DISTRIBUTION OF MAXIMUM ANNUAL WATER EQUIVALENT OF SNOW ON THE GROUND*

H. C. S. THOM

Environmental Data Service, ESSA, Washington, D.C.

ABSTRACT

For later development of design snow loads, the water equivalent of snow on the ground appears to be the best meteorological variable for determining design values. The appropriate climatological series for this is the winter season maximum accumulated water equivalent series. Among many distributions investigated, the lognormal distribution provided the best fit to these climatological series. Distributions fitted to 140 stations provided data for preparing contour maps of the parameter estimates. In southern areas of the United States, where snow does not occur every year, the climatological series is a mixture of zeros and water equivalent maxima. This requires the fitting of mixed distributions in these areas. A contour map is provided of the mixture parameter estimate. Methods for determining confidence intervals for quantiles from both distributions are developed.

1. INTRODUCTION

The main objectives of this study are to provide analyses and results which will furnish the means of later obtaining design snow loads. One of the requirements which modern design data must meet is that they should provide the designer with a choice of risks which are related to the use of the structure. Hence, the analysis must provide frequency distributions in which selected probabilities may be used to determine design values.

In the winter of 1952–53, the U.S. Weather Bureau began measuring the water equivalent of the snow on the ground each day. This provided the first extensive data from which the weight of the snow on the ground could be determined directly without reference to density. This weight included any rainfall added to the snow mantle which must be included in the total load. The elimination of the consideration of density was a great help, for the density of a snow pack can be highly variable and little information on its magnitude is available.

There have been a number of studies of snow depth with the objective of obtaining snow loads. As far as is known, all of these began with some form of snow depth statistics and applied a density average to obtain snow weight. Since the density itself is a random variable with a distribution having a relatively large scale, the distribution of water equivalent cannot be obtained by applying an average density to the snow depth distribution. The distribution obtained by this procedure underestimates the probabilities of the higher values being exceeded. It must be said, however, in justification for this procedure, that it furnishes design values which are certainly much better than no design values at all.

*This study was supported in part by the Agricultural Engineering Research Division, ARS, U.S. Department of Agriculture.

2. DENSITY PROPERTIES

Although it is commonly believed that snow pack density and depth are related, this is a fallacy. In reference [1], figure 7.06 gives the results of an investigation the author made of the snow pack in New York State on March 1, 1940. The graph of density against depth shows no relationship, although depths varied from about 2 in. to 44 in. The mean density for the 218 observations was about 0.22 in. of water per in. of snow. Similar independence between density and depth was found in an extensive series of Russian density-depth observations. In connection with the present study, 363 maximum annual snow accumulations were related to their corresponding densities. Again, there was no evidence of a correlation of density with depth up to 50 in. The correlation coefficient for these data is −0.0698, which is not significantly different from zero at the 0.05 level. The highest density observed was at a depth of 10 in., and the mean was about 0.15. In extremely deep packs, greater than 90 in., there seemed to be a shift of the mean density to higher values, as shown by ten values ranging in depth from 90 in. to 180 in. with mean density of about 0.44. The greatest density for these values, however, was less than the maximum observed at 10 in. depth. A possible explanation for this is that such great depths are always the result of long accumulations and extreme melting periods, whereas the lower depths can occur in a single fall or relatively short periods of accumulation. This shift of the mean density does not affect our results because our methods will treat only normal local densities. Conditions where accumulations of such great magnitude are prevalent require special consideration in design problems.

The availability of extensive water equivalent data opens the possibility of studying density more closely. The density of the snow pack is taken to be d with units inches of water per inch of snow depth. The reciprocal of this will be called the specific depth y. Since d varies between 0 and 1, y must vary between 1 and ∞. The climatological series of both d and y obtained from the maximum annual water equivalent of the snow pack and the corresponding snow depth have been found to follow lognormal distributions with unit bounds. The distributions are, of course, reflections of each other, but the specific depth distribution is a bit more conventional. The maximum snow depth accumulation z has also been found to follow a lognormal distribution. Since the product zd is water equivalent w, it must also follow a lognormal distribution, although not necessarily the one sought because the maximum depth does not necessarily go with the maximum water equivalent. There are interesting possibilities here of using the snow depth distribution based on a longer record to increase the precision of the water equivalent distribution. For the present, the water equivalent distribution has been treated directly, leaving the further development of the distributions of z and y to a later study.

