TORNADO PROBABILITIES

H. C. S. THOM

Office of Climatology, U.S. Weather Bureau, Washington D.C.

Manuscript received July 2, 1963; revised August 7, 1963]

ABSTRACT

The frequency distributions of tornado path width and length are developed using data series from Iowa and Kansas. From these, the distribution of path area is derived. Direction of path and annual frequency are discussed. It is found that all but about 1 percent of Iowa tornadoes had path directions toward the northeast and southeast quadrants. The annual frequency for a group of Iowa counties is found to have a negative binomial distribution indicating that the climatological series is formed from a Polya stochastic process. This resembles the situation for other types of storms where the events tend to cluster. A new map of annual frequency for the United States is presented for the period 1953–62, during which it is believed tornado observation was fairly stable. The expected value of tornado area is derived from the area distribution. From this and the annual frequency, the probability of a tornado striking a point is found.

1. INTRODUCTION

There have been a large number of studies of tornado climatology, most of which have been simply counts of tornadoes for various areas and time periods. Asp [1] lists 78 references, a few of which are not climatological in nature; not all references have been listed. Many of these studies have recognized the possible incompleteness of the frequency series and the difficulties of observation, but little could be done to correct this deficiency. So far as is known, none of these studies made a direct attack on the problem of tornado probability, which is the object of the present study.

In 1945, William F. Kuffel, then of the Dubuque Fire Marine Insurance Company, asked the writer to develop a system of limiting the loss from a single tornado in a given region for the purpose of preventing liabilities from exceeding reserve funds. This resulted in a limited study for several Iowa counties [2] in which the direction frequency and path length and width distributions were discussed. From this, a directed standard path was devised within whose bounds the insured liability could be totaled. If this exceeded a certain limit related to the reserves of the company, the excess could be reinsured with other companies. It should be noted that the occurrence of more than one tornado in the region is still to be taken care of by the ordinary risk of the business which is not well defined in this type of insurance coverage.

By 1957, these ideas had developed further [3], and after mathematical distributions were fitted to the path length and width it was possible to determine the probability of a tornado striking a point. There still remained a bothersome correlation between path length and width which was not easily taken into account in the area distribution. This prevented obtaining a complete solution to the distribution problem. In 1958, Battan [4] presented a simple frequency diagram of path length, but his objective was to study the duration of a tornado, not its probability of occurrence.

In the present study, we introduce distribution theory which provides a better fit to the basic data and makes possible a more satisfactory solution to the area distribution problem. The distribution of annual frequency is also discussed and several comparisons of data are made, together with a number of statistical tests for homogeneity.

2. PATH LENGTH AND WIDTH DISTRIBUTIONS

Since path width and length cannot be negative, zero must be the lower bound of any distribution assumed, although this need not be a greatest lower bound. As with a number of other physical variables, where the true bound is certainly near zero, but cannot be established to be different from zero, it has proven convenient to assume that the distribution has a zero lower bound. Also, in this instance, it would appear that both variables should have a probability density of zero at the origin, for as the path length and width approach their greatest lower bounds, the probability density should approach zero.

In previous studies [3], a gamma distribution was assumed. While it has a zero bound, it need not have a zero probability density at the origin. When fitted to length and width data, both variables gave shape parameter estimates which indicated non-zero densities at the origin. Furthermore, with this function the distribution of area becomes intractable, and above all, the distribution did not fit the data series particularly well.

The log-normal distribution meets all the above theoretical requirements: It may have a zero bound, its probability density is zero at the origin, and the distribution of area is simply a convolution of the logarithmic distributions of path length and width, the correlation between them entering quite naturally.

Let x be the path length in miles, and w be the path width in yards; then, transforming to logarithms, we have

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

and

$$u = \ln w. \tag{2}$$

And inversely

$$x = e^y, \tag{3}$$

and

$$w = e^u. \tag{4}$$

Let $\alpha_1(\ )$ and $\sigma^2(\ )$ be the mean and variance operators respectively. Then, if y is normally distributed, the distribution of x is log-normal, i.e.,

$$dL(x) = \frac{1}{x\sigma(y)\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2(y)}[\ln x - \alpha_1(y)]^2\right\} dx \tag{5}$$

where $L(x)$ is the distribution function or cumulative distribution. Similarly, for w

$$dM(w) = \frac{1}{w\sigma(u)\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2(u)}[\ln w - \alpha_1(u)]^2\right\} dw. \tag{6}$$

It is well known that the mean and variance estimated in the usual fashion are jointly sufficient statistics for estimating $\alpha_1(y)$, $\alpha_1(u)$, $\sigma^2(y)$, and $\sigma^2(u)$. Since functions of sufficient statistics are also sufficient, the optimum property holds for the transformed sample given by equations (3) and (4). Hence, we may work in the logarithms transforming back after estimation is completed.

