THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 55, No. 8, October 1976
Printed in U.S.A.

Rain-Rate Distributions and Extreme-Value Statistics

By S. H. LIN

(Manuscript received March 17, 1976)

An extension of a method of Chen and Lahlum[1] for estimating the distributions of high rain rates is described. Through application of the statistical theory of extremes to available yearly maximum rain-rate data, a reasonably accurate distribution is obtained. The calculated results agree well with previously obtained 20-year data. Application of this method to Weather Bureau Rainfall Intensity-Duration-Frequency Curves has yielded 50-year distributions of 5-minute rain rates for 36 locations in the western United States; others can be similarly obtained. Such long-term rain-rate distributions are valuable for microwave radio-path engineering, especially in the western United States where shorter-term data sources are inadequate.

I. INTRODUCTION

References 2 and 3 describe a procedure for obtaining long-term ($\geqq 20$ years) distributions of 5-minute rain rates from data published by the National Climatic Center.[4] Such distributions have been obtained for 202 locations in the eastern and midwestern United States and applied to path engineering of 11-GHz radio links.

However, the excessive short duration rainfall data[4] on which these distributions are based contain only rainfalls that exceed an excessive rainfall threshold defined by the National Climatic Center.[2,3] For example, the threshold is 75 mm/hr for 5-minute intervals. In low rain-rate areas, such as Oregon and Washington, almost all rainfalls do not exceed the threshold and, hence, are not included in the excessive short-duration rainfall data. For example, at Spokane, Washington, the 5-minute rain rate exceeded the 75 mm/hr threshold only once in the 20-year period from 1953 to 1972. This data source, therefore, is an unsatisfactory basis for radio-path engineering in such areas. On the other hand, processing other longer-term data—say 50 years—is tedious and costly. This has motivated the search for an alternative method. Fortunately, the statistical behavior of the extremes of a random variable has been extensively investigated.[5–10] This paper describes a method for obtaining distributions of high rain rates by applying this theory to the yearly maximum rain-rate data published by the National Climatic Center.

In an unpublished work, Chen and Lahlum[1] have applied the theoretical distribution of yearly maximum 5-minute rain rates and an empirical extrapolation to obtain the rain-rate distribution in the range of interest to radio-path engineering. In this paper, we extend Chen and Lahlum's method by incorporating the theoretical distributions of yearly kth largest 5-minute rain rates for k ranging from 1 to 12. The application of the higher-order statistics of extremes eliminates the need for empirical extrapolation.

In this paper, a 5-minute rain rate corresponds to the average value of the randomly varying rain rate in a 5-minute interval and is calculated as $\Delta H/\tau$, where ΔH is the 5-minute accumulated depth of rainfall and $\tau = 1/12$ hour $= 5$ minutes. For illustration, only the statistics of 5-minute rain rates are discussed in this report. The method is also applicable to other integration times.

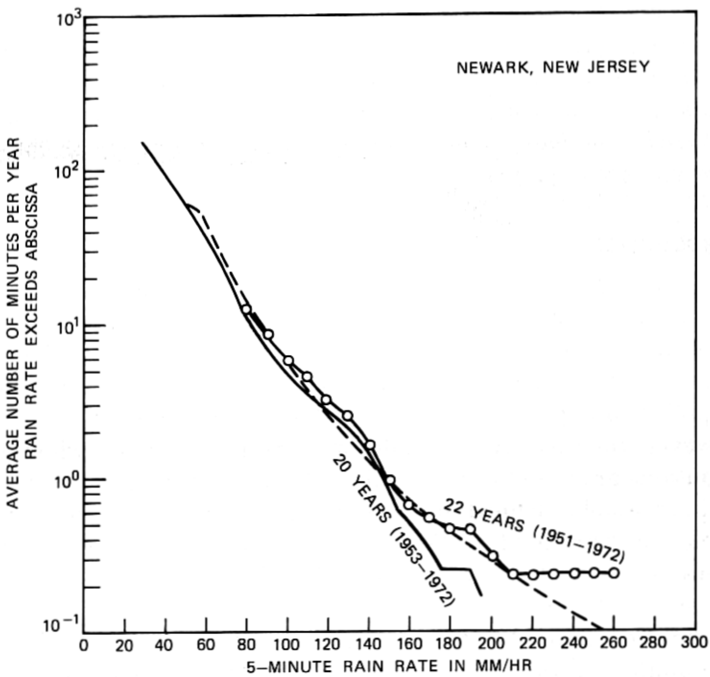

Fig. 1—Comparison of 20-year rain-rate distribution calculated by extreme statistics method (dashed line) with 20- and 22-year data (solid lines) for Newark, New Jersey. The difference between the 20- and 22-year data also indicates the instability of high rain-rate statistics.

