The 3-Parameter Lognormal Distribution and Its Applications in Hydrology

B. P. SANGAL AND ASIT K. BISWAS

Inland Waters Branch and Policy and Planning Branch
Department of Energy, Mines and Resources, Ottawa, Canada

Abstract. The 3-parameter lognormal distribution is a general skew distribution in which the logarithm of any linear function of a given variable is normally distributed. The distribution is applied to the frequency analysis of floods, annual flows, and monthly flows, and a comparison with other commonly used methods suggests that it can be successfully used for this purpose. A procedure for its application has been suggested using only the median, the mean, and the standard deviation of the original data. The Gumbel distribution is a special case, and any straight line on the Gumbel probability paper can be transformed into a straight line on the lognormal probability paper by the 3-parameter lognormal distribution. The sample of 10 stations used in this study all exhibited negative skewness for the logarithms of the data and therefore the lognormal distribution that assumes this skewness equal to zero has generally given high estimates.

INTRODUCTION

Few hydrologic populations can be represented by the normal distribution, and the degree of skewness depends upon the type of the data. For example, the distribution of daily flows is much more skewed than the distribution of annual flows. Since many of the statistical conclusions are based on the normal distribution, it has often been found advantageous to transform a skew distribution into the normal one. Theoretically it is always possible to determine a function that would yield such a transformation [*Hald,* 1962], even though some transformations may be quite involved.

One of the most important and useful of such transformations is the logarithmic transformation. It has been observed that whereas the original data show considerable skewness, their logarithms are nearly normally distributed, and hence the original data are said to follow the lognormal distribution. The distribution has been extensively used in many areas of science and engineering. *Chow* [1954] made a comprehensive study of this distribution and theoretically derived its parameters. The objective of this paper is to present the theory and some mathematical characteristics of the more general 3-parameter lognormal distribution and to show some of its applications in hydrology.

THEORY

The 3-parameter lognormal distribution represents the normal distribution of log $(x - a)$ where the logarithm is to the base e. The variable x is the random variable, a is a parameter, and $(x - a)$ is the reduced variable. This is a general distribution in which the logarithm of any linear function of x is normally distributed. The parameter a has to be evaluated in terms of the statistical measures of the variable x. It can be positive, zero, or negative. The significance of these signs will be discussed herein.

Matalas [1963] derived an expression for determining the parameter a which involved the coefficient of skewness of the variable x. If the data are limited, as is the usual case in hydrology, the computation of the coefficient of skewness may be subject to error. Its computation is, therefore, discouraged when the sample size is less than 100 events [*Interagency Committee,* 1966]. Although *Matalas and Benson* [1968] attempted to show that the skew coefficient, when computed even from smaller samples, may be reliable, the parameter a can be determined in the following manner without using this coefficient.

For the random variable x, the mean μ_x and variance σ_x^2 are related to the parameter a, and to the mean μ_y and variance σ_y^2 of the transformed variable y, equal to log $(x - a)$ which is normally distributed [*Matalas*, 1967]. Thus

$$\mu_x = a + e^{(\sigma_y^2/2+\mu_y)} \tag{1}$$

and

$$\sigma_x^2 = (e^{\sigma_y^2} - 1)\cdot e^{(2\mu_y+\sigma_y^2)} \tag{2}$$

Let ζ_y represent the median of y. Since y is normally distributed, $\mu_y = \zeta_y$. If ζ_x denotes the median of x, then $(\zeta_x - a)$ denotes the median of $(x - a)$. Therefore $\zeta_y = \log(\zeta_x - a)$. Thus

$$e^{\mu_y} = \zeta_x - a \tag{3}$$

By using this value of e^{μ_y} and by eliminating $e^{\sigma^2_y}$ from equations 1 and 2, we get

$$\begin{aligned}
&2a^3(\mu_x - \zeta_x)\\
&+ a^2(\sigma_x^2 + \zeta_x^2 - 5\mu_x^2 + 4\mu_x\zeta_x)\\
&+ 2a(2\mu_x^3 - \zeta_x\sigma_x^2 - \zeta_x\mu_x^2 - \mu_x\zeta_x^2)\\
&+ \zeta_x^2\sigma_x^2 - \mu_x^4 + \mu_x^2\zeta_x^2 = 0
\end{aligned} \tag{4}$$

This is the basic equation that relates the parameter a to the mean μ_x, the median ζ_x, and the standard deviation σ_x of the variable x. This is a cubic equation in terms of a. It must have one real and two imaginary roots that would be proved later.

If $\alpha = a/\mu_x$, $\beta = \zeta_x/\mu_x$, and $\gamma = \sigma_x/\mu_x$, then equation 4 is transformed to

$$\begin{aligned}
2\alpha^3(1 - \beta) + \alpha^2(\gamma^2 + \beta^2 - 5 + 4\beta)&\\
+ 2\alpha(2 - \beta\gamma^2 - \beta - \beta^2)&\\
+ \beta^2\gamma^2 - 1 + \beta^2 = 0&
\end{aligned} \tag{5}$$

This is in dimensionless form. In the special case of the lognormal distribution when $a = 0$ and thus $\alpha = 0$, the equation gives

$$\beta = 1/(1 + \gamma^2)^{1/2} \tag{6}$$

Thus if x is lognormally distributed, equation 6 has to be satisfied. This is a stringent condition, andl field data, especially hydrologic data, will rarely satisfy this condition exactly.