3. STATISTICAL METHODS

In a previous paper [2], it was proposed to use the Fisher-Tippett Type I extreme value distribution to fit the annual maximum water equivalent climatological series. Later in a preliminary report to the American Society of Agricultural Engineers (Paper 62–903) in Chicago, December 1962, it was proposed to use the Fréchet distribution in an attempt to reflect the scale change from snow depth to water equivalent. After much further study, it was found that neither of these asymptotes provided a satisfactory fit to the water equivalent series; the Type I seemed to underestimate the upper quantiles whereas the Fréchet distribution greatly overestimated the upper quantiles. No exact explanation of the failure of those two asymptotic distributions was found, although it is suspected that it may be due to the cumulative nature of the snow pack maxima, and the fact that they are drawn from a relatively small sample. The gamma distribution was also fitted to a number of series but proved to be no improvement over the extreme value distributions.

Several stochastic processes as models for the accumulation of the snow pack were tried, but none gave a clue to the extreme value distribution. An examination of the moments of a number of sample series of water equivalent maxima showed that the lognormal distribution should give a good fit. This distribution was then tried on about 50 stations distributed over the United States and found to fit annual maximum water equivalent series very well. This distribution was therefore employed in the climatological analysis. In all of this investigation, the classical work of Gumbel [3] was employed extensively.

There is another statistical problem which occurs in snow distributions. This arises from the fact that in more southern locations of the United States, it does not snow every year, and hence there is no snow pack in some years. This results in climatological series of zero and non-zero water equivalent values which form a mixed distribution. The general problem has been treated before [4]. Let $u=\ln w$ and assume a lognormal distribution for w, then u is normally distributed with distribution function $N[u;\ \alpha_1(u),\ \sigma(u)]$. Let q be the probability of a year with *no snow*, then $p=1-q$ is the probability of a year *with snow*. Since the zero component with probability q is at the start of the distribution, the mixed distribution function may be expressed as

$$G(u)=q+pN(u). \tag{1}$$

Solving for $N(u)$ and transforming to the unit normal distribution gives

$$\frac{G(u)-q}{p}=N\left[\frac{u-\alpha_1(u)}{\sigma(u)}\right]. \tag{2}$$

Inverting and solving for u gives

$$u(G)=\sigma(u)N^{-1}\left[\frac{G(u)-q}{p}\right]+\alpha_1(u) \tag{3}$$

where N^{-1} is the inverse normal distribution function. If $\sigma(u)$ is estimated by the sample standard deviation s, $\alpha_1(u)$ is estimated by the sample mean $\overline{u}$, and p is the proportion of years with snow, the sample quantile for the probability $G(u)$ is given by

$$u(G)=sN^{-1}\left[\frac{G(u)-q}{p}\right]+\overline{u}. \tag{4}$$

This equation may be used to provide the design quantiles $u(G)$ for specific values of $1-G(u)$, the probability of exceeding a design value $u(G)$.

4. CLIMATOLOGICAL ANALYSIS

The three most important factors limiting the precision of a climatological analysis are accuracy of the basic data, the length of the climatological series, and the adequacy of the statistical model.

As is well known, snow in general is a difficult element to measure, and snow on the ground is particularly difficult. Not only is the snow pack extremely variable in depth, but it is also extremely variable in density, as shown by the survey mentioned earlier [1]. An exception, of course, is the occurrence of snow which is relatively easy to measure, and thus the snow or no snow series is probably highly accurate.

Only ten years of records were available for the water equivalent series. This tailed off to no record at all for many southern stations because the water equivalent observations began only in 1952–53 for Weather Bureau first-order stations. The snow or no-snow series was adequate in all areas since long records of snow on the ground were available, and 30 years, the international standard, was used in every instance for estimating p.

