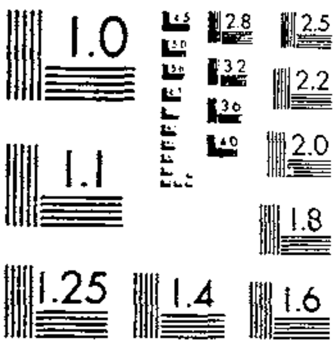
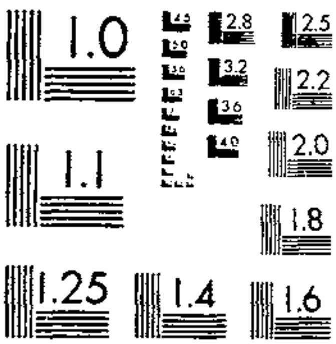
TB 814 (1942) USDA TECHNICAL BULLETINS UPDATA
DROUGHT IN THE UNITED STATES ANALYZED BY MEANS OF THE THEORY OF
BLUMENSTOCK, G., JR 1 OF 1

START



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Technical Bulletin No. 819 • April 1942

UNITED STATES DEPARTMENT OF AGRICULTURE
WASHINGTON, D. C.

Drought in the United States Analyzed by Means of the Theory of Probability[1 2]

By George Blumenstock, Jr., *Junior Soil Conservationist, Climatic and Physiographic Division, Office of Research, Soil Conservation Service*[3]

CONTENTS

INTRODUCTION

The ability to forecast weather conditions not only 1 or 2 days but also weeks, months, or even years in advance is of increasing importance in our modern economic life. The farmer is interested in knowing whether climatic conditions will enable him to produce a good crop during the coming season. The hydrologist, concerned with the construction and operation of dams or reservoirs, is interested in detailed knowledge of the rainfall conditions expected in the years to come in order that he may intelligently gage the necessary strength or capacity of his structures. Economists, soil conservationists, and many other scientists are directly concerned with problems necessitating the examination and analysis of climatic data.

Through the use of synoptic weather maps it is now possible to make weather forecasts one to several days in advance. But the

[1] Submitted for publication July 14, 1941.

[2] Assistance in the preparation of these data was furnished by the personnel of Works Progress Administration Official Projects No. 765-42-3-10 and No. 65-2-42-282.

[3] This study was prepared under the direction of C. W. Thornthwaite, Chief, Climatic and Physiographic Division, Office of Research, Soil Conservation Service. Gratitude is also expressed to W. G. Cochran, Statistical Laboratory, Iowa State College, for reviewing the manuscript and offering suggestions and criticisms and to Paul Kaser, project supervisor, Works Progress Administration, for supervising the tabulation and computation of the statistical data.

accuracy of these predictions diminishes as the range of the forecast lengthens. At present successful application of synoptic techniques in forecasting does not extend much beyond 5 days.

This does not mean, however, that nothing is known about the weather most likely to occur 10, 20, or even 200 days in the future. Though weather conditions vary from day to day and from week to week, certain conditions are more likely than others, the likelihood of the occurrence of an event depending upon the mean pattern of atmospheric circulation and the variations which occur about that mean. These variations are reflected in climatic data collected at several thousand stations in the United States for periods of a few years to many decades. From statistical analysis of such data, the climatologist may determine the most likely weather conditions for any period of time within the year. Frequency and probable occurrence of such important climatic factors as excessive rainfall intensities, drought, or high and low temperature extremes can be obtained with a determined degree of accuracy. In this phase of statistical meteorology probability analysis is particularly valuable.

Application of probability theory to climatic data had its beginnings, however, only in recent years. This is in part due to the fact that results of theoretical research into such phenomena as areal and temporal persistence or the random occurrence of sequences of similar or dissimilar events in time series have been available only recently in such work as that of Bartels (*2*)[4] or Mood (*11*). An example of the application of probability theory in the field of climatology is the investigation of the apparent persistence of one type of weather by Gold (*7*) and Cochran (*4*).

Most studies dealing with the frequency of occurrence of climatic events have consisted primarily of examining available meteorological data for a station or group of stations and expressing the results of the examination in a concise and readily usable form. For example, charts and maps of the United States indicating past frequency of various extremes in rainfall intensity (*10*, *16*), drought occurrence,[5] and climatic type (*17*) have been prepared. These charts and maps offer a clear picture of past occurrence and therefore supply valuable data for further analysis such as correlation with flood damage or crop failure during the same years of record. These data, however, are not so valuable when used in predicting future climatic conditions, since no technique has been developed for evaluating the sampling error[6] in the results. Some doubt must therefore exist concerning the accuracy of considering these observed frequencies as representative of true probabilities.

In estimating the likelihood of extreme climatic events rather than the likelihood of those that most frequently occur, the observed frequencies become highly unreliable. The primary reason for this is the large standard error in the observed frequencies, which results from the fact that the frequency measurements are based only upon a

---

[4] Italic numbers in parentheses refer to Literature Cited, p. 62.

[5] Blumenstock, David I. DROUGHT EXPECTANCIES IN THE UNITED STATES AS RELATED TO SOIL AND MOISTURE CONSERVATION. [Unpublished.]

[6] The term "sampling error" has been used here to indicate uncontrollable or chance variation and does not imply mistakes in the observation, collection, and tabulation of the data. Measurement of a climatic factor at a given station for 1 year only would yield a very unreliable estimate of the mean conditions at the station. Similarly, observation over a longer period of time will yield a measurement that is still only an estimate of the true mean of the parent population at the station. The difference between an observed sample and the parent population is sampling error. In general, the larger an unbiased sample is, the more accurately it will represent the parent population.

few observations and disregard the more frequent, less extreme occurrences. To take a simple example, it is as if the frequency of a run of 10 or more heads in tossing at random an unbiased coin were estimated by the number of times such runs were observed in a limited number of consecutive tosses. The frequency obtained by such an approximation might differ considerably from the true a priori frequency. As a measure of likelihood, the a priori frequency calculated on the basis of probability would provide a much more reliable measure than a simple a posteriori frequency count.

Where a priori probabilities are unknown, it may be possible to formulate a "law of occurrence" of a frequency distribution on the basis of observed data. However, in so doing, the entire distribution must be considered, and not any restricted part thereof. If a satisfactory law of occurrence can be deduced, this law can then be used in calculating frequencies of extreme events. Such calculated frequencies, with their defined confidence limits, will offer more stable and more valuable measurements of likelihood.

This fact has, of course, long been realized by statisticians dealing with frequencies. It has been recognized that so long as one has no knowledge of the law of occurrence of a climatic factor, no definite statement can be made concerning the probability of the occurrence of future events. By "law of occurrence" is meant no exact and restricting rule, but rather a statement of the mean conditions and the manner in which variation from the mean occurs. Thus, in tossing coins, the mean condition is that one-half of the tosses shall be heads and one-half tails. Variation from this mean agrees with the theory of random sampling, assuming no bias in the experiment. Knowing the mean condition and the variance, one can set up probabilities for any event. In dealing with climatic data, if analogous "laws" could be demonstrated, probabilities could be calculated.

Little has been done to determine the character of climatic laws of occurrence. Nevertheless, that climatic events occur in accordance with some law has been implicitly accepted by any number of investigators. Thus, for example, in many fields of climatic research, an effort has been made to develop a "type curve" applicable to a given climatic frequency distribution. Since the constants of the curve are derived from the observed data at the station to which it is applied, the type curve becomes merely a smoothed representation of the frequency distribution at that station. The search for type curves indicates that the investigators believe, though they seldom explicitly make the statement, that the climatic factor they are studying is governed by some complex law of occurrence, which tends to insure a basic similarity between frequency distributions at widely separated stations, although the constants describing the distributions may vary from area to area.

Much work has been done in developing type curves. In the field of hydrology, Charles W. Sherman (*13*), Merrill M. Bernard (*3*), and others have succeeded in obtaining a type curve for rainfall intensity-frequency distributions that, although still open to modification, seems to be basically adequate. Preliminary to the investigation summarized in the following pages, the author developed a type curve for drought frequency. Howard Ross Tolley has shown that type-frequency curves can be set up for temperature extremes and for certain other climatic phenomena (*15*). These type curves throw no real light on the laws of the frequency distributions they represent beyond indicating that such laws may be operative. The use of a type curve does not insure greater accuracy in estimating the likelihood of climatic events; such greater accuracy will come only as a result of further knowledge of the character of the distributions themselves.

An important question, therefore, is whether progress can be made toward this end. Granting our theory that for any given climatic phenomenon, such as drought or rainfall, there is a law of occurrence, it should be possible to make at least some progress toward determining its character through conventional statistical procedures. The usual method is to set up certain minimum assumptions concerning the probability of occurrence of the events to be analyzed. Theoretical distributions are then derived and these distributions are compared with the data observed over a period of years. By such comparisons, conclusions can be drawn concerning the validity of the original assumptions. Then, if necessary, adjustments can be made in the assumptions, and the technique repeated.

Analyses of this type are not new in statistical literature. To cite a well-known example, Major Greenwood and G. Udny Yule were able to set up theoretical distributions that were based on a theory of variable probability and adequately represented frequency distributions of the occurrence of repetitious or multiple events (*8*). This was done by imposing certain specified limitations on randomness of occurrence. The assumptions made were logical and were carefully chosen to fit the problem at hand. The mathematical derivation of the theoretical distribution followed directly from these assumptions. The final step was to compare the theoretical and observed distributions and to demonstrate their similarity.

It is important to note the differences between analysis by probability theory and by simple frequency determination. In estimating the likelihood of future occurrence on the basis of observed frequency we can merely say that a particular value seems to be most likely, but this value may be greatly in error since it is based on only a few extreme cases. On the other hand, if we can discover a law of occurrence that can be accepted as applicable to the distributions, probabilities and confidence limits can be set up for any problem whatsoever.

Probability analysis can readily be applied to climatic problems. In a study of climatic data it is altogether likely that our original assumptions, even though carefully chosen, will be complex and may prove inadequate to a complete explanation of the frequency distributions observed. However, adjustments in the assumptions can be made and the data retested.