Three sets of data were analyzed with distributions (5) and (6): Iowa tornadoes for the period 1937–56, Iowa tornadoes for the period 1953–62, a kind of standard period to be discussed later, and Kansas tornadoes for this standard period. Tornadoes with paths longer than 100 mi. and paths wider than 1000 yd. were rejected as doubtful observations. The relative frequency of these is small and, if included, would have little effect on the distributions considering the accuracy of the observations. The fits to the path length are all excellent, none of the three distributions having an absolute departure significant at the 0.05 level on the Kolmogorov-Smirnov distribution. The fits to the path width are also good, there being no absolute departure significant at the 0.01 level. The somewhat greater precision of the path length observations was to be expected since the path width is more difficult to estimate.

The sample means and standard deviations are shown in table 1. Since all of the distribution fits were very good, it was decided to show only those for the Iowa 1953–62 record which consisted of 106 observations of path length and 103 observations of path width. These samples were large enough to confuse the plotting, so only the middle point of each series of repeated values was plotted on figures 1 and 2. While this might seem to make the fit appear better, this criticism can not be important because the fits were shown to be good by the significance test. The theoretical lines on the graphs were obtained from the statistics for Iowa 1953–62 in table 1.

TABLE 1.—*Path statistics (logarithms)*

	Path Length		Path Width	
	$\bar{y}$	$s(y)$	$\bar{u}$	$s(u)$
Iowa 1937–56	1.99	1.25	5.41	1.01
Iowa 1953–62	1.42	1.43	5.15	0.91
Kansas 1953–62	1.32	1.44	4.92	1.12
Means 1953–62	1.37	1.43	5.04	1.02

The variance of the three records of table 1 were compared by Bartlett's test and no significant difference at the 0.05 level was found for either path length or width. However, an analysis of variance on the three records showed significant differences in the means for both path length and width. Since this could be due to poorer observing during the 1937–52 period, it was decided to discard the earlier record for present purposes. t–tests on the Iowa and Kansas means for the 1953–62 record then showed no significant difference at the 0.05 level for either path length or width. An F–test on the variances for path length showed no significant difference at the 0.05 level as the variances are nearly identical. The variances for path width, however, tested significantly different. Inasmuch as the difference is small and the number of degrees of freedom large (103 and 125), and considering the accuracy of the observations, it was decided to average the variances as if they, like the means, were from the same population. The resulting means and standard deviations of path length and width are shown at the foot of table 1.

Although expected superiority of path width and length observations over frequency observations was the basis for the study, it was a surprise to the writer to find an apparently crude observational technique producing such a remarkable agreement between data for Iowa and Kansas. It should not be forgotten, however, that the observational technique still may be biased and methods should be devised for testing this. Data from other areas should also be analyzed to determine whether the tornado characteristics presented here are invariant over larger areas or possibly physically invariant.

3. THE DISTRIBUTION OF PATH AREA

The convolution of the distributions of the logarithms is the transformed distribution of the product. Thus, the distribution of $y+u=v$ is the transformed distribution of