The statistical model assuming a lognormal distribution for the water equivalent annual maxima was shown to be adequate by an examination of the moments and the graphical plotting of climatological series for 50 stations. Figure 1 shows the standardized data for the first 10 stations of the alphabetical list of 140 stations employed, plotted on lognormal probability paper. In view of the small sample, 10 points for each station, the Blom plotting positions have been employed. The Kolmogorov-Smirnov test applied to each of the ten-point series showed no significant departure at the 0.20 limit for any of the samples. The reader should keep in mind that the larger the significance limit probability, the better is the fit. The probability of a good fit for all ten distributions is therefore high.

The snow or no-snow climatological series was assumed to have a binomial distribution. The only thing which could cause departure from this model would be correlation from year to year. This would, of course, be negligible.

As is customary, each of the parameters of the distribution (1) is assumed to form a climatological field and hence can be represented by contours on a map. The estimate of p is simply the proportion of the years with snow and is shown mapped in figure 2. West of the 100th meridian records were scarce, so, to augment the information on p, weekly maps of snow cover prepared for another purpose were employed. Figure 2 does not attempt to reflect exceptional local conditions such as occur in mountainous areas, but it is thought to depict conditions

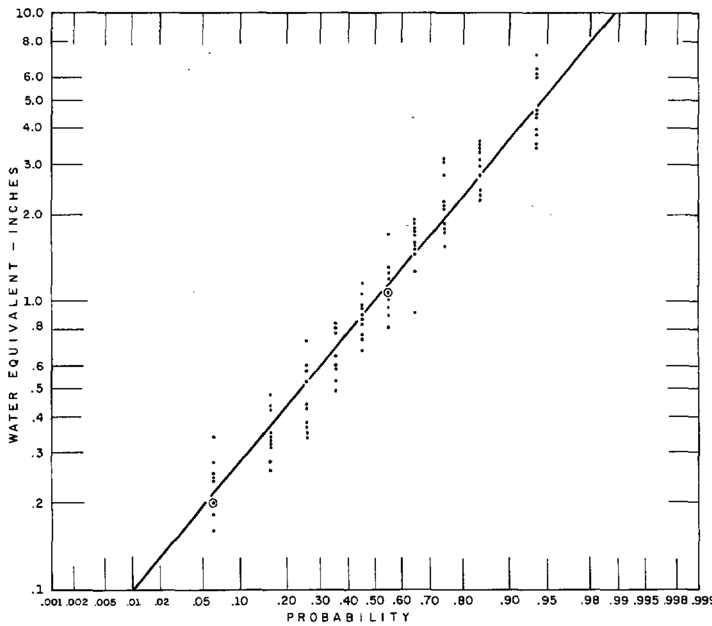

FIGURE 1.—Standardized distribution function for 10 stations.

where first-order meteorological stations might be located. The unit contour separates the region to the east and north where a snow pack occurs every year from the complementary region where it does not occur every year.

Figures 3 and 4 give the contours of $\bar{u}$ and $s(u)$, or the mean and standard deviation of the logarithms of the water equivalent series, which were estimated for the 140