If σ_y is assumed to be small, an approximate but simple expression for a can be obtained by neglecting higher powers of σ_y. This assumption is not unrealistic and most hydrologic data can be analyzed with this assumption (appendix).

From equations 1 and 2

$$\mu_x = a + e^{\mu_y}(1 + \sigma_y^2/2) \tag{7}$$

and

$$\sigma_x^2 = e^{2\mu_y}\cdot\sigma_y^2 \tag{8}$$

By substituting the value of e^{μ_y} from equation 3 and by eliminating σ_y^2 from equations 7 and 8, we get

$$a = \zeta_x - \frac{\sigma_x^2}{2(\mu_x - \zeta_x)} \tag{9}$$

This equation can be conveniently used in determining the parameter a. Its dimensionless form is

$$\alpha = \beta - \frac{\gamma^2}{2(1 - \beta)} \tag{10}$$

which can be used in regional studies.

CHARACTERISTICS OF THE DISTRIBUTION

The characteristics of the 3-parameter lognormal distribution are analogous to those of the lognormal distribution. In the former case the distribution applies to the reduced variable $(x - a)$, whereas in the latter case it applies to the variable x. For the lognormal distribution *Chow* [1954] derived the following basic relation between the coefficient of variation and the coefficient of skewness:

$$g_x = 3v_x + v_x^3 \tag{11}$$

where g_x and v_x are the coefficient of skewness and the coefficient of variation of x. For the 3-parameter lognormal distribution the relation becomes

$$g_{(x-a)} = 3v_{(x-a)} + v_{(x-a)}^3 \tag{12}$$

where $g_{(x-a)}$ and $v_{(x-a)}$ are the coefficient of skewness and the coefficient of variation of the reduced variable $(x - a)$.

The coefficient of skewness and the standard deviation of the $(x - a)$ series are the same as those of the x series since the parameter a has no effect on these measures. But the mean of the $(x - a)$ series differs from that of the x series by an amount equal to a. Therefore for

3-parameter lognormal distribution

$$g_x = 3v_{(x-a)} + v_{(x-a)}^3 \tag{13}$$

It gives a cubic equation in $v_{(x-a)}$

$$v_{(x-a)}^3 + 3v_{(x-a)} - g_x = 0 \tag{14}$$

Equation 14 has three roots. According to the Descartes rule of signs it has only one positive root. The other two roots are either negative or imaginary. Since the equation does not have a $v_{(x-a)}^2$ term it cannot have negative roots and therefore the other two roots must be imaginary. This is also evident from the fact that $v_{(x-a)}$ must be positive. Since there is only one positive value of $v_{(x-a)}$ to satisfy equation 14, there can be only one real value of a, which proves the earlier statement that a can have only one real value.

In any given series of x, three cases are possible

$$\text{(a)} \quad g_x > 3v_x + v_x^3 \tag{15}$$

$$\text{(b)} \quad g_x = 3v_x + v_x^3 \tag{16}$$

and

$$\text{(c)} \quad g_x < 3v_x + v_x^3 \tag{17}$$

In case (a) the series would plot curved upward on the lognormal probability paper. The coefficients of skewness of the x and the log x series would both be positive. In case (b) the series would plot a straight line on the lognormal probability paper. The coefficient of skewness of the x series would be positive, and the coefficient of skewness of the log x series would be zero. This is the required condition of the lognormal distribution. In case (c) the series would plot curved downward on the lognormal probability paper and the coefficient of skewness of the log x series would be negative, but the coefficient of skewness of the x series can either be positive, zero, or negative. However, only the series with positive skewness can be treated by the 3-parameter lognormal distribution. The case of $g_x = 0$ with $v_x = 0$ is unimportant, since the variable x is constant.

For these three cases of positive skewness, the role played by the parameter a will now be examined. In case (a) the role would be to make the coefficient of variation of the $(x - a)$ series greater than that of the x series so that equation 13 is satisfied. Since it would require a reduction in the mean of the x series, a would be positive in this case. The case (b) represents a condition of lognormal distribution and a would be equal to zero. In case (c) the effect of a would be to make the coefficient of variation of the $(x - a)$ series less than that of the x series. Since it would require an increase in the mean of the x series, a would be negative in this case. Thus the sign of a can be predetermined from the plot of the original data on the lognormal probability paper. If the plot is curved upward, a would be positive. If the plot is a straight line, a would be zero. If the plot is curved downward, a would be negative.

In the case of lognormal distribution the three measures of location are given by [*Hald*, 1962]

Mode

$$\log x = \log \xi - 2.3026\, s_{\log x}^2 \tag{18}$$

Median

$$\log x = \log \xi \tag{19}$$

Mean

$$\log x = \log \xi + 1.1513\, s_{\log x}^2 \tag{20}$$

For the 3-parameter lognormal distribution these equations become

Mode

$$\log (x - a) = \log \xi - 2.3026\, s_{\log(x-a)}^2 \tag{21}$$

Median

$$\log (x - a) = \log \xi \tag{22}$$

Mean

$$\log (x - a) = \log \xi + 1.1513\, s_{\log (x-a)}^2 \tag{23}$$

In these equations the logarithms are to the base 10. *Thöni* [1969] has given a more accurate method of estimating the mean of a lognormal distribution. However, if the sample size is 30 or more, equation 20 is sufficiently accurate.

