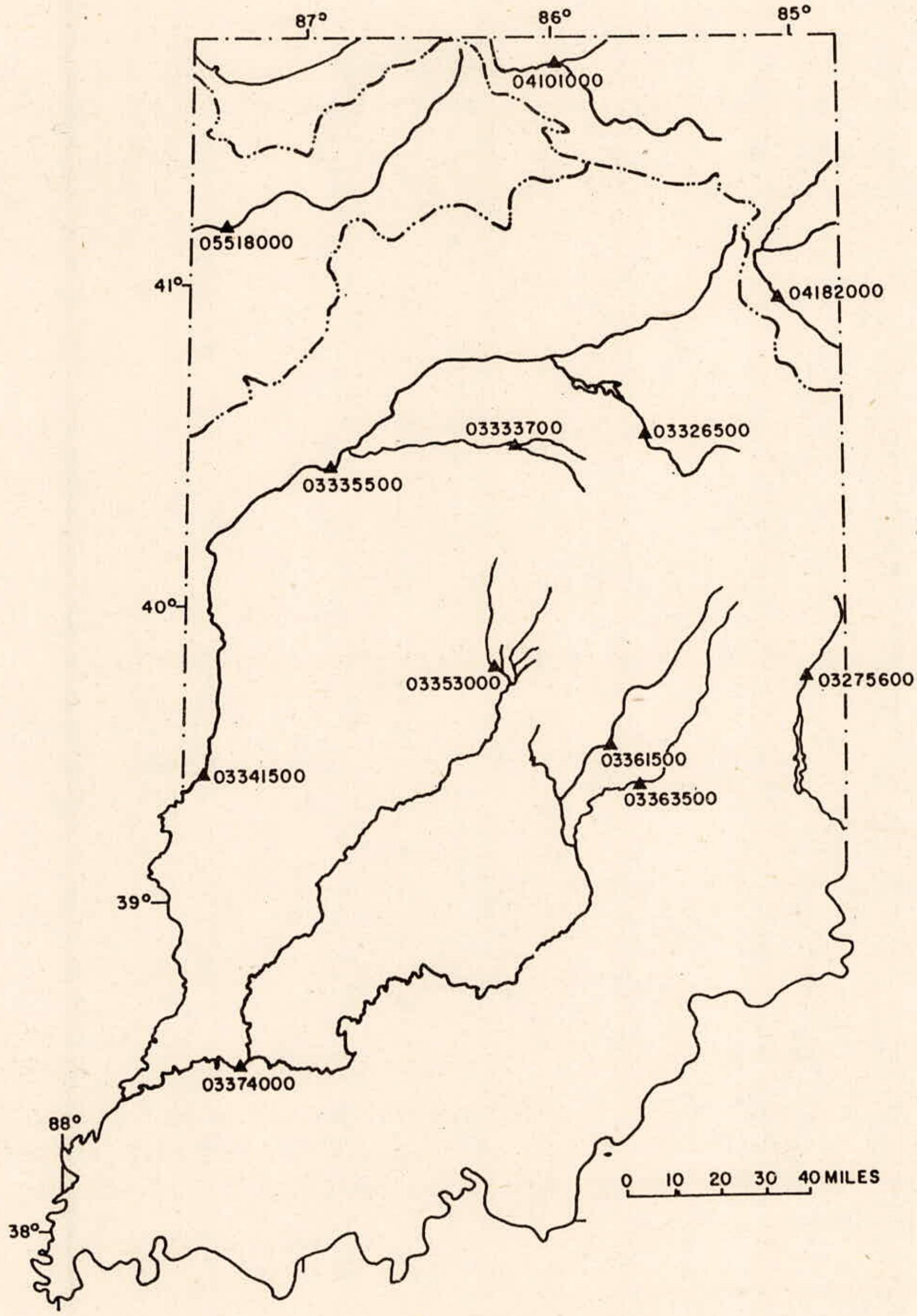

Figure 2.1 Location of Stations from Which Low Flow Data Are Taken

III. THEORETICAL ASPECTS

The three parameter Weibull or Gumbel's type III distribution is used in the present study. Some of the theoretical aspects of this distribution is discussed herein.