This bulletin represents an attempt to apply probability theory to a particular climatic frequency distribution. For this purpose, frequency of occurrence of drought has been chosen, primarily because at the time this study was undertaken an exhaustive tabulation and analysis of past drought occurrence was already in progress.[7] The study has been divided into six main sections, the first four of which are concerned primarily with the derivation of theoretical distributions based upon assumptions concerning the nature of drought occurrences. The developed distributions are then compared with observed data,

[7] See footnote 5.

and the applicability of the assumptions thereby tested. Since the constants of the theoretical formulas are obtained from the same data with which the theoretical distributions are later compared, this does not constitute an a priori approach. Allowance has been made for this fact in interpreting the results.

PRELIMINARY ANALYSIS OF THE PROBLEM

Method of Collection and Organization of Observed Drought Distributions

This probability analysis of drought frequencies was made in connection with and as a corollary to a detailed study of drought occurrence in the continental United States. Drought data for over 2,000 stations were compiled and tabulated by the Works Progress Administration for the Climatic and Physiographic Division, Office of Research, Soil Conservation Service. This compilation made available for analysis a large number of drought frequency distributions. Not all these data have been utilized here. However, in conducting this study, the author had at his disposal a vast reservoir of data that could be tapped at will. The initial investigations on drought probability were made with a small number of stations chosen for contrasting climatologic characteristics. A large number of stations was not considered until the technique of the analysis finally used had been developed and had demonstrated its adequacy for the more limited sample. The records of 216 stations were wholly or partly analyzed. These stations are widely scattered over the United States and include all climatic areas.

In tabulating the data, a drought was considered terminated by the occurrence of a minimum of 0.10 inch of precipitation in 48 hours or less. Weather Bureau records for the years 1898–1937 were scanned, and for every station the length of each drought in days was recorded. Frequency tables were then prepared that gave the total number of occurrences of drought of each specified length from a single day to the maximum number of days observed.[8] Such distributions were prepared for monthly, seasonal, and annual intervals. When intervals of a month or a season were under consideration, droughts were considered terminated at the margins of the given time interval, irrespective of whether the following or preceding day continued the sequence of consecutive days of drought beyond the limits of the month or season.

The frequency distributions thus calculated were the basic form of presentation of the drought data. Two additional sets of data were computed, however. The cumulative distributions were obtained directly from the original frequency distributions and indicated the number of observed occurrences of a drought of a given number of days or more. The mean frequencies were computed as the ratio of the number of years of record to the number of occurrences in the cumulative distributions, and hence represented the average time interval between occurrences of a drought of a given minimum length.[9] Table 1 gives an example of the method of presenting these distributions.

---

[8] The term "drought" has been used to include time intervals as short as a single day.

[9] One of the purposes of the probability analysis of the drought data was to obtain some measure of the accuracy with which the mean frequencies could be used in forecasting drought hazard. The results of the following analysis have been used for this purpose by David I. Blumenstock. (See footnote 5.)

In analyzing the data through the use of probability theory an attempt was first made to deal directly with the mean frequencies. This approach was soon abandoned, however, and the basic frequency distributions were considered in all further work.

TABLE 1.—*Observed frequency of drought occurrence, North Head, Wash., for the month of August, for the 34 years of record between 1898 and 1937*

Length of drought (days)	Frequency: Noncumulative	Frequency: Cumulative	Mean frequency	Length of drought (days)	Frequency: Noncumulative	Frequency: Cumulative	Mean frequency
	Number	*Number*	*Years*		*Number*	*Number*	*Years*
1	11	91	0.37	17	2	16	2.13
2	8	80	.43	18	1	14	2.43
3	6	72	.47	19	0	13	2.62
4	5	66	.52	20	1	13	2.62
5	5	61	.56	21	0	12	2.83
6	4	56	.61	22	1	12	2.83
7	3	52	.65	23	0	11	3.09
8	2	49	.69	24	3	11	3.09
9	7	47	.72	25	2	8	4.25
10	5	40	.85	26	1	6	5.67
11	6	35	.97	27	0	5	6.80
12	4	29	1.17	28	1	5	6.80
13	3	25	1.36	29	0	4	8.50
14	4	22	1.55	30	0	4	8.50
15	2	18	1.89	31	4	4	8.50
16	0	16	2.13				

EXAMINATION OF THE DROUGHT DATA

In preparing the tabulated drought data in final form, graphs were made of the mean frequencies. It was observed that when the mean frequency was plotted logarithmically as the abscissa and the length of drought was plotted as the ordinate, a curve approximating a straight line was obtained. It was observed, further, that the closeness with which the plotted points approached a straight line on semilogarithmic paper (that is, an exponential curve on arithmetic paper) was apparently roughly proportional to the constancy of drought hazard at the station under consideration. However, because the sample from individual stations was small, irregularities in the data made it difficult to draw any definite conclusions concerning the relationship between these two factors.

To establish more stable curves, groups of 10 closely spaced stations were taken in several different climatic areas. Mean frequencies for annual time intervals were computed for the average of the 10 stations in each area, and arithmetic graphs were prepared. It was found that the data thus obtained could be closely fitted by a curve of the general type

$$y = ax + b \log x + c.$$

Figure 1 shows the observed mean frequencies and the fitted curves for four widely separated areas.[10] The curves shown were

[10] Following is a list of the stations used in obtaining the 4 sets of data: (1) Eastern Ohio and Kentucky—Ironton, Portsmouth, and Waverly, Ohio; Scott, Mount Sterling, Farmers, Lexington, Frankfort, Maysville, and Williamstown, Ky.; (2) southern Louisiana—Houma, Abbeville, Covington, Grand Coteau, Lake Charles, Jennings, Schriever, Donaldsonville, New Orleans, and Lafayette; (3) eastern Kansas—Concordia, Newton, Burlington, Emporia, Wichita, Toronto, Fort Scott, Norwich, Burr Oak, and Atchison; (4) San Joaquin and Sacramento Valleys, Calif.—Stockton, Newman, Merced, Fresno, Hanford, Visalia, Porterville, Bakersfield, Sacramento, and Marysville.

fitted to the plotted points by least squares. Since the magnitude of the constant a is proportional to the curvature of the equation plotted on semilogarithmic paper, we would expect, and, indeed, can observe

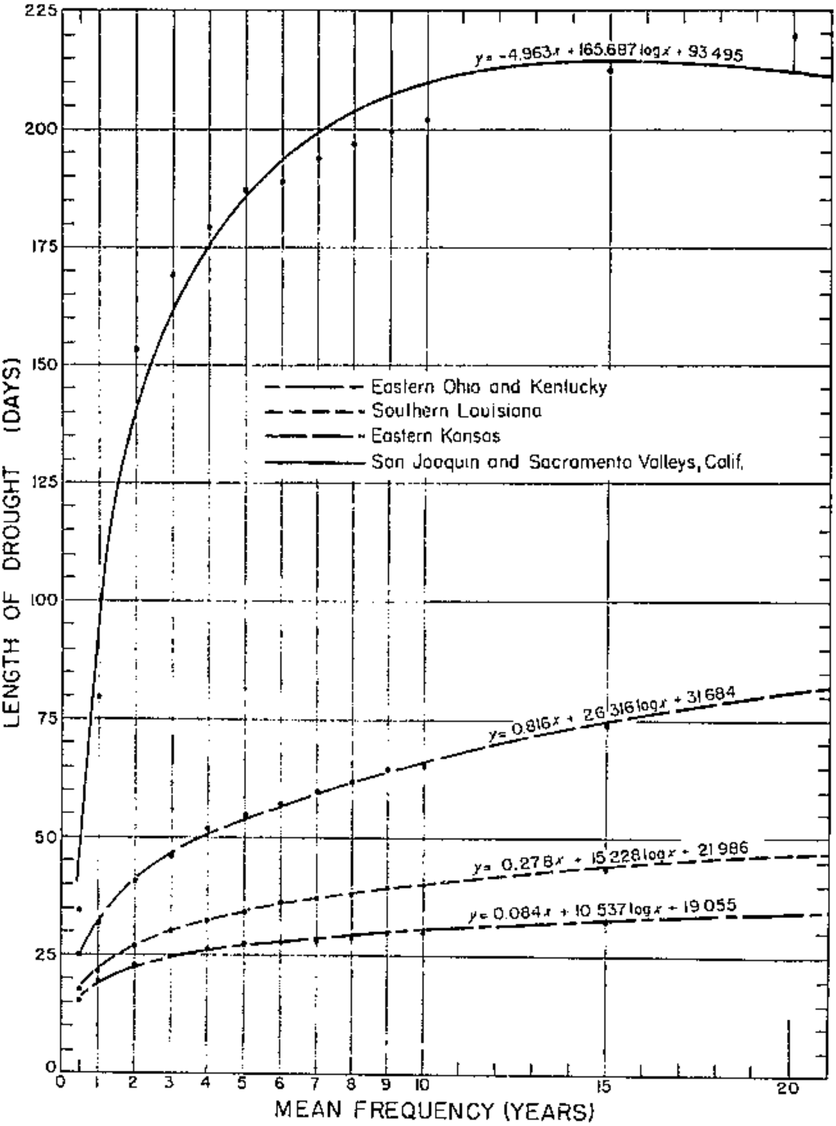


FIGURE 1.—Curves fitted by least squares to annual mean-frequency drought data for four areas in the United States.

that a varies inversely as the seasonal constancy of drought hazard for the areas.

The two points to be emphasized with respect to the mean-frequency curves are (1) the approximation of the data to an exponential curve and (2) the increasing departure from exact exponential character with the introduction of variability in drought hazard within the time interval considered. Both these conditions would obtain if occurrence of drought were random with respect to a variable daily probability of drought. Fortunately, a rough test of the correctness of this assumption is immediately available. It is possible, with little difficulty, to derive a theoretical formula for a mean frequency curve based upon the assumption of a constant daily drought probability. In an area where drought hazard does not vary extensively with the season, such a curve should be a close approximation to the observed data.

It can be shown that where p is the probability of a day of drought, and where p is held constant over the observed period of M years, the number of expected occurrences of a drought of n or more days (O_N) may be written:

$$O_N = Sp^{n-1},$$

where S is the number of droughts in the M years. Let E be the mean frequency. Then,

$$E = \frac{M}{O_N}$$

or, substituting and developing,

$$n = \left(\frac{\log M + \log p - \log S}{\log p}\right) - \left(\frac{1}{\log p}\right) \log E.$$

Note that this formula is of the form

$$y = b \log x + c,$$

whereas the fitted curve was of the form

$$y = ax + b \log x + c.$$

Deviation from simple exponential character would be caused by error in assuming constant drought probability. Hence we would expect variation in drought hazard from season to season to require the actual fitted curve to be more complex. The introduction of the term ax into the equation of the fitted curves and the observed increase in a with variability of drought hazard would seem to verify this expectation.