ESTIMATION OF PARAMETERS

The parameters used in the present study have been estimated by the following formulas:

Mean

$$\langle x \rangle = \frac{\sum x}{n} \tag{24}$$

where n is the number of items in the series.

Standard deviation

$$s_x = \left(\frac{\sum (x - \langle x \rangle)^2}{n - 1} \right)^{1/2} \tag{25}$$

Coefficient of skewness

$$g_x = \frac{n \sum (x - \langle x \rangle)^3}{(n - 1)(n - 2) s_x^3} \tag{26}$$

The median represents the value of the variable at 50% probability. If the sample is large, the determination of the median is simple. However, with hydrologic data a basic difficulty is the sample size since this is usually small. For a small sample the arithmetic interpolation would not yield reliable results; therefore the graphical interpolation is necessary. The data lying between 30% and 70% probabilities may be plotted corresponding to their probabilities on arithmetic paper to a large scale, and the median may be estimated from a smooth curve through these points. Generally the curve between 40% and 60% probabilities is almost a straight line. If it is assumed that the frequency is uniformly distributed over this interval, then the arithmetic mean of the data lying between this interval would give a fairly accurate value of the median. Therefore the median can be computed as equal to the mean of the middle fifth of the data, and unless the data show appreciable curvature in the 40% to 60% interval this method of median computation can be adopted for the sake of uniformity.

EXAMPLES

The usefulness of any theoretical method is dependent on its practical use. To test the applicability of the 3-parameter lognormal distribution to flow frequency analysis, it was decided to compare it with several other methods that have been in common use. A group of 10 long-term stations was selected where flows were not affected by artificial regulation. All data were considered, and no effort was made to avoid the high floods, the so-called outliers.

Table 1 gives the 10 test stations, their Water Survey of Canada inventory numbers, drainage areas in square miles, and the number of years of record. The original streamflow data are available on request to the authors.

The different statistical measures that have been used in the analysis of the data by various methods include the mean, the standard deviation, the coefficient of variation, and the coefficient of skewness. These measures were computed for the original data as well as for their logarithms to the base 10. For the 3-parameter lognormal distribution a was computed from equation 9. The median was computed as the mean of the middle fifth of the data. For regional considerations the ratio of the median to the mean (M) and the ratio of the parameter a to the mean (R) were also computed. Having computed a, the different measures were computed for the log $(x - a)$ series. All these results are listed in Table 2. The data analyzed include the annual flood flows of all the 10 stations, and the annual flows and the April flows of station 2GD-1 only. The last two series were included to demonstrate the applicability of the 3-parameter lognormal distribution to the annual flows and to the monthly flows. The analysis of April flows is presented as an example

TABLE 1. Ten Test Stations

Water Survey of Canada Station No.	Location	Drainage Area, mi^2	Years of Record
8MF-5	Fraser River at Hope, British Columbia	83,700	57
2FC-1	Saugeen River near Port Elgin, Ontario	1,570	52
2FC-2	Saugeen River near Walkerton, Ontario	850	52
2GD-1	Thames River near Ealing, Ontario	519	51
2HB-1	Credit River near Cataract, Ontario	82	51
2HL-1	Moira River near Foxboro, Ontario	1,040	51
2CE-2	Aux Sables River near Massey, Ontario	524	46
4LJ-1	Missinaibi River near Mattice, Ontario	3,450	46
5PB-14	Turtle River near Mine Centre, Ontario	1,880	46
5QA-1	English River near Sioux Lookout, Ontario	5,240	45