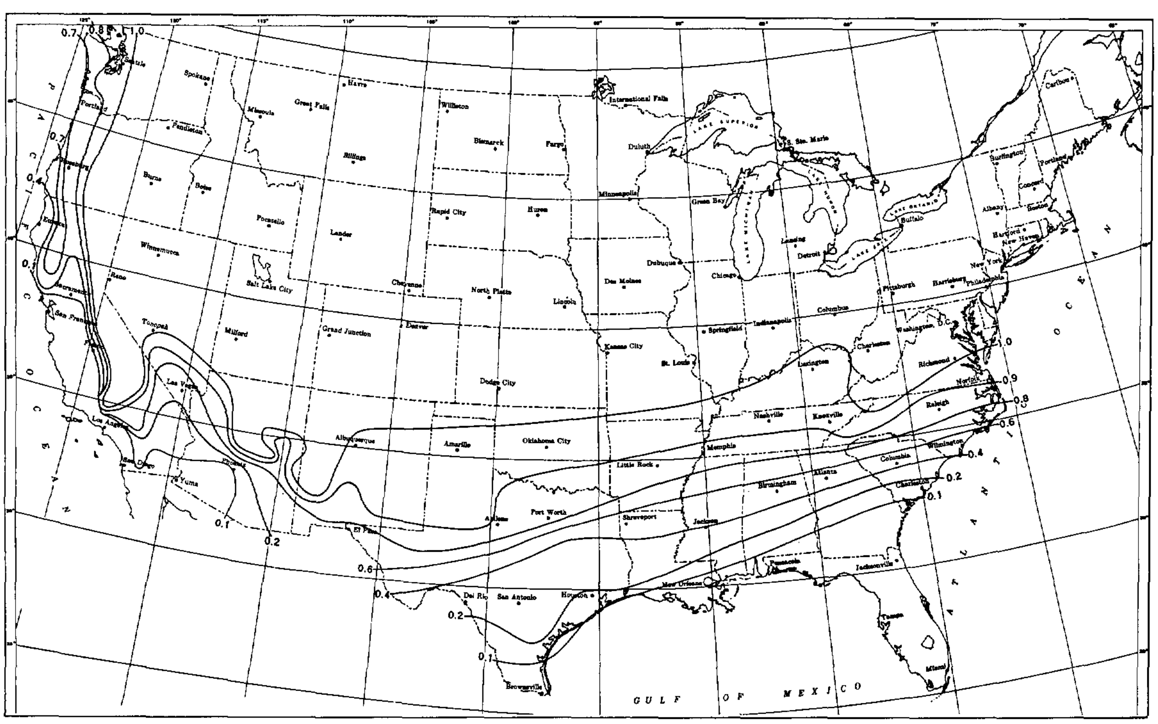

FIGURE 2.—Probability of annual occurrence of snow.