The probability density and cumulative distribution functions of the Weibull distribution are given in eqs. 1 and 2 respectively.

$$f_W(x) = \frac{\beta}{\gamma}\left[\frac{x-\alpha}{\gamma}\right]^{\beta-1} \exp\left[-\left(\frac{x-\alpha}{\gamma}\right)^{\beta}\right] \tag{1}$$

$$x>\alpha \; , \beta \; , \gamma>0, \; -\infty<\alpha<\infty$$

$$F_w(x) = 1-e^{-\left(\frac{x-\alpha}{\gamma}\right)^{\beta}} \tag{2}$$

The Gumbel's type III distribution is of the form given in eqs. 3 and 4.

$$\text{p.d.f.:} \; f_E(x) = \frac{\alpha'}{\beta' - \gamma'}\left[\frac{x-\gamma'}{\beta' - \gamma'}\right]^{\alpha' - 1} \exp\left[-\left(\frac{x-\gamma'}{\beta' - \gamma'}\right)^{\alpha'}\right] \tag{3}$$

$$\text{c.d.f.:} \; F_E(x) = 1-e^{-\left[\frac{x-\gamma'}{\beta' - \gamma'}\right]^{\alpha'}} \tag{4}$$

By defining α', γ' and $(\beta' - \gamma')$ in eqs. 3 and 4 to be respectively equal to β,α and γ in eqs. 1 and 2, it is easy to see that the Weibull and Gumbel's type III distributions are the same. For a given series of observations, x_i, $i = 1, 2, ..., N$, the ML and MPS estimates are discussed below.

Maximum likelihood estimates

The log-likelihood function of the Weibull distribution is given in eq. 5.

$$\ln L = N \ln \beta - N \beta \ln \gamma + (\beta-1) \sum_{i=1}^{N} \ln (x_i - \alpha) - \gamma^{-\beta} \sum_{i=1}^{N} (x_i - \alpha)^{\beta} \tag{5}$$

The derivatives of the log-likelihood equation with respect to the parameters of α, β and γ are given respectively in eqs. 6, 7, and 8.

$$L_\alpha = \frac{\partial \ln L}{\partial \alpha} = -(\beta-1)\sum_{i=1}^{N}(x_i-\alpha)^{-1} + \beta\gamma^{-\beta}\sum_{i=1}^{N}(x_i-\alpha)^{\beta-1} \quad (6)$$

$$L_\beta = \frac{\partial \ln L}{\partial \beta} = \frac{N}{\beta} + \sum_{i=1}^{N}\ln(x_i-\alpha) - N\ln\gamma \quad (7)$$

$$-\gamma^{-\beta}\left[\sum_{i=1}^{N}(x_i-\alpha)^{\beta}\ln(x_i-\alpha) - \ln\gamma\sum_{i=1}^{N}(x_i-\alpha)^{\beta}\right]$$

$$L_\gamma = \frac{\partial \ln L}{\partial \gamma} = -\frac{N\beta}{\gamma} + \beta\,\gamma^{-(\beta+1)}\sum_{i=1}^{N}(x_i-\alpha)^{\beta} \quad (8)$$

Solution of these equations give the maximum likelihood estimates of α, β and γ. Numerical methods must be used for solving these equations.

MPS estimates

For the maximum product of spacings (MPS), the logarithmic spacings function H defined in eq. 9 is used. y_i and Z_i are defined in eqs. 10 and 11.

$$H = \sum_{i=1}^{N}\left[\ln(-Z_i) - y_i\right] \quad (9)$$

$$\text{where} \qquad y_i = \frac{x_i-\alpha}{\gamma} \quad (10)$$

$$Z_i = 1-e^{y_i-y_{i-1}} \quad (11)$$

For the MPS solution an estimate α is first selected and the functions f_1 and f_2 given in eqs. 12 and 13 are defined. S_i and T_i values used in eqs. 12 and 13 are defined in eqs. 14 and 15.

$$f_1 = \sum_{i=1}^{N}(S_i+y_{i-1}) = 0 \quad (12)$$

$$f_2 = \sum_{i=1}^{N}\left\{T_i-\beta^{-1}y_{i-1}\ln y_{i-1}\right\} = 0 \quad (13)$$

$$S_i = \frac{y_i - y_{i-1}}{Z_i} \tag{14}$$

$$T_i = \frac{1}{\beta Z_i} \left[y_i \ln y_i - y_{i-1} \ln y_{i-1} \right] \tag{15}$$

Equations 12 and 13 are solved for β and γ. Then, another α value is selected and eqs. 12 and 13 are solved for β and γ and the procedure is repeated for different α values. The criterion function H is evaluated by using each set of α, β and γ values. Estimates which give the maximum H value are selected as the MPS estimate. The moment or the maximum likelihood estimates may be used as initial estimates for the MPS method.