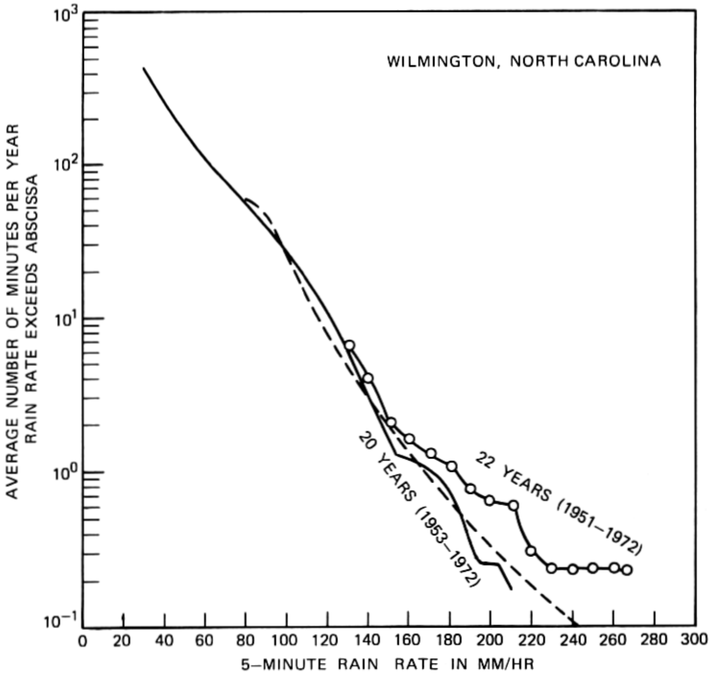

Fig. 2—Comparison of 20-year rain-rate distribution calculated by extreme statistics method (dashed line) with 20- and 22-year data (solid lines) for Wilmington, North Carolina. The difference between the 20- and 22-year data also indicates the instability of high rain-rate statistics.

II. THE STATISTICS OF EXTREMES

Let R be the randomly varying 5-minute rain rate and R_k be the kth largest 5-minute rain rate in a given year. In other words, R_1 is the yearly maximum 5-minute rain rate, R_2 is the yearly second largest 5-minute rain rate, etc. The value of R_k varies from year to year, and the probability distribution of R_k is the subject of the statistics of the kth extreme.

Many sets of rain-rate data[11–15] indicate that rain-rate distributions in the moderate and low rain-rate region can be closely approximated by the lognormal distribution. In the tail region, the time bases are usually insufficient to yield stable results for testing the lognormal hypothesis. Figures 1 and 2 show instability at extreme values occurring in 20-year time bases. We will assume that the rain-rate distributions are lognormal and proceed to show that the calculated distributions of extreme rain rates agree well with the data as displayed in Figs. 1 through 9.

Let

$$x = \ln R, \tag{1}$$

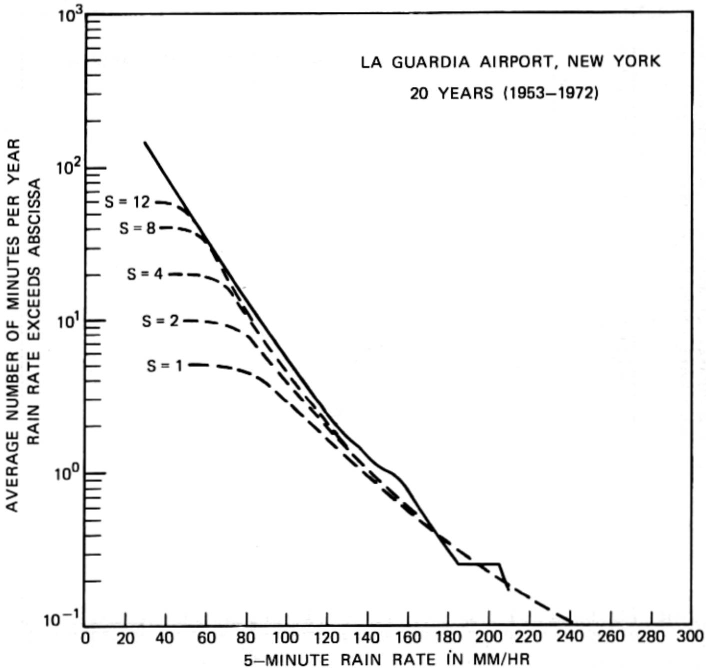

Fig. 3—Comparison of 20-year rain-rate distribution calculated by extreme statistics method (dashed lines) with the 20-year data (solid line) at La Guardia Airport, New York, New York.