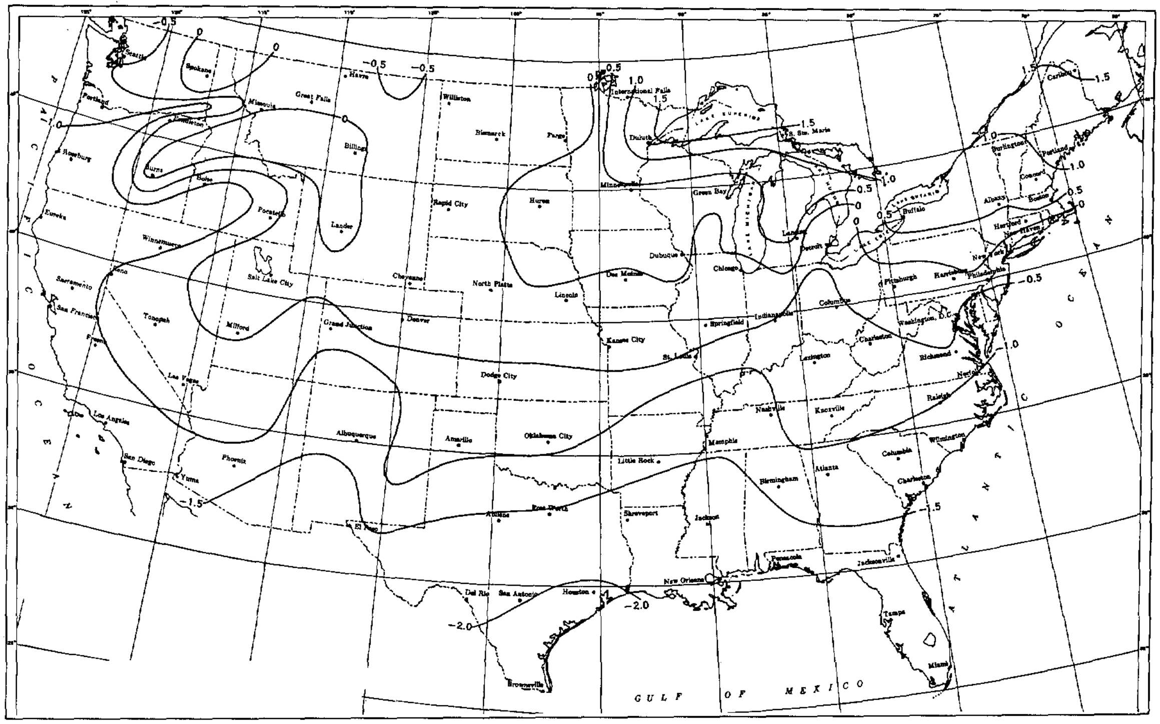

FIGURE 3.—Mean of the logarithms of the water equivalent series. Values are based on water equivalent of snow accumulation on ground for general elevations such as those near meteorological stations. Any effect for unusual conditions such as for high elevations, drifting, etc. must be taken into account by further analysis.

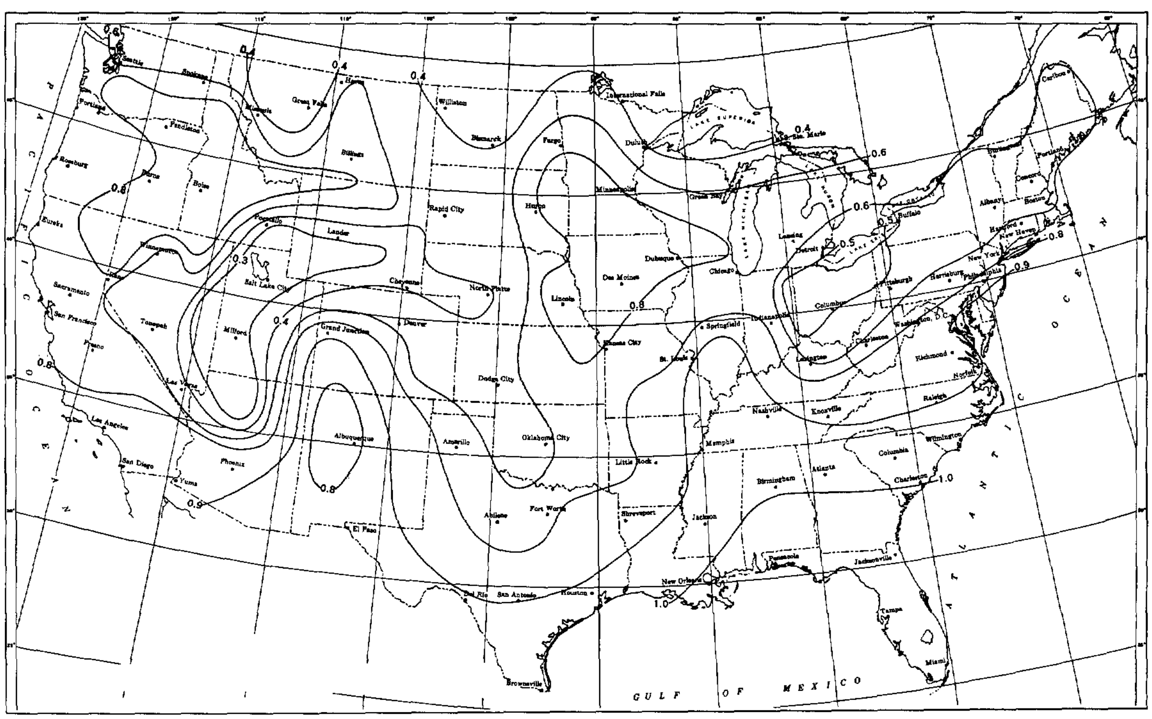

FIGURE 4.—Standard deviation of the logarithms of the water equivalent series. See legend to figure 3 for note on inapplicability of analysis for unusual conditions.

stations by the usual methods. Here again the contours are meant to depict only local conditions associated with normal exposure. Any unusual circumstances which are common for snow variables must be given special consideration.

5. CONFIDENCE INTERVALS

For the mixed distribution, it is difficult to take the parameters of the normal distribution into account in determining confidence intervals for u. A normal approximation could be used, but this gives symmetrical confidence intervals which are unrealistic at high distribution function values. As a substitute, a method was developed for adjusting the binomial probabilities to account for the mixing of unequally precise probabilities.