In the formula for the curve based upon random occurrence the constants M and S are obtained directly from the observed data. Before calculating the expected distributions, however, it will be necessary to obtain an estimation of p. Julius Hann (*9, v. 1, p. 68*) has suggested that such a probability should be taken as the ratio D/T, where D is the total number of days of drought and T is the total number of days examined.[11] If this approximation is used, the distribution expected on the basis of a random occurrence of drought can be determined.

[11] Hann deals directly with rainfall probabilities and suggests the ratio $(T-D)/T$ as the probability of a day of rain. This is the equivalent, of course, of a probability of D/T for a day of drought.

Figure 2 shows the results obtained for the average of 10 stations in eastern Connecticut and Rhode Island.[12] For these 10 stations there were 10,085 days of drought observed in 13,651 days of record;

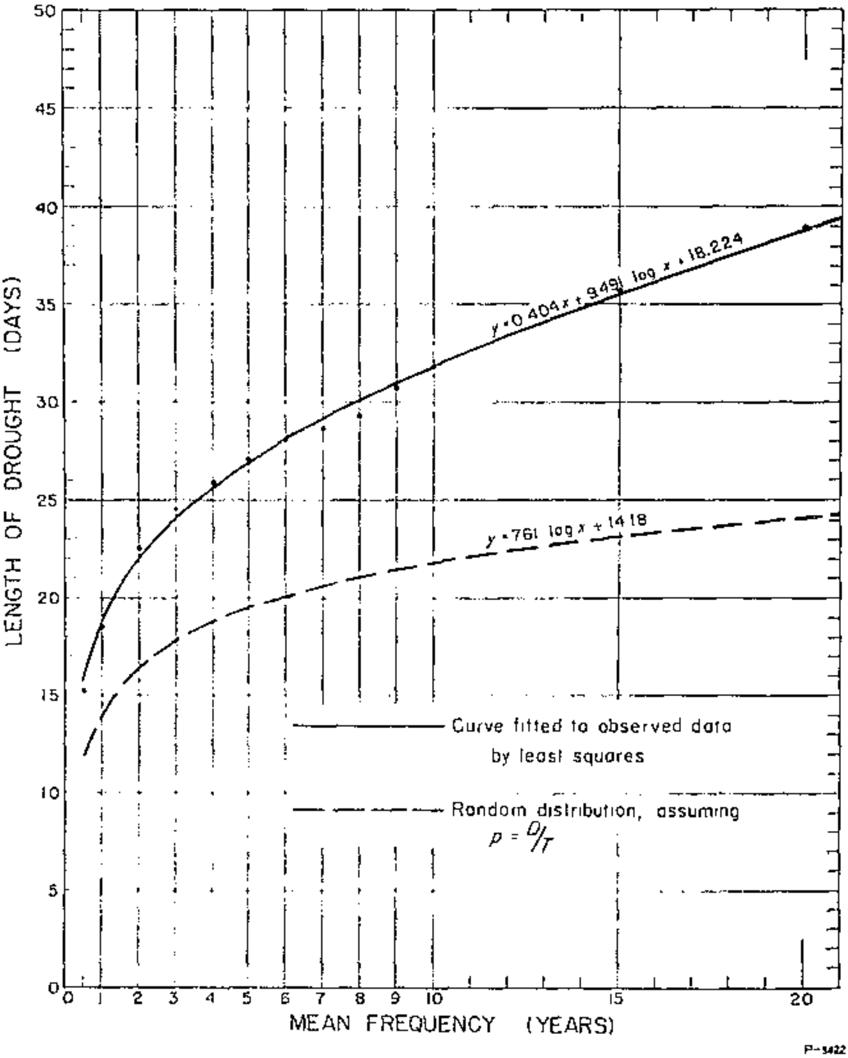


FIGURE 2. Observed data and the expected and fitted curves for the annual mean frequency of drought for 10 stations in eastern Connecticut and Rhode Island.

hence p was taken as 10,085/13,651, or 0.7388. It is obvious from inspection of figure 2 that there is a great disparity between the curve representing the expected occurrence and the observed data. This comparison is typical of the results obtained; in no case did the theo-

[12] The station at Springfield, Mass., is included in this group since it is representative of the eastern Connecticut-Rhode Island area. The other nine stations are: Colchester, Hartford, New Haven, New London, N. Grosvenor Dale, and Storrs, Conn.; Kingston, Providence, and Block Island, R. I.

427625—42——2

retical distribution calculated in this manner adequately represent the observed mean-frequency data.

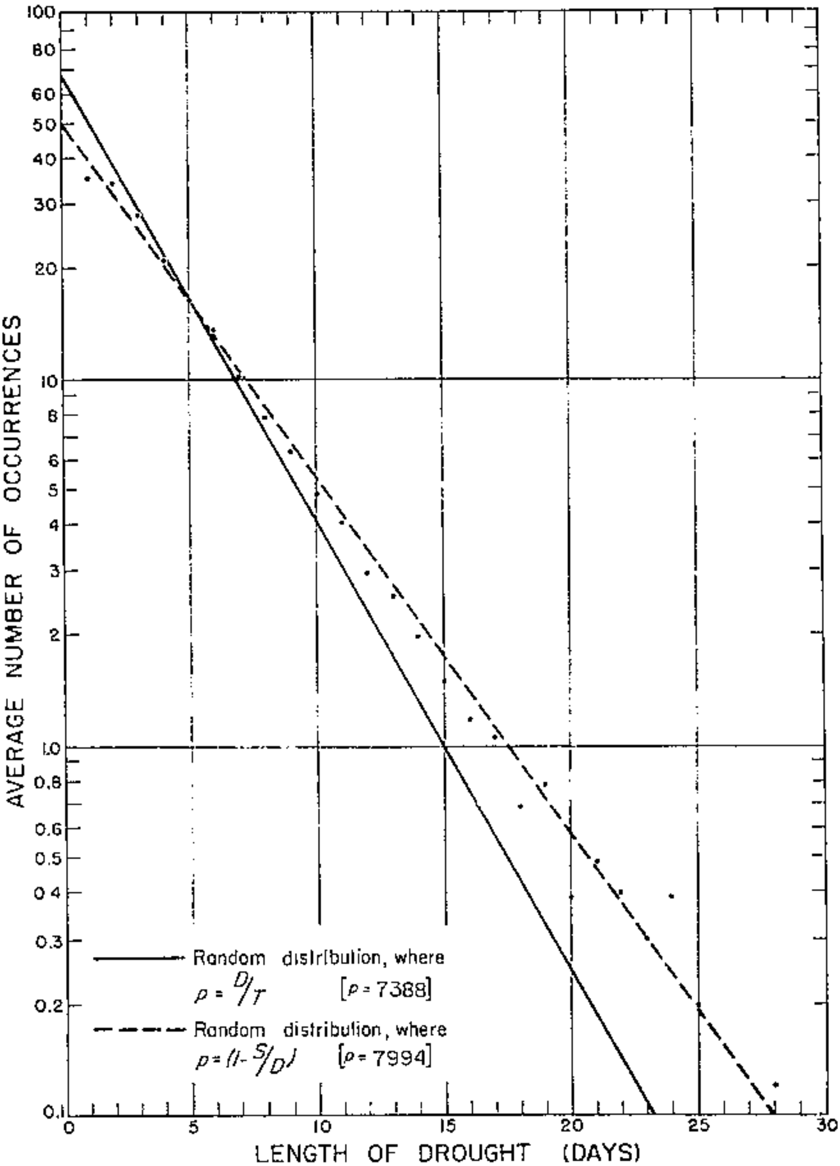


FIGURE 3. Observed data and expected annual drought frequency distributions for stations in eastern Connecticut and Rhode Island obtained by two different approximations of p.

In continuing the analysis, use of the mean frequency will be abandoned and reference made directly to the basic frequency distributions, which compare number of occurrences with exact drought length.

These distributions are more representative in that they cover the entire range of drought occurrence and do not give undue emphasis to droughts of excessive length. Figure 3 shows the drought-frequency distribution for the same 10 stations represented in figure 2. The scale of ordinates has been plotted logarithmically. The heavy line represents the exponential curve expected from random occurrence, for which the probability, p, has again been taken as constant at D/T, or 0.7388. The formula for the number of expected occurrences of droughts exactly n days in length (O_n) has been taken as:

$$O_n = S(1-p)p^{n-1},$$

where, as before, S is the number of observed droughts in the M years of record and p is the constant probability of drought. It is obvious, even from a superficial examination of the graph, not only that the fit of this expected distribution is far from satisfactory but also that there does exist some other value of p that will yield a much more adequate fit, as is indicated by the dotted curve. The reason for the inadequacy of using $p=D/T$ must be found by considering the assumptions implicit in this equality.

In considering occurrence of drought two independent distributions are dealt with—the distribution of days of drought, with which we are primarily concerned, and the distribution of days of rainfall. Granting the hypothesis that the former is random in occurrence with respect to some probability, p, it does not immediately follow that the latter is random in occurrence with respect to $1-p$. In fact, it is altogether reasonable that the occurrence of days of rainfall would vary markedly from such randomness.

In tabulating the drought data, the time unit has been restricted to 1 day, and rain has been defined as the occurrence of 0.10 inch or more of precipitation in 48 hours or less. Thus, for example, if a rain of 0.50 inch started at noon of one day and lasted until noon of the following day both days would be considered days of rainfall, even though the time interval of precipitation was less than 24 hours. This method of tabulation has undoubtedly led to a "clustering" of days of rainfall in excess of that which would be expected from random occurrence with respect to $1-p$. Some cognizance must be taken of this in the analysis.

In addition, aside from the effect of rainfall occurring "across" two or more days, it is possible that further clustering of days of rainfall may occur because, as is well known from observation, many types of storms usually extend over more than 1 day. Given a single day of rainfall, there may be an increase in the probability that the following day will be a day of rain. Such variation in the probability of rainfall from day to day could not be determined unless a separate study were made of the distribution of rainfall sequences. This is outside the scope of this paper. However, in analyzing drought, unwarranted assumptions concerning the character of distribution of days of rainfall must not be made.