and

$$x_k = \ln R_k. \tag{2}$$

The lognormal hypothesis for R is equivalent to the statement that x is (approximately) normal. The distribution of the kth extreme, x_k, as derived by Cramer[8] is

$$P(x_k \geqq X) = 1 - e^{-e^{-y}} \sum_{N=0}^{k-1} \frac{e^{-Ny}}{N!}, \tag{3}^*$$

where

$$y = \alpha(X - U) \tag{4}$$

and is called the reduced variate. In this expression, α and U are scale and location parameters, respectively, and are related to the sample mean and sample standard deviation of x_1. Notice that the distribution, $P(x_k \geqq X)$, for any k, is completely determined by the two parameters α and U. These two parameters can be calculated from the measured

* The cumulative distribution function (3) is obtained by integrating the probability-density function derived by Cramer in Ref. 8.

yearly maximum rain-rate data. Let

$$R_1(j),\ j = 1, 2, 3, \cdots, M \tag{5}$$

be the measured yearly maximum 5-minute rain rate in M years of measurements. For example, $R_1(7)$ represents the yearly maximum 5-minute rain rate observed in the seventh year of an M year experiment. Let

$$x_1(j) = \ln\left[R_1(j)\right] \tag{6}$$

be the yearly maximum value of x in the jth year. From the measured data of $x_1(j)$, $j = 1, 2, 3, \cdots, M$, we can obtain an approximate distribution of x_1. The parameters α and U can be estimated by fitting the theoretical $P(x_1 \geqq X)$ to the measured data. Gumbel[6] has shown that a least-square fit of $P(x_1 \geqq X)$ to the data leads to the following formulas for calculating α and U:

$$\alpha = \frac{\sigma_z}{\sigma_x}, \tag{7}$$

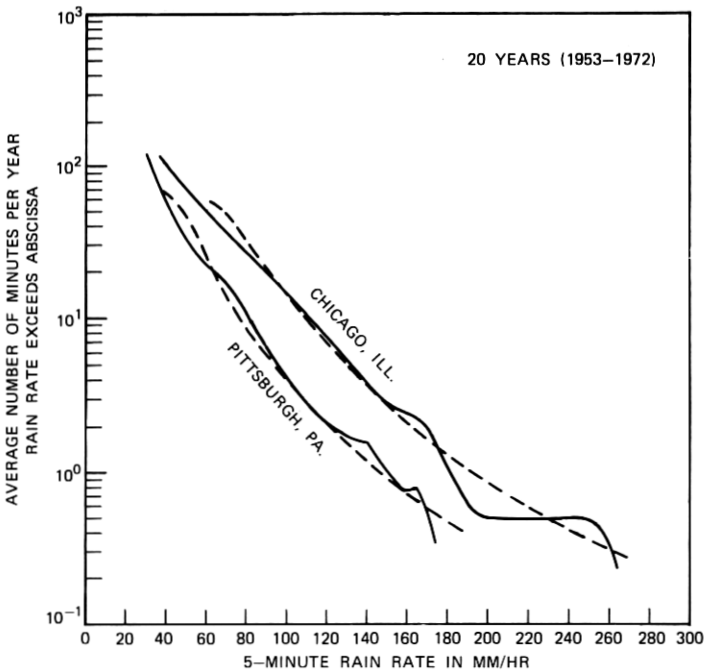

Fig. 4—Comparison of 20-year rain-rate distribution calculated by extreme statistics method (dashed line) with 20-year data (solid line) for Pittsburgh, Pennsylvania, and Chicago, Illinois.