TABLE 2. Statistical Measures

Station No.	Series	Mean	Standard Deviation	Coefficient of Variation	Coefficient of Skewness	Median with Ratio to Mean (M)	Parameter a with Ratio to Mean (R)
8MF-5	x	317,000	62,700	0 198	0 685	311,000 (0 980)	−6,460 (−.020)
	log x	5.49310	0.08558	0 01558	−0 19541	—	—
	log $(x - a)$	5 50219	0.08378	0.01523	−0.19859	—	—
2FC-1	x	17,160	6,410	0.374	0.387	16,850 (0.982)	−50,150 (−2.923)
	log x	4.20208	0.17521	0 04170	−0.49986	—	—
	log $(x - a)$	4.82617	0.04106	0.00851	−0.08171	—	—
2FC-2	x	10,060	4,500	0.447	1.101	9,090 (0.903)	−1,230 (−.123)
	log x	3.96274	0 18860	0.04759	−0.00478	—	—
	log $(x - a)$	4 02177	0.16508	0.04105	0.13454	—	—
2GD-1	x	6,620	3,610	0.545	1.781	5,980 (0.904)	−4,260 (−.644)
	log x	3.76091	0 23892	0.06353	−0 52635	—	—
	log $(x - a)$	4.01654	0 13083	0.03257	0.37617	—	—
2HB-1	x	827	441	0 534	1 318	732 (0 885)	−292 (−.353)
	log x	2 85784	0.23925	0 08372	−0.56127	—	—
	log $(x - a)$	3 01873	0 16163	0.05354	0 12390	—	—
2HL-1	x	7,280	2,220	0 306	−0.009	7,200 (0.990)	−25,280 (−3.475)
	log x	3.83898	0.14932	0.03890	−0 87706	—	—
	log $(x - a)$	4.51166	0.02986	0 00662	−1.12765	—	—
2CE-2	x	3,770	1,320	0.349	0 555	3,650 (0 968)	−3,620 (−.959)
	log x	3.54861	0 16317	0 04598	−0.69850	—	—
	log $(x - a)$	3 86173	0 07674	0 01987	−0.03036	—	—
4LJ-1	x	31,690	9,350	0 295	0.006	31,370 (0 990)	−105,700 (−3.336)
	log x	4 48009	0 14118	0.03151	−0.71110	—	—
	log $(x - a)$	5 13701	0 02977	0.00579	−1 45705	—	—
5PB-14	x	4,400	2,150	0 489	0 811	4,020 (0 914)	−2,090 (− 474)
	log x	3 59122	0.21907	0 06100	−0 15953	—	—
	log $(x - a)$	3 78939	0.14024	0 03701	0 15723	—	—
5QA-1	x	10,320	5,140	0 498	1 159	8,950 (0.867)	−661 (− 064)
	log x	3 96336	0 21457	0 05414	−0.17820	—	—
	log $(x - a)$	3 99697	0.19822	0 04959	−0 06861	—	—
2GD-1	x*	474	139	0 293	0 495	460 (0 970)	−230 (−.485)
	log x	2 65753	0 12940	0 04869	−0 20293	—	—
	log $(x - a)$	2 83961	0 08493	0.02991	0.05552	—	—
	x†	1,045	597	0 571	1 052	897 (858)	−307 (−.294)
	log x	2 94973	0.25501	0 08645	−0 26476	—	—
	log $(x - a)$	3 09247	0.18378	0 05943	0 14828	—	—

x* Annual flow in cfs.
x† April flow in cfs.
All other x's are flood flows in cfs.

although the method is applicable to other months also.

The following five methods of frequency analysis were applied to the flood series: (1) Gumbel distribution; (2) Pearson type 3 distribution; (3) log-Pearson type 3 distribution; (4) lognormal distribution; and (5) 3-parameter lognormal distribution. To the annual series and the monthly series the following three methods were applied: normal distribution, lognormal distribution, and 3-parameter lognormal distribution. All these methods, except the 3-parameter lognormal distribution, are now in common use, and the procedures for their applications have been fairly standardized. However, a brief description of the procedures used in this study is given below.

The parameters of the Gumbel distribution were obtained by the method of maximum likelihood [*Panchang and Aggarwal*, 1962; *Panchang*, 1967]. In the Pearson type 3 distribution, the skewness was computed by equation 26 and no adjustment for the length of record was made. The Pearson type 3 coordinates were computed using the approximate formula given by *Beard* [1965]

$$K = \frac{2}{g_x}\left\{\left[\frac{g_x}{6}\left(T - \frac{g_x}{6}\right) + 1\right]^3 - 1\right\} \quad (27)$$

where K is the Pearson type 3 coordinate; g_x is the coefficient of skewness of x; and T is the standard normal deviate.

The log-Pearson type 3 distribution is of special significance since it has been recommended as a base method for analyzing flood flow frequencies for the U. S. federal agencies [*Benson*, 1968]. Benson has discussed its application. The suggested procedure was used, except that the Pearson type 3 coordinates were computed by equation 27.

For the lognormal distribution, the logarithms of the discharges at the selected probabilities were obtained from

$$\log x = \langle \log x \rangle + Ts_{\log x} \quad (28)$$

where $\langle \log x \rangle$ and $s_{\log x}$ are the mean and the standard deviation of the logarithmic series.

The procedure for the 3-parameter lognormal distribution was similar to the one used for the lognormal distribution. In this method the logarithms of the discharges are given by

$$\log (x - a) = \langle \log (x - a) \rangle + Ts_{\log (x-a)} \quad (29)$$

where $\langle \log (x - a) \rangle$ and $s_{\log (x-a)}$ are the mean and the standard deviation of the log $(x - a)$ series.

Equation 9, used to determine a, is sensitive to the condition of the median being very close to the mean. In this situation one of the following two methods may be used:

(1) Plot the data on the normal probability paper. In most cases this plot would very nearly be a straight line.

(2) Assume a value of the median close to but less than the mean, say 99% of the mean, and compute a. Two examples of this method are the analyses of flood flows of stations 2HL-1 and 4LJ-1. Both these basins are affected by appreciable natural lake storage. In both cases the computed median exceeded the mean. For a positively skewed distribution, the median should be less than the mean. Therefore the analyses were made with the assumed median being equal to 99% of the mean. Plots for both these stations were very nearly straight lines on the normal probability paper. This method helped in fitting the 3-parameter lognormal distribution without causing any significant change in the distribution.