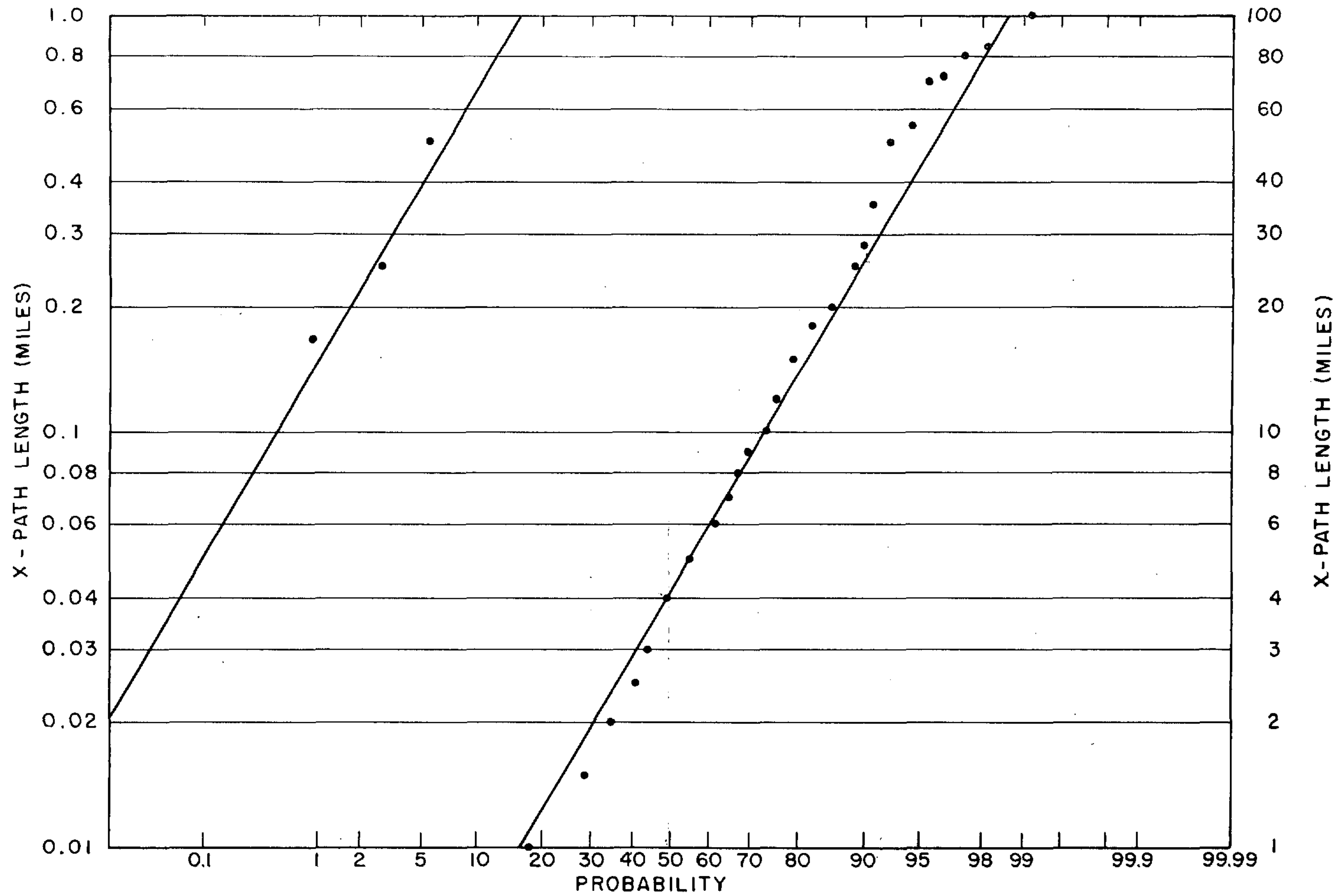

FIGURE 1.—Distribution of path length for Iowa tornadoes during 1953–62.

$xw=z$, the path length times the path width, or the tornado path area in yard-miles. The equations of transformation are similar to those of path length and width, i.e.,

$$v=\ln z \tag{7}$$

and

$$z=e^v.$$

Since y and u are normally distributed, their sum v is also normally distributed, and z has the log-normal distribution

$$dQ(z)=\frac{1}{z\sigma(v)\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma^2(v)}[\log v-\alpha_1(v)]^2\right\}dz. \tag{8}$$

The value of $\alpha_1(v)$ is given easily by the sum of the means of y and u or

$$\alpha_1(v)=\alpha_1(y)+\alpha_1(u) \tag{9}$$

The variance of v is more complicated because the path length and path width are positively correlated. This introduces a covariance term and the variance is then

$$\sigma^2(v)=\sigma^2(y)+\sigma^2(u)+2\rho(y,u)\sigma(y)\sigma(u), \tag{10}$$

where $\rho(y, u)$ is the correlation between the logarithm of path length and width. The correlation coefficient between 96 pairs of path lengths and widths for Iowa was found to be 0.39. Inasmuch as the correlation would be expected to vary less areally than the variances, we use this correlation coefficient with the average path length and width variances of table 1. This gives a sample path area variance of $s^2(v)=4.233$ and standard deviation $s(v)=2.057$. Substituting the sample means from table 1 in equation (9) gives $\bar{v}=6.41$.