On the assumption that the occurrence of drought is random it will be necessary to make some adjustment for the nonrandom clustering of days of rainfall. In effect, this grouping has resulted in a decrease in the number of independent observations—independent days observed. Hence, we must set $p=D/T_e$, where T_e is a constant to be calculated and where $(T-T_c)$ is a measure of the extent to which rainfall days are grouped in excess of such grouping as would normally be expected.

It is impractical to calculate the value of T_c directly. However, it must be kept in mind that the total number of independent days of record is actually T_c and not T. In obtaining the value of p, reference must be made only to the distribution of drought sequences. It is comparatively simple to evaluate p for annual distributions. It has been shown that for such distributions the number of expected droughts of exactly n days in length is

$$O_n = S(1-p)p^{n-1},$$

where, again, p is assumed to remain constant throughout the year. Since

$$D = \sum_{n=1}^{D} nO_n,$$

we may write

$$D = \sum_{n=1}^{D} nS(1-p)p^{n-1}.$$

Summing, and solving for S,

$$S = \frac{D(1-p)}{1-p^D(1+D-D\bar{p})}.$$

But D will always be very large, and p will always be less than 1. Therefore, $p^D=0$, nearly, and

$$S = D(1-p).$$

From which we may obtain

$$p = 1-\frac{S}{D}.$$

This value of p may be easily computed since both D and S are known from the observed sample.[13] In figure 3 the dotted line is the expected curve obtained by using this revised approximation of p. The fit is apparently good. Investigation of the distributions for a

[13] It can be readily shown that this value of p is also the statistic of maximum likelihood. For, multiplying the logarithm of the number of expected occurrences in each class by the number observed, and summing for all classes, we have

$$f(p) = \sum_{n=1}^{D} \log [Sp^{n-1}(1-p)](\text{observed } O_n),$$

which reduces to

$$f(p) = S \log S + D \log p - S \log p + S \log (1-p).$$

Differentiating with respect to p, and setting $f'(p)$ equal to 0,

$$0 = \frac{D}{p} - \frac{S}{p} - \frac{S}{1-p},$$

from which it follows that $f(p)$ is a maximum when

$$p = 1-\frac{S}{D}.$$

number of other areas indicates that distributions calculated from this revised probability are greatly preferable to distributions calculated from the hypothesis $p=D/T$. The fit is apparently adequate for those areas exhibiting no great seasonal variation in drought hazard; however, for such areas as the San Joaquin Valley, Calif., the fit of the revised curve is not good because the observed frequencies varied from a simple exponential relation.

It becomes apparent, therefore, that for longer periods the simple concept of a constant daily drought probability is inadequate. However, the effect of variability of drought hazard can be minimized or avoided by considering p constant over only very short intervals. This fact is of particular utility in devising a more adequate test of the randomness of drought occurrence.

Basic Assumptions

In the further conduct of the probability analysis three basic assumptions are made:

1. That for each station every day of the year can be assigned a value that will represent the average year-to-year probability of that particular day being a day of drought.
2. That since no significant long-term trend, cycle, or other annual change in drought occurrence has been proved, the probabilty values assigned will be considered a priori as being equally applicable to any year or group of years.
3. That although the daily probability values may vary widely from season to season, the values for the days of the year, plotted consecutively, form a smooth curve.

It is to be noted that the third assumption does not specify in any manner the actual magnitude or character of the change in drought probability from season to season. The great difficulty of dealing mathematically with a complex probability curve makes it advisable to introduce approximations of drought variation that will allow more simple treatment. Although these approximations will introduce a certain amount of error, it is felt that ordinarily the error will be too small to necessitate more accurate handling of the data. Two different approximations of variation in drought probability are used:

1. For time intervals as short as a month, the daily probability of drought is considered constant.
2. For time intervals such as a season, the daily probability of drought is considered as varying linearly.

Both of these approximations may be justified statistically. It can be demonstrated that variation of probability from a constant value or from linear change would have to be great before the distribution of drought occurrence expected from a random sample would be materially affected. Examination of the drought-probability data indicates that such variation would be the exception rather than the rule. Taking into consideration the fact that these approximations were set up for mass use (that is, for application to a large number of stations) it will be realized that the net effect of exceptional cases will be small. If the investigator is concerned with only one station, however, it might be advisable to examine the drought data more closely to determine whether assumption of a constant monthly drought probability or a linear seasonal change in drought probability is suitable.

Whenever time intervals longer than a season are considered, the assumption of linear change in probability is rejected. In addition, the approximations relating to both monthly and seasonal probability will be checked by comparing the observed frequency data with expected distributions derived from and based on the approximations. If a large error is introduced, it will become evident as a result of inadequacy of fit of the theoretical distributions. Derivation of the theoretical distributions is shown in the following section.

THEORETICAL DISTRIBUTIONS FOR MONTHLY TIME INTERVALS

In considering drought occurrence within a time interval limited to 1 month, it is assumed that the probability of any day within this time interval being a day of drought is some constant value, p. Proceeding from this assumption it has been possible (1) to determine an approximation of p from the observed data; (2) from this approximation to calculate an expected distribution, assuming random occurrence of drought; and (3) by comparing a large number of expected distributions with the observed data, to determine whether assumption of random occurrence and constant probability will permit adequate description of monthly drought occurrence.

CALCULATION OF DROUGHT PROBABILITY

Before deriving the formula for the expected distribution, it will be necessary to obtain some measure of p from the observed data. As in the case of the annual data, reference must be made only to the distribution of drought occurrence. For this purpose, the number of droughts observed in the M years of record will again form a convenient measure. Let the mean number of independent days of record in any month be N_e. Then,

$$N_e = \frac{T_e}{M}.$$

The error will not be appreciable if this mean value is considered to apply to each of the M years.

The value of S, the number of drought sequences observed, can easily be calculated in terms of M, N_e, and p. We would expect Mp droughts to start on the first day of the time interval. Also, we would expect Mpq droughts to start on each subsequent day, where $q = (1-p)$. Since there are N_e independent days of record in each month, it follows that

$$S = Mp + Mpq(N_e - 1).$$

By using the relationships $N_e = T_e/M$ and $p = D/T_e$, this reduces to

$$S = D - Dp + Mp^2,$$

from which

$$p = \frac{D - \sqrt{D^2 - 4M(D-S)}}{2M}.$$

Here D, S, and M are known from the observed data, and hence p can be easily evaluated. Also, it can be shown that

$$\sigma_p = \sqrt{\frac{p(1-p)}{D}}.$$

This value will give us a measure of the accuracy with which the years of recorded data describe the actual characteristics of a given station month.

Calculation of Expected Distribution

It will be our next task to derive a general formula for the distribution of drought sequences based upon random occurrence of drought and the constant probability, p. Before doing so, however, it will be necessary to make certain modifications in the procedure outlined above. In obtaining p, it was most convenient to consider the number of years constant at M, and the number of independent days of record in each time interval as the mean value, N_c. In considering occurrence of droughts of various length, however, we may very well be interested in sequences of days of drought longer than N_c but, of course, equal to or less than N. Consequently, it will no longer be possible to consider only N_c independent days of record in each month.

The needed adjustment can be made by considering each monthly time interval to consist of N full independent days of record and by reducing the number of independent years of record to M_c so that

$$M_c = \frac{T_c}{N}.$$

The error introduced by this adjustment will be slight.

Using the symbols previously presented, it will be possible to derive the formula for the expected distribution. Consider a drought of n days. It is desired to calculate the number of occurrences, O_n, expected. This may be done as follows:

1. The probability that a drought of exactly n days will begin on the first day of the time interval is p^nq, where $n < N$. Hence, in M_c years of record, M_cp^nq droughts will be expected to start on the first day.

2. The probability that a drought of exactly n days will begin on any particular day from the second to the $(N-n)$th day of the time interval, inclusive, is qp^nq, provided, again, that $n < N$. Hence, the total number of droughts of exactly n days expected to begin between the second and $(N-n)$th day of the time interval will be $M_c(N-n-1)p^nq^2$.

3. The probability that a drought of exactly n days will begin on the $(N-n+1)$st day of the time interval is qp^n, where $n \leq N$. Hence, in M_c years of record, M_cqp^n droughts of n days will be expected to begin on the $(N-n+1)$st day.

4. No droughts of n days can possibly begin on or after the $(N-n+2)$nd day of the time interval.

Summing the above expected amounts,

$$O_n = M_e p^n q + M_e(N-n-1)p^n q^2 + M_e p^n q,$$

which reduces to [18]

$$O_n = M_e p^n q[2 + q(N-n-1)].$$

This formula is applicable for all values of n, except $n=N$. When $n=N$, a similar evaluation will show that

$$O_n = M_e p^n.$$

These two formulas have been used in calculating expected distributions for comparison with the observed data. In these calculations p has been taken as its most likely value and the standard error of p has been disregarded. An increased correspondence between the expected and observed distributions may therefore be expected. This fact will, of course, be taken into consideration in the analysis. The standard error of p will be considered later in the calculation of likelihood of drought occurrence.

Method of Comparing Observed and Expected Distributions

Table 2 gives an example of a calculated distribution and shows also the observed data. As can be seen by comparing the distributions, the expected distribution closely approximates the observed. In making a statistical comparison for a large number of months, however, it will be necessary to employ some accurate measure of goodness of fit such as χ^2.