For the normal distribution the discharges were obtained from

$$x = \langle x \rangle + Ts_x \quad (30)$$

where $\langle x \rangle$ and s_x are the mean and the standard deviation of the x series.

Table 3 gives the estimated mode, median,

TABLE 3. Measures of Location Estimated from Various Distributions

Station No.	Distribution	Mode	Median	Mean
8MF-5	G	288,000	308,000	320,000
	LN	299,000	311,000	317,000
	3LN	300,000	311,000	317,000
2FC-1	G	14,100	16,130	17,300
	LN	13,530	15,930	17,280
	3LN	16,270	16,860	17,160
2FC-2	G	8,080	9,300	10,000
	LN	6,840	8,780	9,940
	3LN	7,870	9,280	10,070
2GD-1	G	5,090	6,050	6,600
	LN	4,260	5,770	6,710
	3LN	5,230	6,130	6,610
2HB-1	G	635	754	822
	LN	532	721	839
	3LN	617	752	827
2HL-1	G	6,170	6,950	7,400
	LN	6,130	6,900	7,320
	3LN	7,050	7,200	7,280
2CE-2	G	3,150	3,570	3,820
	LN	3,070	3,540	3,800
	3LN	3,440	3,660	3,770
4LJ-1	G	27,060	30,280	32,150
	LN	27,180	30,210	31,850
	3LN	30,730	31,380	31,700
5PB- 14	G	3,420	4,020	4,370
	LN	3,030	3,900	4,430
	3LN	3,460	4,070	4,400
5QA-1	G	8,070	9,450	10,250
	LN	7,200	9,190	10,380
	3LN	7,400	9,270	10,360
2GD-1*	LN	435	474	495
	3LN	435	461	475
2GD-1†	LN	647	897	1,060
	3LN	727	930	1,050

* Annual flow in cfs.
† April flow in cfs.
All other data are flood flows.

G, Gumbel;
LN, lognormal;
3LN, 3-parameter lognormal.

and mean from the Gumbel, the lognormal, and the 3-parameter lognormal distributions for the flood flows. It also shows these measures for the annual flows and for the April flows of station 2GD-1 corresponding to the last two distributions. In the Gumbel distribution, the mode, the median, and the mean occur at 1.58 year, 2 year, and 2.33 year recurrence intervals, respectively. In the lognormal and the 3-parameter lognormal distributions these measures have been computed from equations 18 to 23.

Table 4 lists the results of the computations of flow magnitudes at the selected probabilities of 20%, 10%, 2%, 1%, 0.1%, and 0.01%, which for annual data respectively correspond to 5, 10, 50, 100, 1000, and 10,000-year recurrence intervals. Computations at small probabilities are chosen to illustrate the basic differences in various methods which do not become quite evident at higher probabilities.

Figure 1 is a probability plot of flood flows of station 2GD-1 on lognormal probability paper. The probability for each item was computed as $m/(n + 1)$, where m is the order number of the descending series and n is the total number of events. The transformed line represents the distribution of the reduced variable $(x - a)$, which must be a straight line. The fitted curve through the plotted points represents the curve obtained from the transformed line by algebraically adding a to the values of the transformed line at some selected probabilities. These points are denoted by crosses. This curve represents the distribution of the variable x. The position of the 1937 flood is of interest since its cumulative probability is 0.1%; that is, it is a 1000-year flood.

Figures 2 and 3 show the analyses of the annual flows and the April flows of the same station. The method of fitting the curve through the plotted points was the same as that used for the flood flows.

RESULTS AND DISCUSSION

Since the streamflow records used in the present study are of reasonable length, there is a good possibility that the parameters of the various distributions have been computed with some degree of accuracy.

In all cases the skewness of the logarithms of the data was negative. That is why the lognormal distribution, which assumes the skewness of the logarithms equal to zero, generally gave high flow estimates. For the same reason, no attempt was made to evaluate the log-Gumbel distribution that assumes the skewness of 1.139 for the logarithms of the data. An examination of the results presented by *Benson* [1968] also suggested that the log-Gumbel estimates are unrealistic and should be avoided.

There is no standard method of comparing the merits of one distribution with another. In the central part of the frequency curve, where plotting positions are more reliable, all distributions have given almost the same results, which are close to the data values. However, at small probabilities where no basis for comparison is available, the divergence is quite substantial. Generally at these probabilities the 3-parameter lognormal distribution has occupied the middle position.

In the analyses of the annual and April flows of station 2GD-1, the lognormal distribution has given the highest estimates followed by the 3-parameter lognormal distribution. The normal distribution, as expected, gave the lowest estimates.

In all cases good results were obtained by the 3-parameter lognormal distribution. Table 3 shows that the theoretical median and the mean as obtained from this distribution are quite close to the median and the mean as computed from the original data. It proves the validity of the method.

The 3-parameter lognormal distribution is a general skew distribution. The Gumbel distribution, which has a fixed skewness, is a special case, and any straight line on the Gumbel probability paper can be transformed into a straight line on the lognormal probability paper by the 3-parameter lognormal distribution. It may be noted that this flexibility is not available with the lognormal distribution. The Gumbel and the lognormal distributions are identical only when the coefficient of variation of the original data is 0.364. This is a very limiting condition and would rarely be encountered in practice. However, with the 3-parameter lognormal distribution, the coefficient of variation of the reduced variable $(x - a)$ is automatically adjusted to 0.364, irrespective of the coefficient of variation of the original data.