We shall need to know the mean value of z. This is given by the expectation operator

$$E(z)=E[e^v]$$
$$=\int_{-\infty}^{\infty} e^v dN(v) \tag{11}$$

where $N(v)$ is a normal distribution. By the moment generating properties of the normal distribution, (11) gives

$$E(z)=\exp\{\alpha_1(v)+1/2\sigma^2(v)\} \tag{12}$$

It is interesting to note that although the expected value

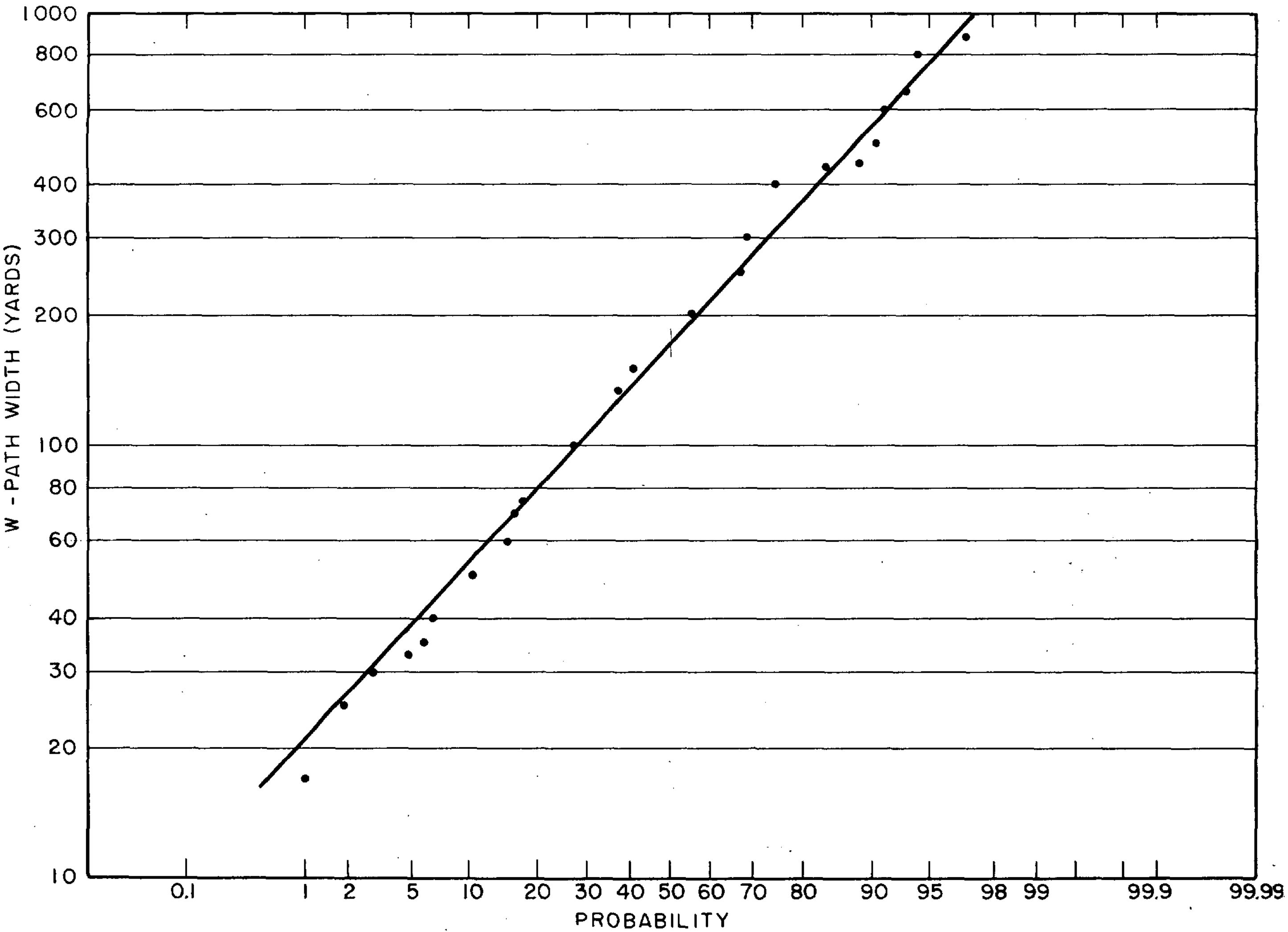

Figure 2.—*Distribution of path width of Iowa tornadoes, 1953–62.*

of a product is not the product of the expectations when the factors are correlated, in case the factors have log-normal distributions the adjustment for the correlation is the covariance term of equation (10) which defines the variance term in (12).