and

$$U = \bar{x}_1 - \frac{\bar{z}}{\alpha}, \tag{8}$$

where

$$\bar{x}_1 = \frac{1}{M} \sum_{j=1}^{M} x_1(j) \tag{9}$$

is the sample mean of x_1,

$$\overline{x_1^2} = \frac{1}{M} \sum_{j=1}^{M} [x_1(j)]^2, \tag{10}$$

$$\sigma_x = \sqrt{\overline{x_1^2} - \bar{x}_1^2} \tag{11}$$

is the sample standard deviation of x_1,

$$Z(j) = -\ln\left(-\ln \frac{j}{M+1}\right), \tag{12}$$

$$\bar{Z} = \frac{1}{M} \sum_{j=1}^{M} Z(j), \tag{13}$$

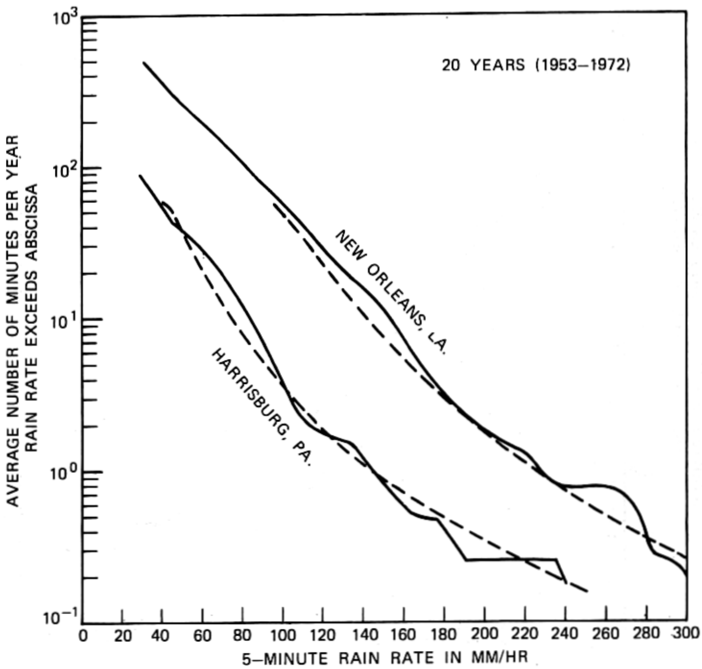

Fig. 5—Comparison of 20-year rain-rate distribution calculated by extreme statistics method (dashed line) with 20-year data (solid line) for Harrisburg, Pennsylvania, and New Orleans, Louisiana.

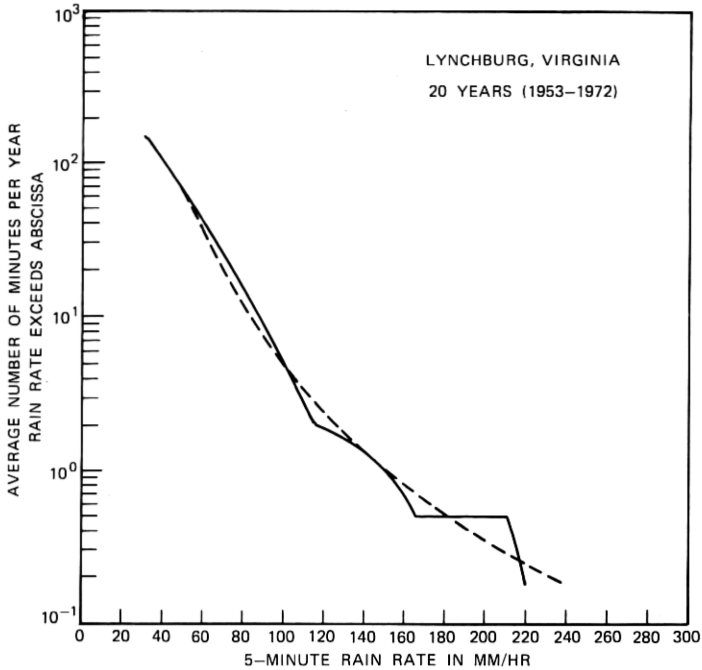

Fig. 6—Comparison of 20-year rain-rate distribution calculated by extreme statistics method (dashed line) with 20-year data (solid line) for Lynchburg, Virginia.

$$\overline{Z^2} = \frac{1}{M} \sum_{j=1}^{M} [Z(j)]^2, \tag{14}$$

and

$$\sigma_z = \sqrt{\overline{Z^2} - \bar{Z}^2}. \tag{15}$$

Thus, we have all the necessary formulas for calculating $P(x_k \geqq x)$. To obtain the rain-rate distribution, we substitute (2) into (3) to yield

$$P(R_k \geqq r) = 1 - e^{-e^{-y}} \sum_{N=0}^{k-1} \frac{e^{-Ny}}{N!}, \tag{16}$$

where

$$y = \alpha[(\ln r) - U]. \tag{17}$$

Therefore, the time that R_k will exceed the threshold r, on long-term average, is

$$\begin{aligned} T(R_k > r) &= \tau \times P(R_k \geqq r) \\ &= \tau \times \left\{1 - e^{-e^{-y}} \sum_{N=0}^{k-1} \frac{e^{-Ny}}{N!}\right\}, \end{aligned} \tag{18}$$

where $\tau = 5$ minutes is the rain-gauge integration time. Furthermore, in any given year, R_k and R_l will never occur in the same 5-minute