It is well known that the probability at any point on the sample distribution function has a binomial distribution. This has been used to obtain nonparametric confidence intervals by Clopper and Pearson [5]. Their results have been extended to confidence values of interest in climatology by Deckinger (see [6]) and are reproduced by Dixon and Massey [7].

The difficulty in using the binomial distribution directly is that p and $N(u)$ in equation (1) are estimated from different sample sizes. An adjustment to the sample size will be used to obtain a binomial approximation to the distribution of the estimates of $G(u)$: For simplication, let $N(u)=p'$, then remembering that $p+q=1$ and $p'+q'=1$, we may express equation (1) as

$$G(u)=1-pq' \quad (5)$$

which is also its mean value. Taking the variance of (5), since p and q' are independent, yields

$$v(G)=p^2v(q')+q'^2v(p) \quad (6)$$

q' and p are associated with binomial distributions with $v(q')=p'q'/n_1$ and $v(p)=pq/m$ where n_1 and m are number of years of record available for estimating $v(q')$ and $v(p)$. Let $max(n_1)=n$ which is associated with $p=1$, hence $n_1=pn$ to the nearest integer. Thus, the first variance becomes $v(q')=p'q'/pn$. Substituting for $v(q')$ and $v(p)$ in (6) gives

$$v(G)=\frac{p^2p'q'}{n}+\frac{q'^2pq}{m}. \quad (7)$$

Since $max(pq)=0.25$ and q' will usually be less than 0.10, the second term in (7) will be less than 1/12000. For q' larger than 0.10, the binomial distribution for sample size n should be used directly on $G(u)$. When the second term in (7) is ignored, it is seen that variance of G has an effective sample size n/p to the nearest integer. When $p=1$, there are no years *without* snow and $v(G)$ approaches its value for sample size n. As p becomes small, there are no years *with* snow and $v(G)$ approaches zero as it should. Since sample size is inversely proportional to variance, n/p can be assumed to be an effective sample size. This may be used together with the mean of $G(u)$ or $1-pq'$ to define an approximate binomial distribution for determining confidence intervals. Thus, when $q'\leqq 0.10$, approximate confidence intervals are available from Dixon and Massey's [7] tables A–9; the abscissa is entered with $1-pq'$ and the curves for sample size n/p are used to the nearest integer. If one is satisfied with confidence intervals which are somewhat too wide for all values $p<1$, one may enter the tables with sample size n which was proposed above for $q'>0.10$. In all cases, the confidence intervals obtained for probabilities may be converted to confidence intervals on $u(G)$ by substitution in the inverse of the distribution function $G(u)$. Those for w in all instances will be obtained from $w=\exp u$. An example will be given at the end of the discussion of confidence intervals.

As will be noted in figure 2, about two-thirds of the country has $p=1$. In this situation, of course, the distribution function $G(u)$ is normal; hence confidence intervals on u for particular values of G are available from the noncentral t-distribution. Tables are available for this distribution, but they cover only a narrow range of probabilities and are difficult to use. A very close normal approximation due to Jennett and Welch [8] has been adapted for obtaining confidence intervals for normal quantiles [9].

Since u is normally distributed with population mean μ and standard deviation σ, the population standardized value for a fixed G may be written

$$K(G)=(u(G)-\mu)/\sigma \quad (8)$$

which is estimated from a sample by

$$k(G)=(u(G)-\bar{u})/s \quad (9)$$

For a fixed G, $k(G)$, say k for brevity, will vary from sample to sample and thus will have a distribution of its own. If n is the sample size for estimating $\bar{u}$ and s, $\sqrt{n}k$ has been shown to have a noncentral t-distribution with $(n-1)$ degrees of freedom and noncentrality parameter $\delta=\sqrt{n}K(G)$. The probability $P(\sqrt{n}k<t)=F(t)=1-\alpha$ defines the $(1-\alpha)$th quantile of t. This and the αth quantile are expressed as $t[(n-1),\ \delta,\ 1-\alpha]$ and $t[(n-1),\ \delta,\ \alpha]=-t[(n-1),\ -\delta,\ 1-\alpha]$ on the noncentral t-distribution. Thus the confidence limits for $K(G)$ will be given by

$$P\{-t[(n-1),\ -\delta,\ 1-\alpha]/\sqrt{n}<K(G)<t[(n-1),\ \delta,\ 1-\alpha]/\sqrt{n}\}=1-2\alpha \quad (10)$$