TABLE 2.—*Expected and observed drought distributions for North Head, Wash., for the month of August*

Length of drought (days)	Noncumulative distribution		Cumulative distribution		Length of drought (days)	Noncumulative distribution		Cumulative distribution	
	Expected	Observed	Expected	Observed		Expected	Observed	Expected	Observed
1	7.92	11	88.60	91	17	1.84	2	18.16	16
2	7.30	8	80.68	80	18	1.67	1	16.32	14
3	6.70	6	73.38	72	19	1.51	0	14.65	13
4	6.02	5	66.68	66	20	1.38	1	13.14	13
5	5.60	5	60.66	61	21	1.25	0	11.76	12
6	5.07	4	55.06	56	22	1.14	1	10.51	12
7	4.62	3	49.99	52	23	1.03	0	9.37	11
8	4.23	2	45.37	49	24	.93	3	8.34	11
9	3.90	7	41.14	47	25	.84	2	7.41	8
10	3.53	5	37.24	40	26	.76	1	6.57	6
11	3.22	6	33.71	35	27	.69	0	5.81	5
12	2.95	4	30.49	29	28	.62	1	5.12	5
13	2.70	3	27.54	25	29	.56	0	4.50	4
14	2.44	4	24.84	22	30	.51	0	3.94	4
15	2.22	2	22.40	18	31	3.43	4	3.43	4
16	2.02	0	20.18	16					

The χ^2 test will not be used exclusively. This test requires that the number of occurrences expected in each class interval be enough to

[18] The same formula was obtained by W. G. Cochran, but was applied to a different type of investigation (4).

allow the assumption of a normal distribution of sampling error about the expected values. It is therefore frequently necessary to group several terminal classes. In the present study, this would actually mean grouping together a large number of the more lengthy drought sequences, and since these are the sequences in which we are particularly interested, such groupings would be exceedingly undesirable. Although the χ^2 test will be utilized in determining whether the number of long droughts observed is excessive, some other technique is necessary to reveal whether droughts of certain specified lengths are more or less probable than we would expect them to be as a consequence of some as yet undefined meteorologic cause.

The method used in making an original rough comparison of the theoretical and the observed data is based upon calculation of the probable error about the expected values O_n.[15] In this connection, normal distribution of sampling error about O_n has been assumed only for the shorter, more frequent droughts. Whenever longer, less frequent droughts are under consideration it is assumed that the distribution about O_n is a Poisson rather than a normal distribution.

For the shorter sequences, the probable error can be expressed as $0.6745\sqrt{p_s q_s T_s}$, where p_s is the mean probability of obtaining a drought of the required length; q_s is the probability of failure, or $(1-p_s)$; and T_s is the number of trials of the event. The number of trials, T_s, can be evaluated as the number of independent days on which a new drought sequence could possibly start. Hence,

$$(T_s)_n = M_s + M_s q_s (N - n),$$

where $(T_s)_n$ is the number of trials of a sequence of n days.

From this relationship, p_s can be determined as the simple ratio of O_n and $(T_s)_n$; and hence q_s also can be found.[16] Substituting these values in the formula for the probable error, we find

$$P.E. = 0.6745\sqrt{O_n\left(1 - \frac{O_n}{(T_s)_n}\right)}$$

Where the ratio of O_n to $(T_s)_n$ is greater than 0.03, this equation, based upon normal distribution of sampling error, is used. Where the ratio is less than 0.03, however, it is necessary to determine probable error from tables of Poisson's exponential limit. Since we are desirous of obtaining the probability that a given observation will fall within or without the probable error, a method must be devised to give an approximation of this value.

From tables of Poisson's exponential binomial limit *(12, pp. 113–121)* it is possible to determine for any value, O_n, the probability of appearance of 0, 1, 2, and higher numbers of occurrences. From these values, the extreme 25 percent on either side of the center has been separated, and these marginal percentages are considered as

---

[15] Use of the probable error in these comparisons will be of additional value in that it will enable us to estimate the reliability of the probable error figures by comparing the results obtained from such an analysis with the results from a subsequent χ^2 test. Probable error is used rather than standard error since it was employed in the original preparation of the data, and the amount of recalculation necessary to change the tabulation would be prohibitive.

[16] Actually, the probability of obtaining a drought of the required length will not remain constant throughout the entire month, since the probability of obtaining a drought of n days beginning on the first day of the time interval is greater than that of obtaining a drought of that length beginning on any succeeding day. However, the effect of this variation will be negligible.

falling without the probable error. To take a simple example, where O_n is 2.2, the following are the probabilities of the various possible number of occurrences of a sequence of n days.

Probability that a sequence of n days of drought will occur 0 to 8 or more times:

0	0.1108	5	0.0476
1	.2438	6	.0174
2	.2681	7	.0055
3	.1966	8 or more	.0020
4	.1082		

These probabilities have been rearranged into the following tabulation

Probability that a sequence of n days of drought will occur 0 to 8 or more times classified below, within, and above probable error:

Below probable error:		Above probable error:	
0	0.1108	3	0.0693
1	.1392	4	.1082
Within probable error:		5	.0476
1	.1046	6	.0174
2	.2681	7	.0055
3	.1273	8 or more	.0020

It will be seen from an examination of these figures that where no occurrence or four or more occurrences are found in the observed data these points will lie entirely without the range of the probable error, and where two occurrences are found, the point will lie entirely within the probable error. Where either one or three occurrences are observed the points may lie either within or without the probable error, and for these numbers, we will express the point as being partly within and partly without, in accordance with the probabilities. For example, where three occurrences are observed, the observation will be considered to fall 0.1273/0.1966, or 0.648, within the probable error, and 0.0693/0.1966, or 0.352 above the probable error. Similarly, where one occurrence is observed, the point would be computed as 0.429 within the probable error, and 0.571 below the probable error. Tables of these values for all numbers up to 15.00 were prepared and are used in the analysis where the ratio of O_n to $(T_t)_n$ is below 0.03.

It is to be noted that although neither method of evaluating probable error can be considered exact in any one case,[17] the difference should compensate. The procedure we shall follow is to calculate for each station month the distribution expected from a random sample and the probable error about the values in the distribution. These values are compared with the observed data for each O_n from $n=1$ and up, and it is determined whether each observed number of occurrences falls within, above, or below the probable error. Comparisons are made for all O_n equal to or greater than 0.01. Of the total number of observations of O_n, 25 percent would be expected to fall above the probable error; 50 percent within; and 25 percent below. Considering that drought distributions were analyzed for 12 months at each of 40 stations and that the 480 station months yield over 14,000 observations of O_n, the 1:2:1 proportion should hold with a high degree of accuracy.

It must again be pointed out, however, that many of the values of O_n for longer, less probable droughts (where p_s is less than 0.03) are considered as the means of Poisson distributions and that corre-

[17] This is particularly true where p_s is only slightly above or below 0.03.

sponding observations are sometimes considered as being partly above and partly within, or partly within and below the probable error. This fact should cause the totals of points above, within, and below the probable error to correspond more closely to a 1:2:1 ratio than might otherwise be expected. As a result, the probable error test will have, at best, only a negative value. That is, although it is altogether likely that any marked deviation of the observed data from randomness will become obvious, it does not follow that good results from the probable error test (close correspondence to a 1:2:1 ratio) necessarily mean that the assumption of random occurrence is correct.

Table 3 shows the results obtained for a total of 480 station months. Out of the 14,461 observed O_n values, 3,692.66, or 25.53 percent, fell above the probable error; 7,075.63, or 48.93 percent, fell within; and 3,692.71, or 25.54 percent fell below. Although these proportions are close to the expected 1:2:1 ratio the sample is large, and it is desirable to test whether the differences are significant. Testing by the χ^2 method, we find that χ^2 is 6.64 for two degrees of freedom, and $P=0.04$. Hence the deviations from a 1:2:1 proportion are evidently significant.

TABLE 3. *Number of O_n observations falling above, within, and below the probable error about the expected monthly distributions for 480 station months at 40 stations*

Length of drought (days)	Observations of O_n	Above probable error	Within probable error	Below probable error	Length of drought (days)	Observations of O_n	Above probable error	Within probable error	Below probable error
1	480	218.72	185.09	76.19	18	480	117.89	245.69	116.41
2	480	173.08	215.50	91.42	19	480	122.03	237.31	120.66
3	480	1[illegible].29	233.75	[illegible]	20	480	122.11	244.50	113.39
4	480	[illegible]	228.57	132.86	21	479	117.80	238.08	123.12
5	480	125.12	233.66	121.22	22	478	114.09	235.09	128.82
6	480	[illegible]	226.08	[illegible]	23	478	124.54	247.20	106.26
7	480	[illegible]	[illegible]	[illegible]	24	472	125.23	227.57	119.20
8	480	[illegible]	227.91	[illegible]	25	470	128.16	227.96	113.89
9	480	[illegible]	227.94	[illegible]	26	468	119.88	230.68	117.45
10	480	[illegible]	[illegible]	[illegible]	27	462	127.32	227.02	107.66
11	480	[illegible]	218.10	[illegible]	28	458	126.83	218.51	112.66
12	480	[illegible]	247.48	[illegible]	29	418	100.37	212.49	105.14
13	480	[illegible]	243.65	140.70	30	411	111.79	202.09	100.32
14	480	[illegible]	[illegible]	108.41	31	261	71.23	138.98	53.79
15	480	[illegible]	241.[illegible]	131.88					
16	480	[illegible]	247.08	[illegible]	Total	14,461	3,692.66	7,075.63	3,692.71
17	480	129.60	223.42	128.98					

To test whether these deviations may be attributed to any particular drought lengths, the χ^2 test may be applied to the different drought lengths as classes. The value of χ^2 for 60 degrees of freedom was 235.91. The probability of obtaining this large a χ^2 by chance is negligible. However, by far the larger part of the total is contributed in the lower sequences. The χ^2 value for $n=1$ is 109.77, or almost half the total. Further examination of table 3 will indicate that there is a large excess of points above the probable error for droughts of 1 to 4 days and an excess of points below the probable error for medium droughts, 6 to 16 days. Beyond that point the 1:2:1 ratio apparently holds. Considering only values of n equal to or greater than 16, $\chi^2=18.11$ for 30 degrees of freedom. This corresponds to a value of P greater than 0.95. This is, as we would expect, an abnormally high value, and probably reflects the effect of dividing single observations between two or even three classes.

It can be seen that although the results obtained are purely negative, they are, nevertheless, of value. There is a possibility that droughts of 16 days or more may be random in occurrence, but shorter droughts cannot be considered random. If randomness of occurrence is related to length of drought, it remains to explain by some meteorological theory why this should be true.

One theory immediately suggests itself. Excess of short droughts, particularly 2 or 3 days, may be due primarily to the interval between precipitation from successive frontal systems. Although the mean time interval between successive fronts would most likely be 4 to 6 days, this would be disguised by the manner in which the data are tabulated. Precipitation of 0.10 inch or more in 48 hours has been defined as terminating a drought. Hence, for example, a sequence of 6 days in which 0.47, 0.04, 0, 0, 0.01, and 0.19 inch occurred would be tabulated as a 2-day drought although the actual time between fronts might well be 4 full days. The great excess of single days of droughts may be due to the random occurrence of scattered pre- and post-frontal showers, which would tend to reduce a drought of 3 or 4 days to two shorter droughts.