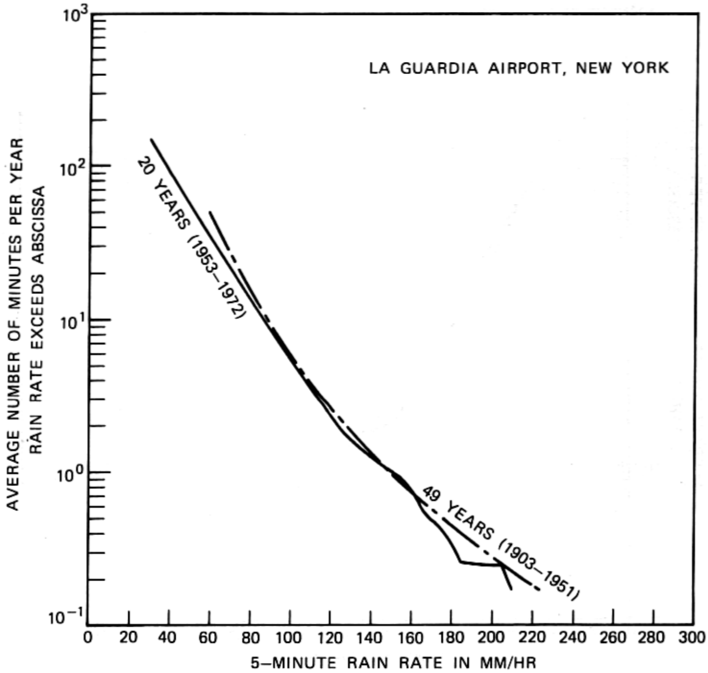

Fig. 7—Comparison of 49-year (1903–1951) distribution of 5-minute rain rates calculated by extreme statistics method with measured 20-year (1953–1972) data at La Guardia Airport, New York, New York.

interval if $k \neq l$. This means $T(R_k \geqq r)$ for various order k can be summed to yield an approximation to the original rain-rate distribution in the extremal region; i.e.,

$$T(R \geqq r) \simeq \sum_{k=1}^{S} T(R_k \geqq r) \tag{19}$$

for high rain rates.

The only input required for the calculation is the yearly maximum 5-minute rain rates,

$$R_1(j),\ j = 1, 2, 3, \cdots, M,$$

which can be obtained from National Climatic Center publications.[4]

III. COMPARISON OF CALCULATED AND MEASURED RESULTS

Figures 1 through 6 display the comparison of the distributions of high rain rates calculated via statistics of extremes from the data for eight locations. The number of years M is 20 (from 1953 to 1972). In these figures, the solid lines represent the data obtained by the method described in Refs. 2 and 3, whereas the dashed lines are the distribution calculated by the theory of extremes.

In Fig. 3, the value of S in eq. (19) is varied from 1 to 12. For $S = 1$, $T(R_1 \geqq r)$ is the distribution of yearly maximum 5-minute rain rate and is approximately equal to $T(R \geqq r)$ *only* in the extremal region (beyond 160 mm/hr). As r decreases, $T(R_1 \geqq r)$ deviates significantly from $T(R \geqq r)$, limiting at the 5-minute-per-year level as r approaches zero. The basis for this saturation is that there is only *one* yearly maximum 5-minute rain rate (with 5-minute duration by definition) in any given year. It is obvious that the yearly maximum 5-minute rain rate can exceed any threshold by no more than 5 minutes per year. Similarly, for $S = 2$, $T(R_1 \geqq r)$ plus $T(R_2 \geqq r)$ is limited to a 10-minute-per-year level as r approaches zero. However, Fig. 3 shows that the applicable range of approximation (19) increases rapidly with S.

For engineering terrestrial radio paths, we are interested in the range of rain-rate distributions below 50 minutes per year, because a single radio hop outage exceeding 50 minutes per year is considered