Substituting sample values $\bar{v}$ and $s^2(v)$ into equation (12) gives a mean path area of 4964.8 mi. yd. Dividing by 1760 gives 2.8209 mi.^{2}

4. DIRECTION AND ANNUAL FREQUENCY

It is well known that the preponderant direction of tornado movement is toward the northeast. Our findings from a tabulation of 230 tornadoes in Iowa are in agreement with this. Table 2 shows the distribution by directions. It is seen in the table that about 63 percent of the tornadoes tabulated have path directions toward the northeast, while about 90 percent of them have an easterly component in their paths. For practical purposes, it might be useful to assume that almost all tornadoes have paths with a component toward the east.

We now consider the annual frequency of tornadoes for a central section of Iowa where it is believed that the population mean annual frequency was constant during the period 1916–56. This series consists of the data for Boone, Story, Marshall, Dallas, Polk, and Jasper Counties. To demonstrate the stability of annual frequency, we made a run test on the annual series. The number of runs of years with frequency of one and zero and those with frequency greater than one was 18, which is approximately at the median frequency. This value has a probability of 0.37 of being exceeded on the run distribution, and is therefore clearly not significant at the 0.05 level. The

Table 2.—*Tornado path direction–frequency*

Toward	NW	N	NE	E	SE	S	Total
Number	3	13	145	35	32	2	230
Percent	1.3	5.6	63.1	15.0	13.8	1.2	100