IV. RESULTS FROM GENERATED DATA

A typical example of the parameters of the Weibull distribution estimated by the MPS and ML methods is shown in Figure 2. For this case $\alpha = 30$, $\beta = 3$ and $\gamma = 100$. The horizontal line at the center of the boxes indicate the mean, the top and bottom lines the one standard error limits and the range is indicated by the dashed lines. The range of the parameter estimates can be very large for small sample sizes and it decreases with increasing sample size. The mean parameter estimate approaches the true value with increasing sample size, although there is a discernible bias even with large sample sizes.

V. ANALYSIS OF REAL DATA

The ML estimates obtained by the present procedure were also compared (Hsieh and Rao, 1988) to those by Condie's program (Condie and Cheng (1980)). Condie and Cheng's program is designed to estimate the parameters of the ML estimates of the Gumbel's type III distribution. These ML estimates computed by the two methods were extremely close to each other in most of the cases (Hsieh and Rao, 1988). On the other hand, the MPS and MLE estimates were frequently different from each other as shown in Table 2. In the Weibull distribution the sum of the parameters α and γ is equal to the "characteristic drought" values. the characteristic drought values estimated by the ML and MPS methods are different but close to each other. All the characteristic drought values are positive and increase with duration as they should. An example of the goodness of fit between observed data and fitted distributions is shown in Figure 3.

The mean squared errors between observed and fitted distributions are given in Table 3. In almost all these cases the MSE between the observed and fitted distribution is larger and higher for the MPS method than for the ML method. However, the ML method did not converge for all cases but the MPS method converged for all the cases. For other 3 parameter distributions, and for many cases, the results from the MPS method were superior to those obtained by the ML method. Consequently it may be preferable to use both the MPS and ML methods to estimate the parameters of probability distributions in the frequency analysis of low flows.

VI. CONCLUSIONS

The results presented herein and supported by extensive studies (Hsieh and Rao, 1988) lead us to conclude that the MPS method complements the ML method. The estimates obtained by the MPS method are superior to those obtained by the method of moments. The MPS estimates are especially useful when ML estimates may not be available.

VII. REFERENCES

1. Cheng, R.C.H. and Amin, N.A.K., "Maximum Product-of-Spacings Estimation with Application to the Lognormal Distribution", Math. Rept. 79-1, The University of Wales, Institute of Science and Technology, Condiff, Wales, pp. 16, 1979.

2. Cheng, R.C.H. and Amin, N.A.K., "Estimating Parameters in Continuous Univariate Distributions with a Shifted Origin", Math. Rept. 82-1, The University of Wales, Institute of Science and Technology, Candiff, Wales, pp. 30, 1982.

3. Cheng, R.C.H. and Amin, N.A.K., "Estimating Parameters in Continuous Univariate Distributions with a Shifted Origin", *J.R. Stat. Soc., (B),* 45, No. 3, pp. 394-403, 1983.

4. Condie, R., and Cheng, L.C., "Lowflow Frequency Analysis - Program Lowflow" Report of Water Resources Branch, Inland Water Directorate, Environment Canada, Ottawa, Ontario, K1A 0E7, 1980.

5. Hsieh, C.H. and Rao, A.R., "MPS and ML Parameters Estimates", Tech. Rept. CE-HSF-88-01, School of Civil Engineering, Purdue University, W. Lafayette, IN 47907, USA, pp. 182, Jan. 1988.