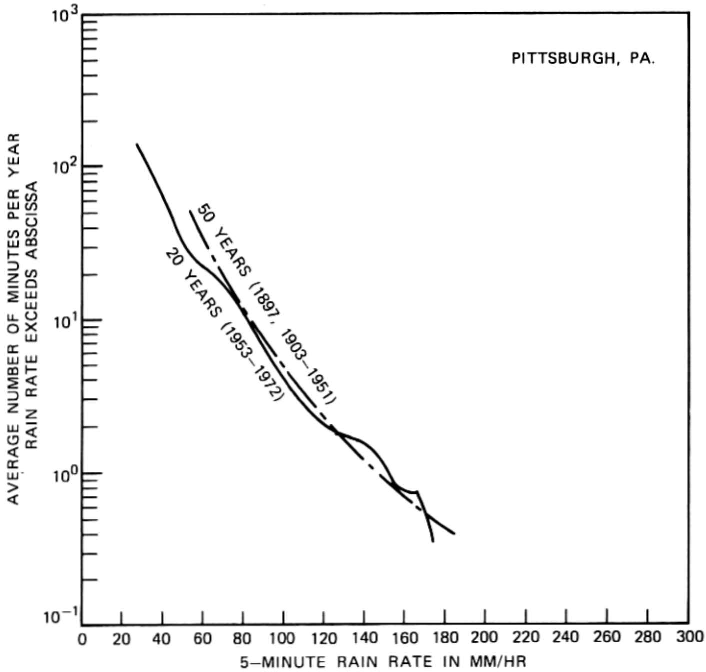

Fig. 8—Comparison of 50-year (1897, 1903–1951) distribution of 5-minute rain rates calculated by extreme statistics method with measured 20-year (1953–1972) data for Pittsburgh, Pennsylvania.

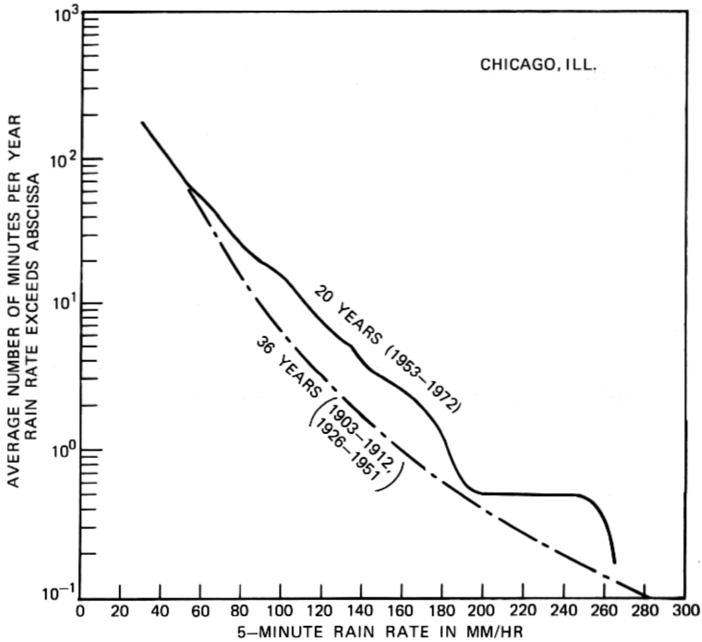

Fig. 9—Comparison of 36-year (1903–1912, 1926–1951) distribution of 5-minute rain rates calculated by extreme statistics method with measured 20-year (1953–1972) data at Midway Airport, Chicago, Illinois.

undesirable. Figure 3 shows that using $S = 12$ in eq. (19) includes the range of interest to terrestrial radio-path engineering.

Figures 1 to 6 show similar close agreement between the calculated results and the data.

The advantage of using the extreme statistics method is its simplicity; i.e., it is much easier to obtain the extreme-statistics results from the set of 20 maximum yearly rain rates than to obtain the other displayed results which require analyzing all heavy rainfalls in each year.

IV. FIFTY-YEAR DISTRIBUTIONS OBTAINED FROM RAINFALL INTENSITY-DURATION-FREQUENCY CURVES

Rainfall intensity-duration-frequency curves published by the Weather Bureau[16] were obtained from approximately 50 years (1900–1950) of rain-rate data processed in accordance with the statistical theory for distribution of yearly maximum rain rate by the Gumbel method.[6] In other words, these curves represent $P(R_1 \geqq r)$, where R_1 is the yearly maximum rain rate. The return period $Q(R_1 \geqq r)$, which is the reciprocal of $P(R_1 \geqq r)$, is labeled on each curve; i.e.,

$$Q(R_1 \geqq r) = \frac{1}{P(R_1 \geqq r)} \text{ years.} \tag{20}$$

These curves cover the range-of-return period from 2 to 100 years and rain-gauge integration time τ (i.e., the duration) from 5 minutes to 24 hours. We considered only the 5-minute rain-rate statistics in this report.