Let $A=\delta^2/2(n-1)$ and $B=1-\zeta^2(1-\alpha)/2(n-1)$ where $\zeta(1-\alpha)$ is $(1-\alpha)$th quantile of the unit normal deviate. The normal approximations of the noncentral t-quantiles are then given by

$$t(1-\alpha)=t[(n-1),\ \delta,\ 1-\alpha]=\delta+\zeta(1-\alpha)\sqrt{A+B}/B \quad (11)$$

and

$$t(\alpha)=t[(n-1), \delta, \alpha]=\delta-\zeta(1-\alpha)\sqrt{A+B}/B \qquad (12)$$

Substitution of these values in (10) gives the confidence limits for $K(G)$. To obtain the confidence limits for $u(G)$ it is only necessary to note that $P[k<(u(G)-u)/s]$ is equivalent to $P[\bar{u}+sk<u(G)]$. Hence the confidence limits are

$$P\{[\bar{u}+st(\alpha)]/\sqrt{n}<u(G)<[\bar{u}+st(1-\alpha)]/\sqrt{n}\}=1-2\alpha \quad (13)$$

It must be emphasized that equations (10) and (13) give only confidence intervals. Several of these for a suitable selection of G's may be used to form a confidence band, but this band must only be used at individual values of G and never jointly at a sample of G values. The joint consideration of several G values to obtain a test for goodness of fit is quite a different problem. Thus the confidence band applies only to individual values of G and *cannot* be used as a test of goodness of fit as has been tried by some.

To obtain confidence intervals for any location above the $p=1$ line of figure 2, the mean $\bar{u}$ and standard deviation s read from figures 3 and 4 are employed. Clearly the value of a contour at a particular point is more precise than a single plotted value, for more than one point has been used in determining it. If it is assumed that the value for any point on the maps is determined from the average of four points whose errors are independent, the sample size n is multiplied by four.

As an example, the 0.98 quantile and its 0.80 confidence interval for Des Moines, Iowa, are determined. From figure 2, Des Moines is seen to be in the $p=1$ region. The mean $\bar{u}$ is close to zero and s is 0.8 as seen in figures 3 and 4. From inverse normal tables, $k(0.98)=2.054$. Substituting these values in equation (9) gives

$$u(0.98)=0+2.054\times0.8=1.64.$$

Transforming to water equivalent yields

$$w=\exp(1.64)=5.16 \text{ in.}$$

Therefore, the chance of an annual extreme water equivalent of the snow pack exceeding 5.2 in. at Des Moines, Iowa, is 0.02.

The 0.80 confidence interval may be obtained as follows: From above $n=40$, $\bar{u}=0$, $s=0.8$, and $G=0.98$. Hence $\zeta(1-\alpha)=\zeta(0.90)=1.2816$, $K(0.98)=2.054$, and $\delta=\sqrt{40}$ $(2.054)=12.992$. From these

$$A=12.992/78=0.1662$$

$$B=1-1.642/78=0.9789$$

and

$$\sqrt{A+B}/B=\sqrt{1.146}/0.9789=1.094$$

Therefore, by equations (11) and (12)

$$t(0.90)=12.992+1.2816\times1.094=14.394$$

and

$$t(0.10)=12.992-1.402=11.590$$

The first and last terms in the left of equation (13) are

$$[\bar{u}+st(0.10)]/\sqrt{n}=(0+0.8\times11.590)/6.325=1.466$$

and

$$[\bar{u}+st(0.90)]/\sqrt{n}=(0+0.8\times14.394)/6.325=1.821$$

so the confidence interval for $u(0.98)$ is

$$P(1.466<u(0.98)<1.821)=0.80$$

Taking exponentials to convert to water equivalent gives

$$P(4.332<w(0.98)<6.178)=0.80$$

Thus the true value of $w(0.98)$ will be covered by the interval (4.332 in., 6.178 in.) with probability 0.80.