The Pearson type 3 distributions and the log-Pearson type 3 distributions require the compu-

TABLE 4. Computed Flows for Selected Probabilities by Various Distributions

Station No.	Distribution	Cumulative Probability, %					
		20	10	2	1	0.1	0.01
8MF-5	G	372,000	414,000	506,000	546,000	675,000	804,000
	P	367,000	401,000	468,000	494,000	574,000	649,000
	LP	368,000	399,000	457,000	478,000	542,000	598,000
	LN	367,000	401,000	467,000	492,000	572,000	648,000
	3LN	367,000	401,000	466,000	491,000	570,000	645,000
2FC-1	G	22,390	26,530	35,660	39,510	52,260	64,980
	P	22,400	25,600	31,630	33,880	40,600	46,570
	LP	22,490	26,010	29,980	35,090	42,080	47,820
	LN	22,370	26,710	36,470	40,700	55,400	71,430
	3LN	22,420	25,500	31,230	33,350	39,600	45,110
2FC-2	G	13,070	15,560	21,050	23,380	31,050	38,710
	P	13,400	16,060	21,690	23,980	31,730	38,510
	LP	13,230	16,010	22,370	25,160	35,010	45,960
	LN	13,230	16,020	22,390	25,200	35,120	46,170
	3LN	13,250	15,880	21,720	24,220	32,800	42,000
2GD-1	G	8,990	10,950	15,240	17,060	23,060	29,050
	P	8,920	11,320	16,850	19,240	27,370	35,780
	LP	9,240	11,230	15,220	16,750	21,290	25,160
	LN	9,160	11,670	17,850	20,730	31,560	44,640
	3LN	9,130	11,020	15,020	16,670	22,090	27,560
2HB-1	G	1,120	1,360	1,900	2,120	2,870	3,610
	P	1,140	1,410	2,010	2,250	3,060	3,860
	LP	1,160	1,400	1,890	2,070	2,600	3,040
	LN	1,150	1,460	2,240	2,600	3,960	5,600
	3LN	1,140	1,390	1,950	2,190	3,010	3,880
2HL-1	G	9,350	10,940	14,440	15,920	20,810	25,690
	P	9,150	10,120	11,830	12,430	14,110	15,500
	LP	9,250	10,250	11,840	12,330	13,480	14,170
	LN	9,220	10,730	13,990	15,360	19,970	24,800
	3LN	9,140	10,190	12,130	12,830	14,890	16,670
2CE-2	G	4,880	5,740	7,640	8,450	11,100	13,750
	P	4,830	5,510	6,850	7,360	8,900	10,320
	LP	4,880	5,520	6,620	6,990	7,960	8,640
	LN	4,850	5,730	7,650	8,480	11,290	14,310
	3LN	4,830	5,510	6,840	7,360	8,940	10,420
4LJ-1	G	40,240	46,840	61,350	67,490	87,760	108,000
	P	39,560	43,680	50,930	53,480	60,670	66,600
	LP	39,890	44,340	51,830	54,310	60,630	64,990
	LN	39,720	45,820	58,890	69,340	82,480	101,200
	3LN	39,520	43,970	52,100	55,070	63,710	71,190
5PB-14	G	5,900	7,140	9,880	11,040	14,860	18,670
	P	6,070	7,270	9,690	10,640	13,630	16,450
	LP	5,990	7,380	10,530	11,890	16,560	21,520
	LN	5,970	7,450	11,000	12,610	18,540	25,480
	3LN	6,000	7,230	9,870	10,960	14,620	18,380
5QA-1	G	13,720	16,550	22,770	25,400	34,090	42,770
	P	14,090	17,180	23,740	26,420	35,100	43,600
	LP	13,980	17,140	24,170	27,180	37,380	48,040
	LN	13,930	17,320	25,360	29,000	42,300	57,750
	3LN	13,920	17,170	24,700	28,050	40,030	53,590

TABLE 4. (*Continued*)

Station No.	Distribution	Cumulative Probability, %					
		20	10	2	1	0.1	0.01
2GD-1*	NORMAL	591	652	760	798	904	991
	LN	584	666	838	909	1,140	1,380
	3LN	585	658	803	859	1,040	1,200
2GD-1†	NORMAL	1,550	1,810	2,270	2,430	2,890	3,270
	LN	1,460	1,890	2,980	3,490	5,470	7,910
	3LN	1,460	1,820	2,640	3,000	4,270	5,670

* Annual flow in cfs.
† April flow in cfs.
All other data are floods in cfs.
G, Gumbel;
P, Pearson type 3;
LP, log-Pearson type 3;
LN, lognormal;
3LN, 3-parameter lognormal.

tation of the coefficient of skewness. When computed from limited data this measure, whether of the original data or of their logarithms, may not be reliable.