Since these curves represent $P(R_1 \geqq r)$, the two parameters α and U in eq. (17) can be estimated by fitting the theoretical $P(R_1 \geqq r)$ to two points on the intensity-duration-frequency curves. Once α and U are obtained, we can calculate a distribution of high rain rates by eqs. (18) and (19). However, one adjustment is needed in this process. Gumbel[6] has shown that as the number of years M approaches infinity, $\bar{Z}$ and σ_z in eqs. (13) and (15) approach the following asymptotic values

$$\lim_{M \to \infty} \bar{Z} = \gamma = \text{Euler's Constant} \simeq 0.5772 \tag{21}$$

$$\lim_{M \to \infty} \sigma_z = \frac{\pi}{\sqrt{6}}. \tag{22}$$

The corresponding asymptotic forms for α and U are

$$\lim_{M \to \infty} \alpha = \alpha_\infty = \frac{\pi}{\sqrt{6} \cdot \sigma_x}, \tag{23}$$

and

$$\lim_{M \to \infty} U = U_\infty = \bar{x}_1 - \frac{\gamma}{\alpha_\infty}. \tag{24}$$

The rainfall intensity-duration-frequency curves were obtained using the asymptotic values, α_∞ and U_∞, even though the number of years M are 50 or less. Such approximations introduce slight errors and can be corrected by the following relationships among α_∞, U_∞, α, and U:

$$\alpha = \alpha_\infty \cdot \sigma_z \cdot \frac{\sqrt{6}}{\pi} \tag{25}$$

$$U = U_\infty + \frac{1}{\alpha_\infty}\left[\gamma - \frac{\bar{Z}}{\sigma_z} \cdot \frac{\pi}{\sqrt{6}}\right]. \tag{26}$$

To relate theoretical $P(R_1 \geqq r)$ with the intensity-duration-frequency data, we combine eq. (16) and (20) to give

$$Q(R_1 \geqq r) = \frac{1}{1 - e^{-e^{-y}}}, \tag{27}$$

where

$$y = \alpha_\infty[(\ln r) - U_\infty]. \tag{28}$$

The two parameters, α_∞ and U_∞, can be determined by equating (27) to two sets of data, say (r_a, Q_a) and (r_b, Q_b), read from the intensity-duration-frequency curves. It is easily shown that the relationships are

$$\alpha_\infty = \frac{A_a - A_b}{\ln r_a - \ln r_b}, \tag{29}$$

and

$$U_\infty = \frac{A_a \ln r_b - A_b \ln r_a}{A_a - A_b}, \tag{30}$$

where

$$A_a = -\ln\left[\ln \frac{Q_a}{Q_a - 1}\right], \tag{31}$$

and

$$A_b = -\ln\left[\ln \frac{Q_b}{Q_b - 1}\right]. \tag{32}$$

By substituting α_∞ and U_∞ into (25) and (26), we obtain α and U for the 50-year data. Substituting α and U into (18) and (19) gives the 50-year distribution of 5-minute rain rates.

The time bases (i.e., M) in the intensity-duration-frequency curves are mostly 50 years or less. However, several locations have time bases much shorter than 50 years. For example, the time bases are 18, 16, and 17 years for Mt. Tamalpais, California; Tonopah, Nevada; and Yakima, Washington, respectively. Due to this limitation, we have chosen $Q_a = 2$ years and $Q_b = 10$ years for calculations of α and U. Since Q_a and Q_b are fixed, we then need only three numbers: M, r_a, and r_b for each location, read from the intensity-duration-frequency curves to calculate the rain-rate distribution.

For example, for New York City, the three numbers are:

$$M = 49 \text{ years } (1903\text{–}1951)$$
$$r_a = 4.4 \text{ inches/hr} = 111.8 \text{ mm/hr}$$
$$r_b = 6.5 \text{ inches/hr} = 165 \text{ mm/hr}.$$

Substituting r_a and r_b into eqs. (29) and (30) yields

$$\alpha_\infty = 4.828$$
$$U_\infty = 4.64.$$

Substituting M into eqs. (12) through (15) and α_∞ and U_∞ into (25) and (26) gives

$$\alpha = 4.363$$
$$U = 4.63.$$

The 49-year (1903–1951) distribution of 5-minute rain rate calculated from this α, U pair for New York City is very close to the 20-year

(1953–1972) data, as shown in Fig. 7. Figure 8 shows similar close agreement between 50-year and 20-year distributions in Pittsburgh, Pennsylvania.

However, Fig. 9 shows an appreciable difference between 36-year and 20-year distributions for Chicago, Illinois. This appreciable difference, the irregular shape of the 20-year distribution in Fig. 9 and the instability noted in Figs. 1 and 2 indicate that a 20-year time base with a single rain-gauge measurement may not be sufficient to guarantee a stable distribution for extremely high rain rates. If more stable results are required, the 20- and 50-year data may be combined to give a 70-year distribution.