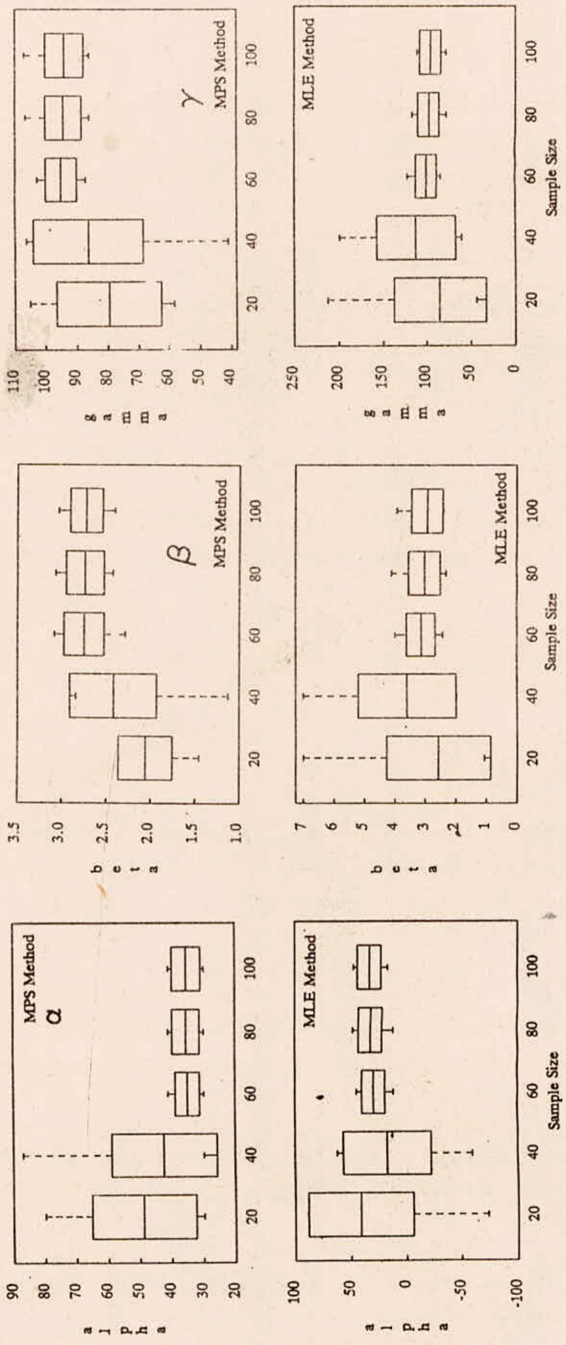

Figure 2. Variation of MPS and ML Estimates with Sample Size. $\alpha = 30$, $\beta = 3$ and $\gamma = 100$