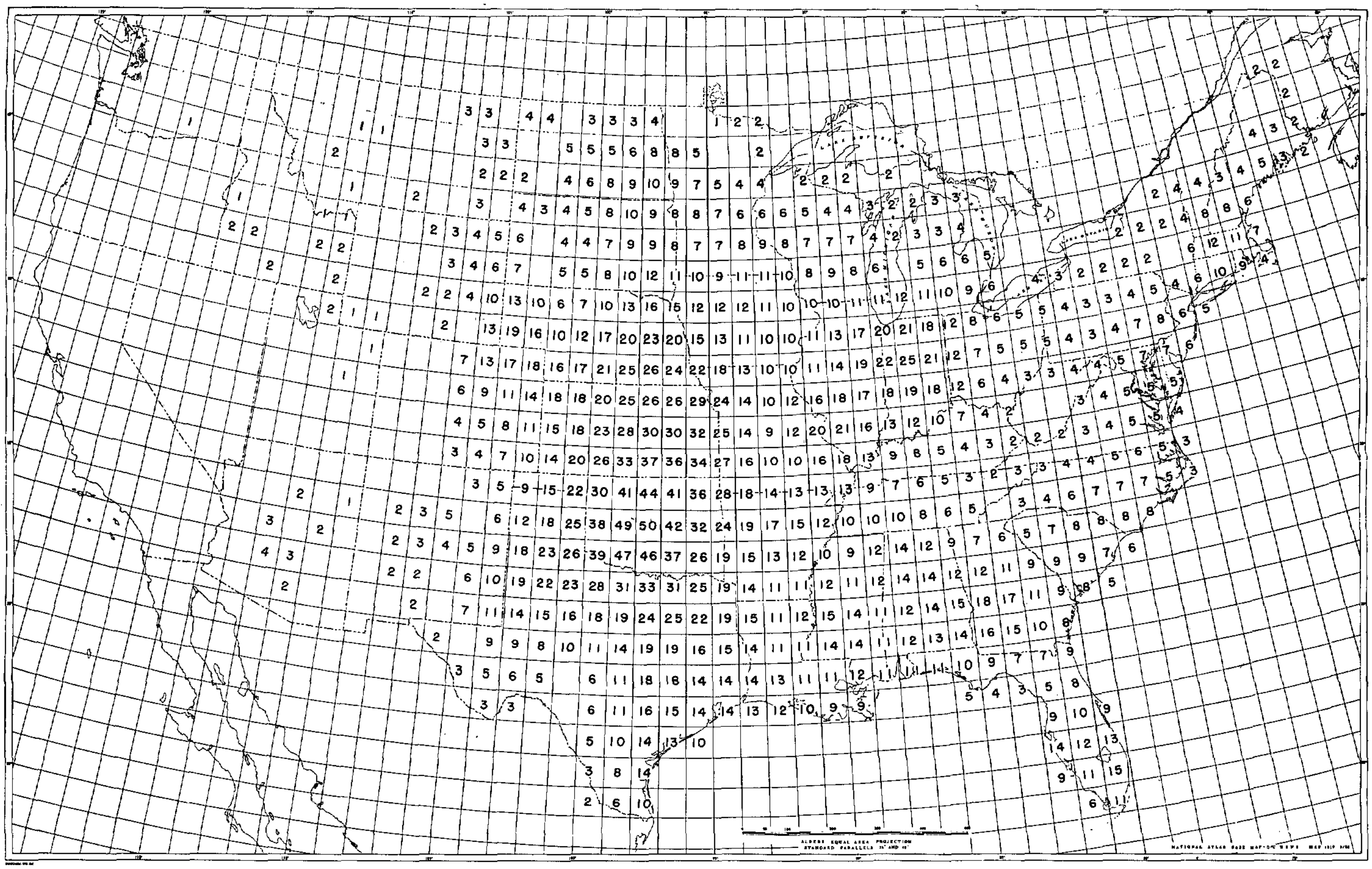

FIGURE 3.—Total frequency of tornadoes, 1953–62

annual frequency series for central Iowa may therefore be treated as a random series.

Since the annual tornado series is discrete, it would appear that some distribution such as that of the Poisson stochastic process would fit the annual frequencies. Poisson-related distributions and their criteria were discussed in [5] and [6]. As described in [6], Sukatme's test was applied to the central Iowa series and showed a probability of less than 0.0005, indicating a significant departure of the series from a Poisson distribution in the direction of a time clustering of tornado occurrences in the annual totals. Hence, the series is of the Polya type having a negative binomial distribution. The probabilities of this distribution are given by

$$f(t)=\frac{\Gamma(t+k)p^t}{\Gamma(t+1)\Gamma(k)(1+p)^{k+t}} \tag{13}$$

where p and k are parameters.

Application of Fisher's criterion as in [5] showed that the method of moments would not produce efficient estimates of p and k; hence, the method of maximum likelihood [6] had to be used. This produced estimates $\hat{p}=1.30$, $\hat{k}=1.06$, using $\bar{t}=1.37$. When these were substituted in equation (13) for various annual occurrences, probabilities at each t were obtained. When these in turn were multiplied by the total number of occurrences for the record, 41, the estimated annual frequency $\hat{g}$ for each t was obtained. This is shown in the comparative table, table 3. Here the observed value g_0 is compared to the estimated value $\hat{g}$. A χ^2-test showed the fit of the negative binomial distribution to be excellent, as is also clear from a comparison of g_0 and $\hat{g}$. As might have been expected, like other convectional storms [7], tornadoes tend to cluster within years and follow a Polya process rather than a Poisson process in areas where frequency of occurrence is high.

TABLE 3.—*Annual tornado frequencies—central Iowa*

t	g_0	$\hat{g}$
0	17	17.0
1	10	10.2
2	6	5.9
3	4	3.4
4	2	2.0
5	0	1.2
6	1	0.7
7	0	0.4
8	1	0.2