The calculated curves (dashed lines) in Figs. 1 through 6 are based on 20-year data from 1953 to 1972 in Ref. 4, whereas the calculated curves (dash-dot lines) in Figs. 7 through 9 are based on approximately 50 years of data in Ref. 16.

V. CONCLUSION

A method has been described for calculating the distribution of high rain rates by applying the statistical theory of extremes to the available yearly maximum 5-minute rain-rate data. Figures 1 through 9 show that the calculated distributions agree closely with the data in the heavy-rain region of interest to radio-path engineering. The virtue of this method is that only yearly maximum rain-rate data are required to generate satisfactory results for radio-path engineering. The rainfall intensity-duration-frequency curves[16] provide approximately 50 years (1900–1950) of such data for 203 locations in the United States. Furthermore, a new publication on "Maximum Short-Duration Precipitation,"[17] for approximately 300 U. S. locations issued annually by the National Climatic Center since 1973, provides additional yearly maximum rain-rate data. Therefore, long-term ($\geqq$50 years) distributions of high rain rates for 203 U. S. locations can easily be obtained by this method.

VI. ACKNOWLEDGMENTS

The author is grateful to W. Y. S. Chen and R. L. Lahlum for valuable discussion on the application of statistics of extremes.

REFERENCES

1. W. Y. S. Chen and R. L. Lahlum, unpublished work.
2. W. Y. S. Chen, "A Simple Method for Estimating 5-Minute Rain Rate Distributions Based on Available Climatological Data," B.S.T.J., *55*, No. 1 (January 1976), pp. 129–134.
3. S. H. Lin, "Dependence of Rain Rate Distribution on Rain Gauge Integration Time," B.S.T.J., *55*, No. 1 (January 1976), pp. 135–141.

4. "Climatological Data, National Summary," annual issues since 1950, U. S. Department of Commerce, National Oceanic and Atmospheric Administration, National Climatic Center, Asheville, North Carolina.
The Excessive Short Duration Rainfall Data prior to 1950 are published in the Monthly Weather Review, the U. S. Meteorological Yearbook (last published for the period 1943–1949), and the Report of the Chief of the Weather Bureau (last published for 1931).
5. R. A. Fisher and L. H. C. Tippett, "Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample," Proc. Cambridge Phil. Soc., *24* (1928), pp. 180–190.
6. E. J. Gumbel, *Statistical Theory of Extreme Values and Some Practical Applications*, National Bureau of Standards, Applied Mathematics series No. 33 (February 12, 1954), pp. 15, 16, and 28.
7. E. J. Gumbel, *Statistics of Extremes*, New York: Columbia University Press, 1958.
8. H. Cramer, *Mathematical Methods of Statistics*, Princeton University Press, 1957, Section 28.6, pp. 370–378.
9. H. Cramer and M. R. Leadbetter, *Stationary and Related Stochastic Processes*, New York: John Wiley, 1967, Section 12.3, pp. 271–272.
10. G. S. Watson, "Extreme Values in Samples From M-Dependent Stationary Stochastic Processes," Ann. Math. Stat., *25* (1954).
11. S. H. Lin, "A Method for Calculating Rain Attenuation Distributions on Microwave Paths," B.S.T.J., *54*, No. 6 (July-August 1975), pp. 1051–1086, Figs. 10–13.
12. L. Hansson, "General Characteristics of Rain Intensity Statistics in the Stockholm Area" (Sweden), TELE, No. 1 (1975), pp. 43–48.
13. L. Hansson, "General Characteristics of Rain Intensity Statistics in the Gothenburg Area," Report USR 75 012, Central Administration of Swedish Telecommunications, January 1975, S-12386, Farsta, Sweden.
14. S. H. Lin, "Statistical Behavior of Rain Attenuation," B.S.T.J., *52*, No. 4 (April 1973), pp. 557–581.
15. P. T. Schickedanz, "Theoretical Frequency Distributions for Rainfall Data," International Symposium on Probability and Statistics in the Atmospheric Sciences, June 1–4, 1971, Honolulu, Hawaii, Sponsored by American Meteorological Society and Co-sponsored by the World Meteorological Organization, Boston, Massachusetts. Preprints of Symposium Papers, pp. 131–136.
16. "Rainfall Intensity-Duration-Frequency Curves," U. S. Department of Commerce, Weather Bureau, Technical Paper No. 25, Washington, D. C. (December 1955).
17. "Maximum Short Duration Precipitation," contained in "Climatological Data, National Summary," annual issues since 1973, National Climatic Center, Asheville, North Carolina.