Thus it is suggested that the 3-parameter lognormal distribution affords a mean of adequately analyzing the hydrologic data. Its application is simple. It can be easily applied with or without the use of computers. The method is directly applicable if the skewness can be reliably computed. An immediate check on the applicability of the method is provided by the straight line plot of the reduced variable ($x - a$) on the lognormal probability paper.

CONCLUSIONS

1. The 3-parameter lognormal distribution is a general skew distribution and can be successfully used for analyzing hydrologic data.

2. The different methods of flow frequency analysis yield different results, especially at low probabilities.

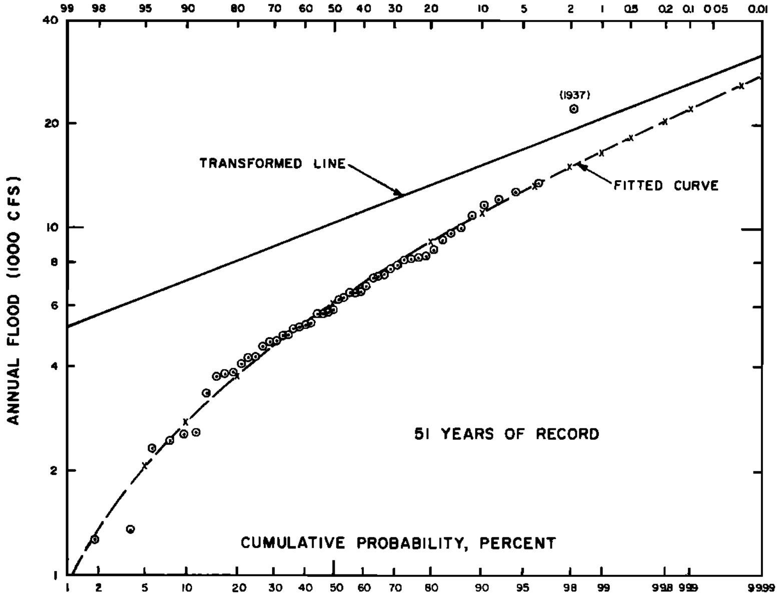

Fig. 1. Distribution of flood flows of Thames River near Ealing, station 2GD-1.

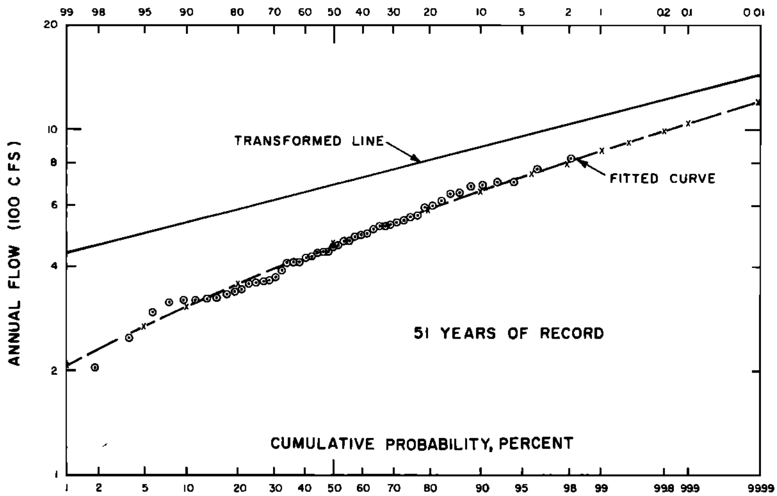

Fig. 2. Distribution of annual flows of Thames River near Ealing, station 2GD-1.

3. The Gumbel distribution is a special case of the 3-parameter lognormal distribution. Any straight line on the Gumbel probability paper can be transformed into a straight line on the lognormal probability paper by the 3-parameter lognormal distribution.

4. The parameter a as computed by the formula given herein using the median, the mean, and the standard deviation of the original data has given reliable results in the analysis of flood flows, annual flows, and monthly flows.

5. The sample of 10 stations used in this study all exhibited negative skewness for the logarithms of the data. The lognormal distribu-

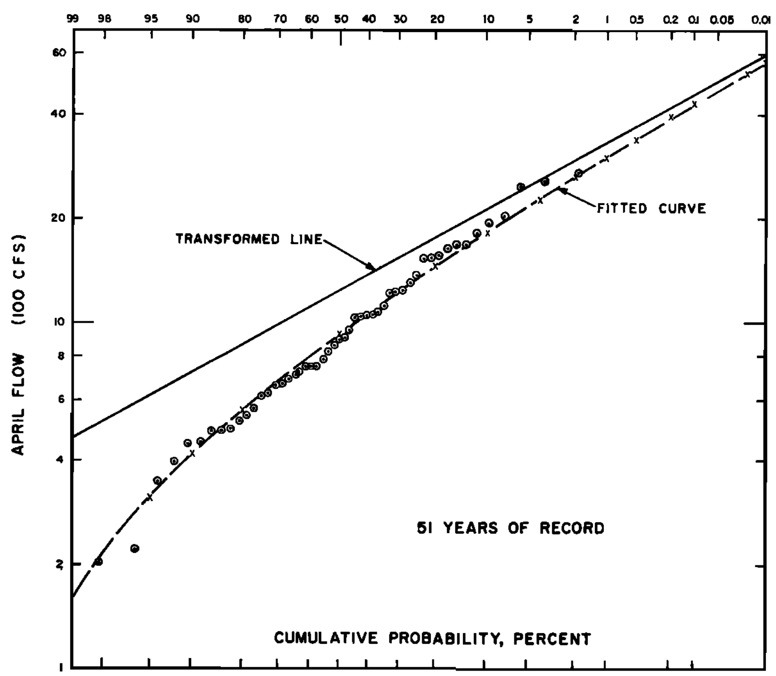

Fig. 3. Distribution of April flows of Thames River near Ealing, station 2GD-1.

tion, which assumes the skewness of the logarithms equal to zero, has therefore given high estimates.