If this interpretation is correct, we would expect to find occurrence of a single day of drought to deviate a maximum amount from the expected figure during the summer months, when scattered showers, superimposed on a frontal pattern, are most frequent. This is observable in the data. During the three summer months an average of 54.8 percent of the O_n observations (where $n=1$) fall above the probable error. This compares with an average of only 45.6 percent for 12 months.

The same theory of recurrent frontal passage is adequate to explain the difference between the observed and expected occurrence of medium-length droughts. In most areas in the United States it would be rather unusual for 7 or more days to go by without passage of a front. That medium-length droughts are less frequent than would be expected can therefore also be explained by the association between precipitation and frontal action. In addition, occurrence of long droughts—about 16 days or more—would indicate either that some unusual meteorological condition caused a break-down in the usual pattern of recurrent frontal precipitation or that the station in question is in an area in which precipitation is infrequent or not generally associated with frontal passage. In either event the effect of the recurrence of fronts would be eliminated and we would expect the number of long droughts observed to approximate closely the expected values.

Interpreting these meteorological conditions statistically, we see that they adequately explain the apparent discrepancies between the observed and expected data. The question remains, however, whether occurrence of droughts over 16 to 20 days in length is actually random. To test this possibility more accurately, it will be necessary to resort to techniques other than the use of the probable error.

In continuing the analysis the χ^2 test is used, primarily to determine whether the number of long droughts observed is in accordance with the theory of random occurrence. However, it is hoped that the conclusions obtained above can also be verified by such an examination. The distributions are set up to include five classes, in which the number of days of drought are 1 to 4, 5 to 9, 10 to 14, 15 to 19, and 20 to 31.

In addition, the distributions for the 12 months of the year were combined for each station to make certain that the number of expected occurrences in each class would exceed a minimum of 5. Table 4 shows the results obtained. The χ^2 for the distribution at each station is shown in the last column. Since the constant, p, was determined from the observed data, not from a priori consideration, the number of degrees of freedom for each χ^2 value must be taken as 4, rather than 5.[15]

TABLE 4. χ^2 *for monthly drought distributions, showing contribution of each class interval to the total*

Station	Length of drought					χ^2
	1-4 days	5-9 days	10-14 days	15-19 days	20 days and over	
Ann Arbor, Mich	0.79	1.05	2.23	0.00	0.05	4.12
Billings, Mont	5.61	1.55	.11	1.11	1.29	10.00
Bisbee, Ariz	5.53	1.50	2.51	.37	1.68	11.59
Boise, Idaho	6.11	1.43	.43	.36	.02	11.35
Bottineau, N. Dak	2.8[illegible]	.17	1.07	1.42	.40	6.18
Brownsville, Tex	3.93	2.23	1.12	.82	.01	8.12
Chico, Calif	14.02	10.11	3.20	.45	3.42	31.20
Colorado Springs, Colo	.81	.07	.58	.08	1.34	2.88
Dalhart, Tex	.61	.00	.27	1.11	.00	2.05
Dallas, Tex	.63	.28	.30	.03	.09	1.33
Duluth, Minn	2.87	.42	.00	.42	.09	3.80
Escanaba, Mich	5.06	.97	.19	.53	.08	6.83
Eureka, Calif	18.76	10.26	.56	3.19	.58	33.35
Franklin, La	.80	.01	.02	.16	2.52	3.60
Fresno, Calif	5.61	2.33	1.93	1.04	1.25	12.10
Greensboro, Ala	.01	1.91	.30	.04	.12	2.38
Hatteras, N. C	.99	1.04	.58	.49	.26	3.76
Hays City, Kans	3.45	2.19	.08	.02	.05	5.79
Independence, Calif	3.78	4.56	1.48	.32	.11	10.25
Key West, Fla	1.32	.00	.16	1.93	.41	3.82
Kingston, Tenn	1.63	.01	2.30	.01	.01	3.96
Lake Placid, N. Y	.16	3.41	.15	.06	.03	3.81
Le Mars, Iowa	4.23	.92	.67	.48	1.06	7.36
Lexington, Ky	2.77	.19	.50	.77	.02	4.25
Marengo, Ill	1.57	.27	.09	10.20	3.58	15.71
Muskogee, Okla	1.31	.21	.09	1.18	.89	4.40
New Haven, Conn	.49	.45	.50	7.19	.08	6.93
Northhead, Wash	15.23	2.46	.01	.09	.43	18.22
North Platte, Nebr	3.95	2.21	.61	3.59	.06	10.42
Omaha, Nebr	4.73	1.53	5.05	3.44	.22	14.97
Pierre, S. Dak	1.75	.60	.02	.35	.02	2.77
Portland, Maine	1.63	.39	.01	1.84	.00	3.87
Port Oxford, Oreg	3.65	.26	.08	.03	.03	4.05
Rockingham, N. C	.85	.26	.95	.03	1.27	3.36
St. Augustine, Fla	1.19	.31	2.31	1.66	2.81	8.28
Salt Lake City, Utah	.97	2.88	.39	.07	1.21	5.52
San Diego, Calif	13.74	5.78	1.37	1.18	.84	26.91
Tucumcari, N. Mex	.15	.32	.01	.16	.00	.94
Winnemucca, Nev	9.38	.40	2.63	2.61	1.26	16.28
Yuma, Ariz	.83	1.73	1.23	4.04	1.12	12.25
Total	151.2[illegible]	68.03	40.45	51.50	28.87	343.06

Examination of table 4 indicates that variation from randomness definitely does exist. For 4 degrees of freedom, $P = 0.05$ where $\chi^2 = 9.488$. It is to be noted that 14 of the 40 χ^2's exceed this limit. In addition, in the entire table, $\chi^2 = 343.06$ for 160 degrees of freedom. The likelihood that this deviation could have occurred by chance is negligible.

[15] R. A. Fisher has pointed out (6, pp. 92, 280) that where an adjustable parameter is determined from the observed data, a corresponding adjustment must be made in applying the χ^2 test. Since the total number of sequences has not been held the same in the expected and observed distributions, the number of degrees of freedom would be exactly equal to the number of classes if p were determined a priori. Since p was not so determined, however, the number of degrees of freedom must be taken as one less than the number of classes. As there are five class intervals, the number of degrees of freedom is four.

It is important to note, however, that the greater part of the χ^2 total of 343.06 is included in the first and second columns. In this table, it is not feasible to separate the contributions made by the individual degrees of freedom. However, it may be noted that for 160 degrees of freedom, P would equal 0.05, where χ^2 was 190.24. This total distributed evenly between the five columns would give approximately 38.05 for the total of each class. In table 4, only class 5, droughts of over 20 days, does not exceed this limit.

The results obtained from the χ^2 test, then, apparently corroborate the conclusions reached through use of the probable error. Droughts under 16 to 20 days definitely do not follow the expected distribution; beyond that point, the expected distribution of drought may adequately represent observed occurrence. Although it would be possible to continue with further analysis, it is felt that such investigation would not be profitable with the monthly data. If we admit that the occurrence of droughts of less than 20 days is not random, observational intervals limited to only 30 or 31 days are too short to be of significant value. In the following section, the investigation is extended to longer periods of time. Thereby, the number of long droughts available for comparison with expected distributions is increased greatly.

THEORETICAL DISTRIBUTION OF DROUGHT FOR SEASONAL TIME INTERVALS

It has been shown in the previous section that the hypothesis of random occurrence of drought is not exactly correct. However, it has also been demonstrated that the hypothesis may be valid for longer drought intervals. To investigate this question, the theory of random occurrence is extended to time intervals of more than 1 month. For these longer intervals, the large differences between mean probability values obtained for successive months at many stations indicate that it is necessary to discard the idea of a constant drought probability.

In the pages that follow, an expected drought distribution for seasonal time intervals is derived. Random occurrence is assumed. Variation in drought probability is approximated as linear. One further approximation is made. In calculating the monthly probabilities, the value M_e was obtained for each month. This value represents the number of independent years of record at a given station. The value M_e varies from month to month not only because of differences in the length of record of the months but also because of differences in the extent of rainfall persistence. In order to simplify calculation of seasonal distributions, the average value of M_e for the months of a season under consideration is taken as applying over the entire season.

The net effect of these approximations will undoubtedly be to increase the differences between the theoretical and observed distributions. However, the errors introduced will not in general be sufficient to impair the representativeness of the calculated distributions. In determining the seasonal distributions, it will be expedient to calculate the cumulative rather than the noncumulative expected distribution. That is, the expected number of occurrences of a drought of n days or more, O_N, will be obtained. These cumulative values are more simply evaluated and, in addition, are of greater utility in estimating future likelihood of drought.

Expected seasonal distributions are evaluated for 8 stations: New Haven, Conn.; Greensboro, N. C.; Marengo, Ill.; Bottineau, N. Dak.; Dalhart, Tex.; Boise, Idaho; Fresno, Calif.; and Yuma, Ariz. Cumulative values were calculated for n equal to 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, and 80, and by taking first differences a series of noncumulative groupings were obtained. By using a limited number of widely separated stations the areal dependence between stations has been reduced or eliminated. Careful examination of drought occurrence at these 8 stations will afford a fair basis for evaluating the validity of our assumptions.

Calculation of Drought Probability

To derive the formula for the expected distribution it will be necessary to define mathematically the assumed linear variation of drought probability. Seasonal time intervals are defined to include 3 months. Let p_1, p_2, and p_3 be the observed monthly drought probabilitites for the 3 months in a given season. Then the linear variation of drought probability for the season will be defined as the straight line fitted by least squares to these three values, considering the monthly drought probabilities as centered within their respective months.