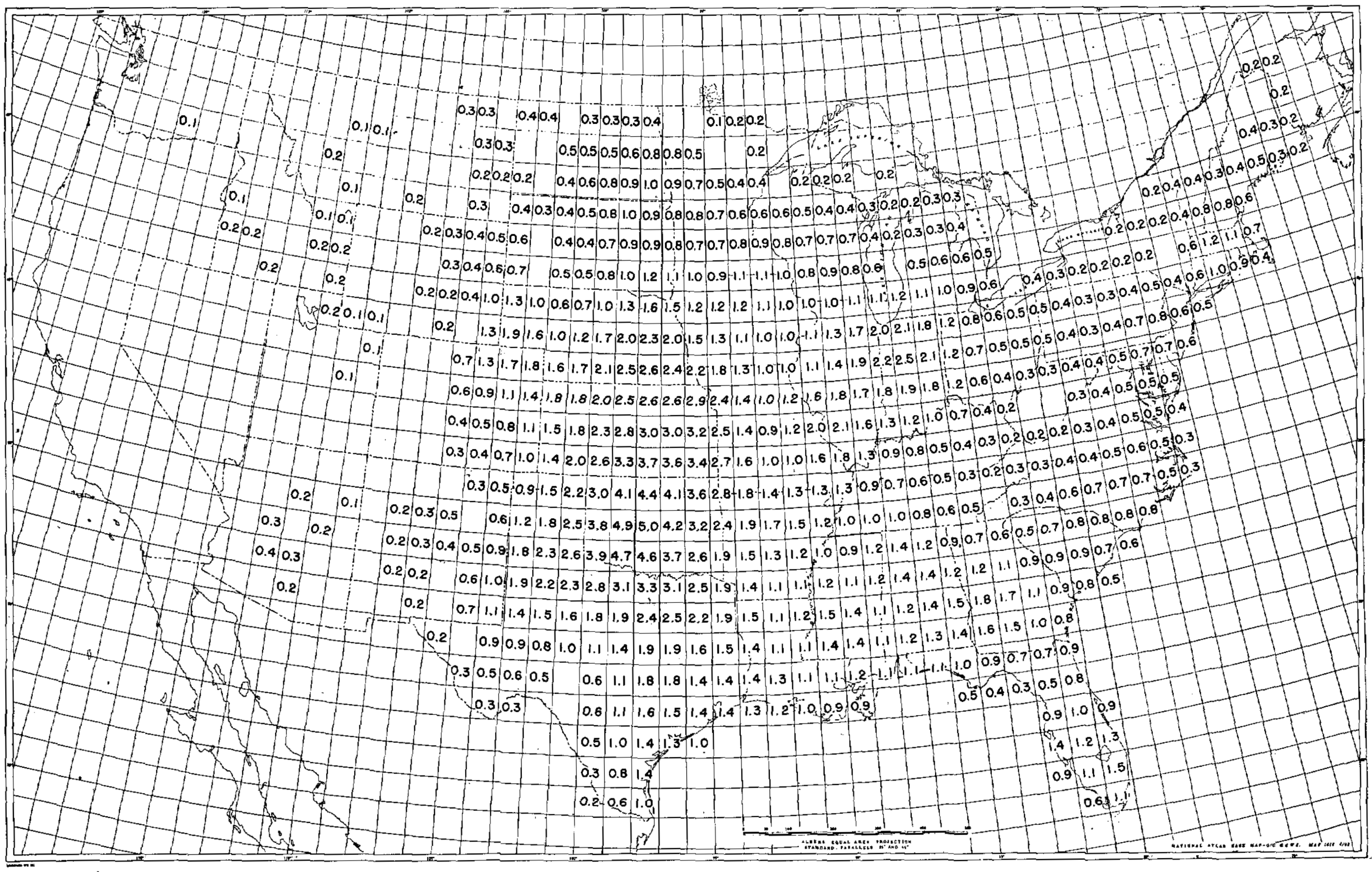

FIGURE 4.—Mean annual frequency of tornadoes, 1953–62.

5. TORNADO FREQUENCY IN THE UNITED STATES

When the interest in local severe storms in the Weather Bureau was heightened by the development of forecasting methods in the late 1940's, it was already fully realized that tornado observations were incomplete in many areas. Beginning in the early 1950's, efforts were made to make the observation of frequency more complete, and by 1953 it was thought that a large proportion of the tornadoes in the areas of high frequency was being reported.

The run test on the central Iowa Counties discussed above did not show a significant lack of runs in the 1916–56 period. This could be accounted for either by the fact that observing was already nearly exhaustive in this area, which appears to be a good possibility, or that not enough of the period after 1952 was included to show in the test.

We shall not go into this further at present, but rather make a test on the annual frequencies for the whole State of Iowa. The mean frequency for the period 1916–52 is 13.8 and the mean frequency for 1953–62 is 16.6. Since the samples are fairly large and the distributions not widely different from normal, the *t*-test may be used to test the difference in the means. There is one difficulty, however, which must be taken care of first. This is caused by the fact that an F-test showed the variances of the two periods to be different. Hence, it was necessary to correct for this, using Geary's method which resulted in a reduction of the degrees of freedom from the original $36+9=45$ to 24. Even this severe reduction in the degrees of freedom did not alter the test result; the difference between the mean frequencies of the two periods was not significant at the 0.10 level. This indicates that for the Iowa record the mean does not appear to have changed, but the variance has. Hence, in all probability there has been a change in the shape of the distribution.