NOTATION

- G, Gumbel distribution;
- K, Pearson type 3 coordinate;
- LN, lognormal distribution;
- LP, log-Pearson type 3 distribution;
- M, ratio of the median to the mean of x;
- P, Pearson type 3 distribution;
- R, ratio of parameter a to the mean of x;
- T, standard normal deviate;
- $3LN$, 3-parameter lognormal distribution;
- a, parameter of the 3-parameter lognormal distribution;
- g_x, coefficient of skewness of x;
- $g_{(x-a)}$, coefficient of skewness of $(x - a)$;
- $\langle \log x \rangle$, the mean of the log x series;
- $\langle \log (x - a) \rangle$, the mean of the log $(x - a)$ series;
- m, order number of events in a descending series;
- n, total number of events;
- $p\%$, cumulative probability, percentage;
- s_x, standard deviation of x;
- $s_{\log x}$, standard deviation of log x;
- $s_{\log (x-a)}$, standard deviation of log $(x - a)$;
- v_x, coefficient of variation of x;
- $v_{(x-a)}$, coefficient of variation of $(x - a)$;
- x, a random variable, flow magnitude;
- $\langle x \rangle$, the mean of x;
- y, logarithmic random variable;
- α, ratio of the parameter a to the population mean of x;
- β, ratio of the population median to the population mean of x;
- γ, population coefficient of variation;
- ζ_x, population median of x;
- ζ_y, population median of y;
- μ_x, population mean of x;
- μ_y, population mean of y;
- $\log \xi$, mean of the logarithmic series;
- σ_x^2, population variance of x;
- σ_y^2, population variance of y;
- σ_x, population standard deviation of x;
- σ_y, population standard deviation of y.

APPENDIX: RELATION BETWEEN σ_y AND $v_{(x-a)}$

In the lognormal distribution [*Chow*, 1954]

$$v_x = (e^{\sigma_y^2} - 1)^{1/2} \quad \text{(A1)}$$

If σ_y is small,

$$v_x = \sigma_y \quad \text{(A2)}$$

In this relation $y = \log_e x$. If $y = \log_{10} x$, then

$$v_x = 2.3026\, \sigma_y \quad \text{(A3)}$$

Thus if v_x is small, then σ_y is small, and its higher powers can be neglected. Also if the parameter a is negative, which is generally the case, $v_{(x-a)}$ would be less than v_x. Therefore in the 3-parameter lognormal distribution the omission of higher than second power terms of σ_y is still more valid. This can be seen from Table 2 where $s_{\log (x-a)}$ values are quite small.

Acknowledgment. Grateful acknowledgment is made to Manuel A. Benson for his comments on an earlier draft of this paper.

REFERENCES

Beard, L. R., Use of interrelated records to simulate streamflow, *J. Hydraul. Div., Amer. Soc. Civil Eng., 91*(HY5), 13–22, 1965.

Benson, M. A., Uniform flood-frequency estimating methods for federal agencies, *Water Resour. Res., 4*(5), 891–908, 1968.

Chow, V. T., The log-probability law and its engineering applications, *Amer. Soc. Civil Eng. Proc. Pap. 536, 80*, 1–25, 1954.

Hald, A., *Statistical Theory with Engineering Applications*, ch. 7, John Wiley, New York, 1962.

Interagency Committee on Water Resources, Methods of flow frequency analysis, *Bull. 13*, Subcommittee on Hydrology, Government Printing Office, Washington, D. C., April 1966.

Matalas, N. C., Probability distribution of low flows, *U. S. Geol. Surv. Prof. Pap. 434-A*, A1–A27, 1963.

Matalas, N. C. Mathematical assessment of synthetic hydrology, *Water Resour. Res., 3*(4), 937–945, 1967.

Matalas, Nicholas C., and Manuel A. Benson, Note on the standard error of the coefficient of skewness, *Water Resour. Res., 4*(1), 204–205, 1968.

Panchang, G. M., Improved precision of future high floods, International Symposium on Floods and Their Computation, Leningrad, August 15–22, 1967, 1-11, 1967.

Panchang, G. M., and V. P. Aggarwal, Peak flow estimation by method of maximum likelihood, *Tech. Memo. HL02*, Government of India, Central Water and Power Research Station, Poona, India, March 1962.

Thöni, Hanspeter, A table for estimating the mean of a lognormal distribution, *J. Amer. Statist. Ass., 64*(326), 632–636, 1969.

(Manuscript received September 2, 1969; revised November 25, 1969.)