Consider the probability, y, as a function of time, $f(x)$, such that $x=0$ is the center of the season. The independent variable, x, will then vary from $-N/2$ to $N/2$, where N is the total number of days in the season. Considering the number of days in a month to be constant at 30, the point p_1 will then be the observed value of $f(x)$, where $x=-30$; the point p_2 will be the observed value of $f(x)$, where $x=0$; and the point p_3 will be the observed value of $f(x)$, where $x=30$. Fitting these points by least squares yields the linear equation

$$y=\frac{p_1+p_2+p_3}{3}+\frac{p_3-p_1}{60}x.$$

Let

$$p=\frac{p_1+p_2+p_3}{3},$$

and

$$m=\frac{p_3-p_1}{60}.$$

Then the equation will read

$$y=p+mx.$$

The constants can, of course, be readily evaluated from the known probability values for the 3 months. The constant, p, represents the mean drought probability for the season; m indicates the slope, or rate of change, of probability from day to day within the season.

The Seasonal Expected Distribution of Drought

Let p_x be the probability of a drought of n days or more beginning on the day $(x+1)$. Then, in M_s independent years of record, we would expect $M_s p_x$ droughts of n days or more to start on the day $(x+1)$.

Note that the values $M_c p_0, M_c p_1, \ldots, M_c p_x, \ldots$ form a Lexis distribution and that the total number of droughts, O_N, expected in the time interval of N days will hence be

$$O_N = M_c \sum_{x=-\frac{N}{2}+1}^{\frac{N}{2}-n} p_x.$$

To evaluate this sum, consider two independent cases: Where $x>[-(N/2)+1]$ and where $x=[-(N/2)+1]$. Only where $x=[-(N/2)+1]$ is it not necessary to assume the condition of a day of rainfall preceding the drought interval.

Where $x>-(N/2)+1$,

$$p_x=[1-f(x)]\prod_{i=1}^{n} f(x+i),$$

or

$$p_x=\prod_{i=1}^{n} f(x+i)-\prod_{i=0}^{n} f(x+i).$$

But the errors will be very small and will almost exactly compensate if we write

$$\prod_{i=1}^{n} f(x+i)=\left[\frac{f(x+1)+f(x+n)}{2}\right]^n,$$

and

$$\prod_{i=0}^{n} f(x+i)=\left[\frac{f(x)+f(x+n)}{2}\right]^{n+1},$$

from which

$$p_x=\left[\frac{f(x+1)+f(x+n)}{2}\right]^n-\left[\frac{f(x)+f(x+n)}{2}\right]^{n+1}.$$

However, since m will always be very small, $f(x)=f(x+1)$, almost, and the error will be small if we write

$$p_x=\left[\frac{f(x+1)+f(x+n)}{2}\right]^n-\left[\frac{f(x+1)+f(x+n)}{2}\right]^{n+1}.$$

To total the number of droughts of n or more days expected to begin on or after the second day of the time interval, we may approximate by setting

$$\sum_{x=-\frac{N}{2}+1}^{\frac{N}{2}-n} p_x=\int_{-\frac{N}{2}+\frac{1}{2}}^{\frac{N}{2}-n+\frac{1}{2}}\left(\left[\frac{f(x+1)+f(x+n)}{2}\right]^n-\left[\frac{f(x+1)+f(x+n)}{2}\right]^{n+1}\right)dx.$$

But since

$$f(x)=p+mx,$$

we have

$$f(x+1)+f(x+n)=2p+2mx+m+mn,$$

and substituting, simplifying, and integrating,

$$\sum_{x=-\frac{N}{2}}^{\frac{N}{2}-n} p_x = \frac{\left[p+\frac{m}{2}(N-n)\right]^{n+1}}{m(n+1)} - \frac{\left[p+\frac{m}{2}(N-n)\right]^{n+2}}{m(n+2)} - \frac{\left[p-\frac{m}{2}(N-n)\right]^{n+1}}{m(n+1)} + \frac{\left[p-\frac{m}{2}(N-n)\right]^{n+2}}{m(n+2)}.$$

Let

$$V = p - \frac{m}{2}(N-n),$$

and

$$Z = p + \frac{m}{2}(N-n).$$

Then:

$$\sum_{x=-\frac{N}{2}}^{\frac{N}{2}-n} p_x = \frac{Z^{n+1}}{m(n+1)} - \frac{Z^{n+2}}{m(n+2)} - \frac{V^{n+1}}{m(n+1)} + \frac{V^{n+2}}{m(n+2)}.$$

By a similar method of approximation, it can be shown that

$$p_{\left(-\frac{N}{2}+1\right)} = V^n,$$

where $p_{\left(-\frac{N}{2}+1\right)}$ is the probability of a drought beginning on the first day of the time interval, N.[19]

Hence

$$O_N = M_e\left[V^{n+1}\left(\frac{1}{V} - \frac{1}{m(n+1)} + \frac{V}{m(n+2)}\right) + Z^{n+1}\left(\frac{1}{m(n+1)} - \frac{Z}{m(n+2)}\right)\right].$$

This formula is used in calculating the expected cumulative distribution of drought sequences for seasonal time intervals. First differences are taken to obtain the noncumulative distribution.

Standard Error in the Expected Distribution

In comparing the seasonal expected drought distributions with the observed data, it is sufficient to use the χ^2 test only. Testing by means of the standard or probable error would be of little value since there are a limited number of comparisons available (only 32 distributions as against 480 in the previous section). However, the standard errors in the expected distributions due to sampling should be calculated as a measure of the reliability of the results.

In obtaining the standard error of the expected figures, use may again be made of the fact that the values $M_e p_0, M_e p_1, \ldots, M_e p_{x_1}$

[19] This approximation introduces a slight error, which is always opposite in sign to the error introduced by writing $f(x+1)$ for $f(x)$, above.

. . . form a Lexis distribution. The standard deviation, σ_L, of the above series can immediately be expressed as

$$\sigma_L = \sqrt{M_e p_s(1-p_s) + \frac{M_e^2 - M_e}{N-n+1} \sum_{i=0}^{N-n} (p_i - p_s)^2},$$

where, as before, p_s is the mean probability of a drought of n or more days, or

$$p_s = \frac{1}{N-n+1} \sum_{i=0}^{N-n} p_i.$$

But the standard error of the mean of the Lexis distribution is

$$\frac{\sigma_L}{\sqrt{N-n+1}}.$$

Hence it follows that the

$$\text{S. E. of } O_N = \sigma_L \sqrt{N-n+1}.$$

The value of σ_L could be obtained and substituted in the above expression. However, the formula for σ_L is too unwieldy for practical use, and hence it is desirable to approximate the standard error by some other method.

The question that immediately arises is whether consideration of the series $M_e p_1$, $M_e p_2$, . . ., $M_e p_x$, . . . as a simple Bernoulli distribution would involve serious error. This would be equivalent, of course, to setting m equal to 0, so that $p_1 = p_2 = p_3 = \ldots = p_x = \ldots = p_s$. The standard error in O_N would then reduce to

$$\text{S. E. of } O_N = \sigma_B \sqrt{N-n+1},$$

which it may easily be shown is the equivalent of evaluating the standard error as

$$\text{S. E. of } O_N = \sqrt{(T_s)_n \frac{O_N}{(T_s)_n}\left(1 - \frac{O_N}{(T_s)_n}\right)},$$

where $(T_s)_n$ is the number of trials of a sequence of n or more days, and $(T_s)_n$ may, again, be obtained from the relationship

$$(T_s)_n = M_e + M_e q(N-n).$$

Note that the percentage error introduced in the value of the standard error by considering drought probability constant is

$$\frac{\sigma_B - \sigma_L}{\sigma_B}.$$

No simple formula can be derived for determining the maximum of this expression. However, repeated evaluation of the ratio under the most unfavorable conditions seems to indicate that in no case would the error be greater than 10 percent and that usually the error would be considerably less. As a result, it is considered accurate enough to set the standard error equal to

$$\sqrt{O_N\left(1-\frac{O_N}{(T_s)_n}\right)}.$$

This formula should be used whenever the ratio $O_N/(T_s)_n$ is greater than 0.03. When the ratio falls below this value, it should again be assumed that the sampling error has a Poisson distribution, and reference should be made to tables of the Poisson exponential binomial limit.[20]

Comparison of Expected and Observed Distributions

In comparing the theoretical and observed data by the χ^2 test, reference is made only to the noncumulative distributions to eliminate dependence between successive groups. Thirty-two distributions are available for analysis. Within each distribution droughts are grouped together in class intervals of 5 days up to $n = 50$, beyond which the class interval is raised to 10. The first question that arises is whether the data as a whole can possibly be considered random in occurrence. In order to make a first rough test, the 32 noncumulative distributions will be grouped together and the χ^2 test applied to determine whether the general theoretical distribution of drought is adequate. Table 5 shows the values obtained by this grouping. The value of χ^2 obtained, 45.62 for 13 degrees of freedom, indicates that the chance of the theoretical distribution being representative of the observed data is negligible. However, it is to be noted that the larger part of this total of 45.62 is accumulated in the first four groups. Far too many very short droughts and too few medium-length droughts are observed. This is entirely in accord with the results indicated in the preceding section.

TABLE 5.—*Total occurrence of drought for 32 seasons at 8 selected stations*

Length of drought (days)	Drought distributions: A (observed)	Drought distributions: B (expected)	$\frac{(A-B)^2}{B}$	
1-4	5,314	5,048.24	13.99	......
5-9	2,632	2,720.04	3.85	......
10-14	1,128	1,171.01	1.58	......
15-19	491	560.19	8.55	......
1-19	9,565	9,499.48	......	0.45
20-24	309	307.36	.01	.01
25-29	163	185.27	2.68	2.68
30-34	140	121.34	2.87	2.87
35-39	67	85.36	3.95	3.95
40-44	58	63.21	.43	.43
45-49	40	47.60	1.21	1.21
50-59	79	67.20	2.07	2.07
60-69	48	43.93	.38	.38
70-79	37	28.75	2.37	2.37
80 and over	100	87.85	1.68	1.68
Total	10,606	10,537.35	45.62	18.10

Grouping together all droughts of 19 days or under reduces the value for χ^2 to 18.10 for 10 degrees of freedom. The corresponding value of

[20] For a more complete explanation, refer to p. 17.

P is slightly above Fisher's suggested "limit of significant deviation" of 0.05 (*6*). Hence it would seem that there is a definite chance that occurrence of droughts of more than 19 days is random. Certainly, more detailed examination of the problem is in order.