It appears likely therefore that since there was a change in the Iowa record there was all the more change in other areas. Consequently, it seems desirable to prepare a United States map for the shorter more complete record for the period 1953–62 and thus, it is hoped, to avoid large biases at the expense of more moderate increases in standard errors.

Figure 3 shows a map of the total number of tornadoes occurring in 1° squares smoothed by Hann areal smoothing, i.e., smoothing in both the north–south and west–east directions with the Hann weights 0.25, 0.50, 0.25. Figure 4 shows the mean annual frequency of occurrence of tornadoes. This is needed in estimating probability of a tornado striking a point.

6. POINT PROBABILITY

For a number of applications the place where a tornado might strike may be approximated by a geometrical point, therefore the probability of a tornado striking a point is of interest. By the principles of geometrical probability the probability of a tornado striking a point is the ratio of the mean area covered by tornadoes per year to the area over which the tornadoes may occur. If we take the mean path area of a tornado to be $\bar{z}$ in square miles and the mean number of tornadoes per year to be $\bar{t}$, then the average area covered by tornadoes per year will be $\bar{t}\bar{z}$. If $\bar{z}$ and $\bar{t}$ are defined for 1° squares as $\bar{t}$ is in figure 4, then the mean probability P of a tornado striking a point in any year in a 1° square with $\bar{z}$, $\bar{t}$, and area A is

$$P=\frac{\bar{z}\bar{t}}{A} \tag{14}$$

If it is assumed further that $\bar{z}$ is invariant, we may substitute the value for $\bar{z}$ previously obtained and equation (14) becomes

$$P=\frac{2.8209\bar{t}}{A} \tag{15}$$

A may be found from geographical tables but for convenience we give in table 4 its value for each 5° for the range of latitude of the United States. Linear interpolation will suffice between the values.

For the square below Des Moines, Iowa, at latitude 40°30′, $\bar{t}=1.3$ and $A=3634$; hence $P=0.0010$. The mean recurrence interval for a tornado striking a point is $R=1/P$ or 1000 yr. for a point in this square.

TABLE 4.—*Areas of 1° squares (sq. mi.)*

	Latitude of middle of square					
	25°30′	30°30′	35°30′	40°30′	45°30′	50°30′
Area	4300	4109	3887	3634	3354	2983

ACKNOWLEDGMENTS

The author expresses his sincere appreciation to Mrs. Betty Galvin for preparing and smoothing the tornado frequency maps and for performing many of the computations, and to Mr. Maurice Kasinoff for performing and checking part of the final computations.

REFERENCES

1. M. O. Asp, "History of Tornado Observations and Data Sources," *Key to Meteorological Records Documentation* No. 3.131, U.S. Weather Bureau, Washington, 1963, 25 pp.
2. H. C. S. Thom, "Survey Report on Tornado Studies in Iowa," Manuscript, U.S. Weather Bureau, Des Moines, Iowa, 1945.
3. H. C. S. Thom, "Climatological Analysis of Tornado Path-Length and -Width," *Bulletin of the American Meteorological Society*, vol. 38, No. 3, Mar. 1957, p. 148 (abstract only).
4. L. J. Battan, "Duration of Tornadoes," *Bulletin of the American Meteorological Society*, vol. 40, No. 7, July 1959, pp. 340–342.
5. H. C. S. Thom, "The Distribution of Annual Tropical Cyclone Frequency," *Journal of Geophysical Research*, vol. 65, No. 1, Jan. 1960, pp. 213–222 (see pp. 217–218).
6. H. C. S. Thom, "The Frequency of Hail Occurrence," *Archiv für Meteorologie, Geophysik und Bioklimatologie*, vol. 8, 1957, pp. 185–94.
7. H. C. S. Thom, "A Method for the Evaluation of Hail Suppression," *Zeitschrift für Angewandte Mathematik und Physik*, vol. 9, No. 1, Jan. 1958, p. 37–63 (see p. 47).