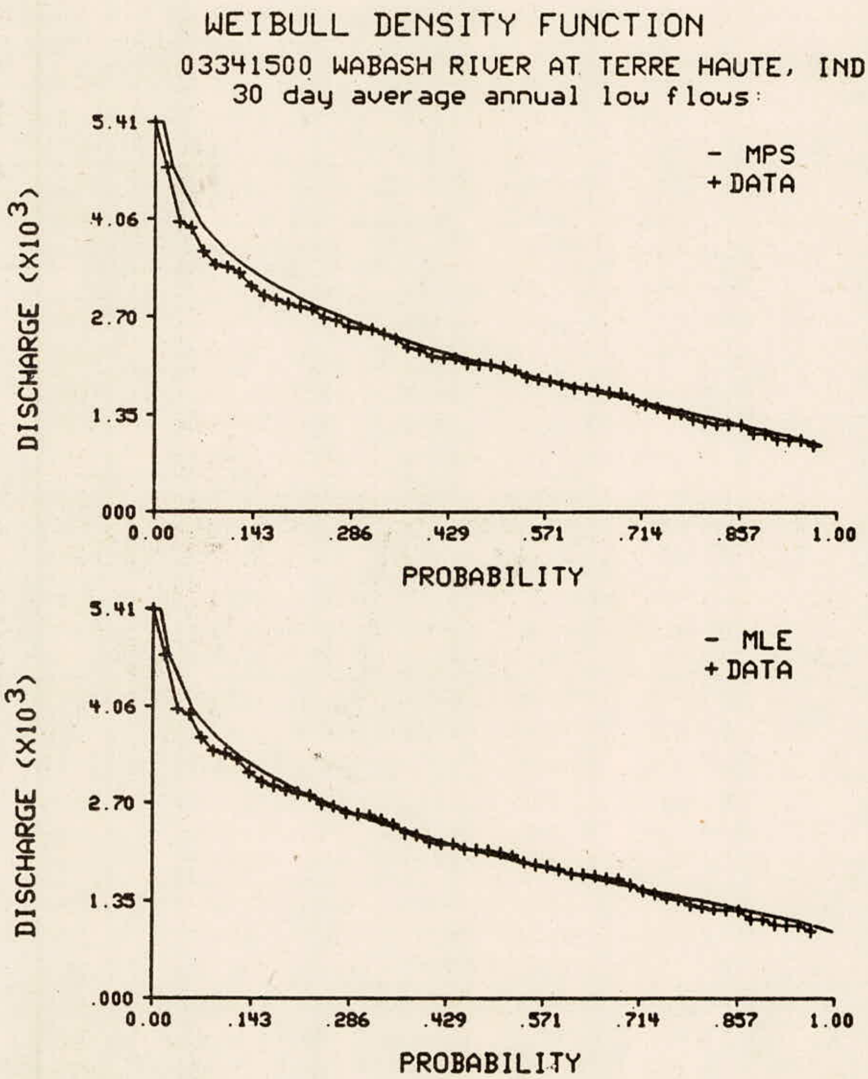

Figure 3. Observed Data and Fitted Distributions

Table 2. Parameters of Weibull Distribution Fitted to Low Flows

Station	Low-flow Day	MPS				MLE			
		α	β	γ	β'	α	β	γ	β'
03275500	1	-0.2	1.34777	19.4008	19	0.631773	1.36607	10.2454	11
	7	0.7002	1.34547	16.7335	18	1.77347	1.35217	10.9088	13
	30	3.12444	1.26389	20.7142	24	2.87700	1.21424	12.2592	16
03333700	1	1.260	2.2280	24.2640	25	4.94724	2.20520	16.0546	21
	7	2.48000	2.14536	27.6691	31	6.08169	2.07013	19.0711	25
	30	5.98000	1.72846	31.1324	37	8.90938	1.65018	22.3567	31
03326500	1	2.48926	1.31894	32.1525	35	2.96418	1.08541	19.4146	22
	7	2.56000	2.08035	48.9462	52	5.98119	1.96029	37.1999	43
	30	11.1760	2.04059	56.5831	68	14.1822	1.87860	43.4877	58
03335500	1	317.398	1.73654	737.341	1054	373.112	1.76376	590.038	963
	7	314.676	1.89499	816.684	1132	375.014	1.76962	729.467	1105
	30	332.550	1.72915	1053.83	1389	403.399	1.66033	958.770	1361
03353000	1	-12.4000	2.16566	163.577	152	0.580380	1.97184	135.975	137
	7	-13.4400	2.23024	188.592	176	2.45007	2.05513	157.251	160
	30	7.28000	1.85734	209.881	217	21.4861	1.71829	183.473	206
04101000	1	170.706	2.08050	1034.64	1206	277.558	2.01474	861.198	1139
	7	378.045	2.17464	1050.71	1429	498.345	2.05831	890.032	1388
	30	561.658	1.85206	1004.54	1567	666.676	1.76822	867.578	1534
05518000	1	196.765	2.46631	444.056	641	236.491	2.31287	377.241	613
	7	287.712	1.91767	357.827	646	316.097	1.99473	311.9800	628
	30	320.80	1.78815	393.929	714	345.511	1.76088	343.061	688
03341500	1	639.004	1.49948	126.533	1904	687.793	1.31825	1031.40	1719
	7	598.478	1.63042	1425.92	2024	690.026	1.54240	1280.34	1970
	30	827.188	1.38466	1542.91	2370	911.683	1.31898	1403.62	2314
03361500	1	24.4744	1.30007	56.6378	81	26.5627	1.27014	44.4078	71
	7	30.0189	1.14922	52.5678	83	31.8208	1.13365	42.6189	75
	30	33.3733	1.16311	68.7202	102	35.7170	1.16362	50.3979	86
03374000	1	538.160	1.14344	102.725	1565	568.97	1.11605	876.573	1446
	7	548.418	1.23139	1099.41	1647	588.881	1.20250	985.852	1575
	30	599.700	1.23042	1355.63	1956	642.671	1.22344	1230.18	1873
04182000	1	5.87018	1.72071	19.9576	26	6.45007	1.47071	11.2306	18
	7	5.39039	1.79045	23.2007	28	6.41930	1.61002	14.5344	21
	30	6.76184	1.45870	33.1871	40	7.69743	1.40086	21.7128	30
03363500	1	-	-	-	-	-	-	-	-
	7	0.761272	1.13737	28.3685	29	-	-	-	-
	30	0.577699	1.08704	35.6632	36	0.995440	1.01188	20.9994	22

Note: β' is characteristic drought

Table 3. Mean Squared Error Between Observed and and Fitted Distributions

Station No.	Day	Weibull MPS	Weibull MLE	Station No.	Day	Weibull MPS	Weibull MLE
	1	55.17	2.52		1	6.25	2.14
03275500	7	27.23	1.81	05518000	7	2.93	1.19
	30	42.16	3.65		30	4.03	1.29
	1	24.24	2.34		1	7.13	1.97
03333700	7	23.23	3.01	03341500	7	1.13	0.69
	30	17.53	2.11		30	0.53	0.53
	1	29.43	0.95		1	11.38	2.05
03326500	7	13.90	1.11	03361500	7	6.96	1.52
	30	17.75	1.68		30	12.77	1.69
	1	4.58	1.01		1	3.13	0.85
03335500	7	1.25	0.75	0337400	7	1.84	0.77
	30	1.15	0.91		30	2.08	0.87
	1	6.15	1.01		1	57.37	1.04
03353000	7	5.03	1.02	04182000	7	41.26	0.87
	30	2.83	0.98		30	33.76	2.30
	1	356.50	4.06		1	--	--
04101000	7	5 30	2.97	03363500	7	46.86	2.67
	30	4.79	3.06		30	30.80	1.28

Note: All the values must be divided by 1000.
*Based on the parameters obtained from Condie's program.