To illustrate the estimation of confidence intervals for the mixed distribution, i.e., when there is no snow pack in some years, Jackson, Miss., may be employed as an example. Again the interval for the 0.98 quantile is required. From figures 3, 4, and 5, find $p=0.4$, $\bar{u}=-1.7$, and $s(u)=0.95$. Substituting these in equation (4) gives

$$u(0.98)=0.95N^{-1}\left[\frac{0.98-0.6}{0.4}\right]-1.7=-0.14$$

Converting to water equivalent yields

$$w(0.98)=\exp(-0.14)=0.869 \text{ in.}$$

Therefore, the probability of exceeding 0.87 in. of water in the snow pack at Jackson, Miss., is 1−0.98 or 0.02.

To obtain the confidence interval for $w(0.98)$, since $p<1$ and $q'<0.10$, the effective sample size concept developed above is used to approximate a binomial distribution. If we assume again that the contour readings have half the sampling errors of the original station values, n is again 40 and the effective sample is 40/0.4 or 100. The 0.80 confidence limits for the probability read from Dixon and Massey's [7] table A–9a at $X/N=0.98$ are 0.952 and 0.987. Since these are G values, the corresponding quantiles $u(G)$ may be obtained from equation (4). Thus

$$u(0.952)=0.95N^{-1}\left[\frac{0.952-0.6}{0.4}\right]-1.7=-0.58$$

and

$$u(0.987)=0.95N^{-1}\left[\frac{0.987-0.6}{0.4}\right]-1.7=0.06$$

Converting to water equivalent gives

$$w(0.952)=\exp(-0.58)=0.56 \text{ in.}$$

and

$$w(0.987)=\exp(0.06)=1.06 \text{ in.}$$

Hence, the confidence interval for the 0.98 quantile is

$$P[0.56 < G(u) < 1.06] = 0.80$$

ACKNOWLEDGMENTS

The author expresses his sincere appreciation to Mrs. Betty Galvin for fitting most of the frequency distributions using several different statistical models, performing many other test computations, and for assisting in preparation of the basic data; to Mr. Maurice Kasinoff for assistance in preparation of the data; and to Mr. C. K. Vestal for investigating an unusually heavy snowfall at Fort Worth, Tex. Dr. R. Sneyers was very helpful in reviewing the manuscript.

REFERENCES

1. U.S. Weather Bureau, "Maximum Possible Precipitation over the Ohio River Basin Above Pittsburgh, Penna.", *Hydrometeorological Report* No. 2, 1941, (see pp. 78–80).
2. R. H. Hahn, H. C. S. Thom, and T. E. Bond, "Livestock Shelter Design," *Agricultural Engineering*, vol. 43, 1962, pp. 704–09.
3. E. J. Gumbel, *Statistics of Extremes*, Columbia University Press, New York, 1958.
4. H. C. S. Thom, "The Distribution of Freeze-Date and Freeze-Free Period for Climatological Series with Freezeless Years," *Monthly Weather Review*, vol. 87, No. 4, Apr. 1959, pp. 136–144.
5. C. J. Clopper and E. S. Pearson, "The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial," *Biometrika*, vol. 26, 1934, pp. 404–413.
6. Churchill Eisenhart, et al., *Techniques of Statistical Analysis*, McGraw-Hill Book Co., Inc., New York, 1947, pp. 331–338.
7. W. J. Dixon and F. J. Massey, *Introduction to Statistical Analysis*, 2d Edition, McGraw-Hill Book Co., Inc., New York, 1957.
8. W. J. Jennett and B. L. Welch, "The Control of the Proportion Defective", *Supplement, Journal of the Royal Statistics Society*, vol. 2, 1939, pp. 80–88.
9. H. C. S. Thom, Discussion of "Flood Probabilities," *ASCE Proceedings*, Sep. 568, Dec. 1954, pp. 11–13.

[*Received January 17, 1966*]