In continuing the investigation, individual χ^2 tests will be made of each of the 32 seasonal distributions. To group the data so that the number of expected occurrences in each class is above 5.00 it will frequently be necessary to consolidate several classes of longer droughts. All droughts of less than 20 days will be grouped together. Above that value, each grouping will be expanded upward until it includes more than 5.00 expected occurrences. If this procedure leaves a final group of less than 5.00 this terminal group will be thrown in with the preceding class.[21]

Table 6 shows the resulting expected and observed distributions for each of the 32 seasons. For each season χ^2 has been evaluated and P has been obtained, the number of degrees of freedom used being 1 less than the number of classes. Most of the values of P are reasonably high; however, a few (notably that for Boise, Idaho, summer season) are excessively low. It is desired to determine whether these extremely low values of P can be traced to some defect in the theoretical distributions.

TABLE 6.—*Expected and observed seasonal drought distributions for eight selected stations*

Length of drought (days)	Spring Expected	Spring Observed	Summer Expected	Summer Observed	Autumn Expected	Autumn Observed	Winter Expected	Winter Observed
DALHART, TEX.								
1–19	182.48	187	299.05	309	123.73	135	64.70	65
20–24	13.95	11	9.75	9	11.53	17	10.00	6
25–29	9.33	5			7.81	8	8.10	2
30–34	6.33	7			5.43	8	6.53	9
35–39			8.62	6	6.35	5	5.25	7
40–44	11.88	8					7.61	6
45–49					5.31	6		
50 and over							12.16	17
χ^2	4.08		1.19		5.22		9.97	
P	.392		.551		.366		.125	
MARENGO, ILL.								
1–19	466.99	478	524.30	527	435.51	443	347.73	355
20–24					8.05	10	13.63	14
25–29	13.00	8	8.57	8	6.30	8	6.75	3
30 and over							6.51	6
χ^2	2.18		0.10		0.71		2.28	
P	.141		.756		.701		.516	
BOISE, IDAHO								
1–19	321.39	331	90.88	95	200.01	213	309.63	331
20–24	13.39	13	11.99	7	17.70	18	10.00	13
25–29	7.01	11	9.24	1	12.28	8		
30–34			7.19	7	8.54	8		
35–39			5.66	7	5.00	3		
40–49	7.50	9	8.01	6	6.91	6	7.39	8
50–59			5.01	5	5.00			
60 and over			7.32	18		7		
χ^2	2.71		25.85		4.72		1.30	
P	.437		.000536		.575		.522	

[21] In 3 distributions the grouping together of all droughts of 1 to 19 days left a final group of less than 5.00. In these distributions, in order to retain at least 1 degree of freedom in the distribution, the 2 classes have been allowed to stand. However, the error introduced will not be significant. In 2 out of the 3 instances the number of expected occurrences in the terminal group is greater than 4.00, although less than 5.00; in the third, the number of expected occurrences is 2.98.

TABLE 6.—*Expected and observed seasonal drought distributions for eight selected stations*—Continued

GREENSBORO, N. C.

Length of drought (days)	Spring Expected	Spring Observed	Summer Expected	Summer Observed	Autumn Expected	Autumn Observed	Winter Expected	Winter Observed
1-19	540.89	532	555.21	552	371.76	368	617.51	595
20-24					11.29	13		
25-29	8.82	9	7.24	8	7.26	5	4.66	4
30 and over					6.95	5		
χ^2	0.41		0.10		1.41		0.91	
P	.525		.756		.703		.343	

FRESNO, CALIF.

Length of drought (days)	Spring Expected	Spring Observed	Summer Expected	Summer Observed	Autumn Expected	Autumn Observed	Winter Expected	Winter Observed
1-19	190.86	187	6.50	8	65.03	80	296.25	292
20-24	13.69	12			10.11	15	15.88	23
25-29	9.19	11			8.95	14	8.95	4
30-34	6.40	11			8.13	11		
35-39					7.45	3		
40-44	7.72	8			6.67	6		
45-49			10.30	6	6.05	2		
50-59					10.14	8	10.77	16
60-69	6.68	15			7.33	5		
70-79					5.05	5		
80 and over			31.87	36	5.63	3		
χ^2	14.15		2.69		17.50		8.53	
P	.015		.261		.004		.036	

YUMA, ARIZ.

Length of drought (days)	Spring Expected	Spring Observed	Summer Expected	Summer Observed	Autumn Expected	Autumn Observed	Winter Expected	Winter Observed
1-19	35.90	20	26.44	36	10.28	47	79.86	76
20-24	5.86	7	5.15	5	7.80	4	12.55	10
25-29					6.88	9	10.50	12
30-34	9.27	13	9.86	11	6.06	8	8.41	4
35-39					5.35	3	6.81	5
40-44	6.81	6	9.76	5			5.78	8
45-49					8.86	5		
50-59	7.32	11	14.29	14	6.81	7	10.09	10
60-69			8.63	5	5.21	4		
70-79	5.86	12	7.41	5			10.80	12
80 and over	13.50	11	15.76	14	14.52	18		
χ^2	17.16		8.69		8.08		4.68	
P	.00871		.275		.421		.699	

NEW HAVEN, CONN.

Length of drought (days)	Spring Expected	Spring Observed	Summer Expected	Summer Observed	Autumn Expected	Autumn Observed	Winter Expected	Winter Observed
1-19	556.94	572	554.45	550	476.26	483	592.86	589
20 and over	4.29	5	5.38	8	11.22	11	2.98	3
χ^2	0.53		1.32		0.10		0.025	
P	.470		.252		.756		.880	

BOTTINEAU, N. DAK.

Length of drought (days)	Spring Expected	Spring Observed	Summer Expected	Summer Observed	Autumn Expected	Autumn Observed	Winter Expected	Winter Observed
1-19	253.09	253	445.07	450	219.71	206	133.50	141
20-24	18.65	16			14.85	21	15.51	16
25-29	12.16	11			8.85	8	11.28	10
30-34	7.86	8			5.13	11	8.24	7
35-39	5.09	3	9.84	12			6.03	4
40-49					8.32	9	7.48	4
50 and over	8.51	11					7.03	15
χ^2	2.08		.74		9.26		10.20	
P	.835		.395		.055		.116	

One of the places where error might be introduced in obtaining the expected distribution is in the assumption of a linear change in drought probability. If the season in question is centered about an emphatic maximum or minimum in the drought probability curve, the straight line fitted to the three values would obscure the true character of the change in drought probability, since m would be equal or almost equal to 0. The fit of the straight line would be inadequate mathematically and climatologically. However, in appraising the adequacy of the fit it should be remembered that a difference of, say, 0.01 for a low value of p would be much less important than a similar error for a high value of p, since the standard error of p is, in general, greater for low values.

To obtain some mathematical measure of the adequacy of fit of the straight line $f(x)$, the values of $f(-30)$, $f(0)$, and $f(30)$, corresponding to the monthly p_1, p_2, and p_3, have been obtained and the sum of the squares of the deviations of these fitted values from the observed probabilities have been calculated in terms of the standard errors of p_1, p_2, and p_3 as units. For most seasons the fit is fairly close. However, it is notable that for Boise, Idaho, summer season, the sum of the deviations squared is 18.15, by far the highest value obtained. This means, of course, that the fit of the straight line is inadequate and that as a result the theoretical distribution must also be erroneous.[22] For every season but two (Marengo, Ill., summer, and Greensboro, N. C., summer) a very poor linear fit was accompanied by a low value of p. For these seasons, the lack of correlation is obviously due to the fact that both of the seasons are humid. There are only two classes in the distributions: Droughts of 1 to 19 days and droughts of 20 or more days. Hence, in both the poor linear fit is comparatively unimportant.

In further analysis of the distributions, all seasons are discarded for which the sum of the squares of the deviations of fitted probabilities from observed probabilities is greater than 5.00. This necessitates disregarding five seasons, as indicated in table 7. There is a considerable break between the adequacy of linear fit for these 5 seasons and for the other 27. The sixth highest value for the sum of the squares of the deviations is 4.54, for Boise, Idaho, spring season. This represents a considerable drop from 8.98. There seems to be no significant correlation between the fit of the straight lines and the values of P for the remaining 27 seasons.

TABLE 7.—*The five seasons for which the sum of the squares of the deviations of fitted probabilities from observed values was greater than 5.00, shown with associated values of P*

Station	Season	Sum of squares	P
Boise, Idaho	Summer	18.15	0.0065
Marengo, Ill	do	10.48	.756
Fresno, Calif	Spring	9.77	.0117
Greensboro, N. C	Summer	9.44	.756
Dalhart, Tex	Winter	8.98	.125

Table 8 presents the values obtained for P by station and season. These figures are combined for further analysis. In table 8, the last

[22] Thus, for Boise, summer season, far too many extremely long droughts were observed and too few medium-length droughts (20 to 30 days). This can be directly traced to erroneous assumption of linear change in drought probability. For this season, the first of the three monthly observed probabilities was low and was followed by two extremely high probabilities. The straight line fitted to these three values had a positive slope but did not reach a significantly high value until the latter part of the season, when, of course, it actually became even higher than the observed p value. Hence, the distribution calculated from the assumed linear fit naturally showed too few very lengthy droughts and too many medium-length droughts, since the probabilities assumed for the latter days of the season were abnormally high.

column indicates the probability for each station that the differences between the expected and observed distributions is no greater than could have arisen from sampling error. These probabilities were obtained by combining the seasonal values of P in the conventional manner.[25] The figures in the bottom row indicate the probability that the five to eight distributions for each season could have been obtained by chance.

TABLE 8.—*Value of P for 27 seasonal drought distributions*

Station	Spring	Summer	Autumn	Winter	Combined probability for station
Marengo, Ill	0.111		0.702	0.516	0.424
Dalhart, Tex	.492	0.554	.386		.511
Fresno, Calif		.264	.064	.036	.021
Yuma, Ariz	.009	.275	.421	.699	.070
Boise, Idaho	.152		.577	.522	.661
Greensboro, N. C	.525		.703	.343	.654
New Haven, Conn	.670	.272	.756	.881	.746
Bottineau, N. Dak	.857	.967	.655	.116	.136
Combined probability for season	.128	.348	.320	.270	

No undue variation from station to station or from season to season can be observed. Combining all 27 P values without reference to season or station yields a probability of 0.129 that the differences between the expected and observed distributions could have occurred by chance. Combining the eight values for the stations, $P = 0.134$, and combining the four values for the seasons, $P = 0.195$. Each of these values is definitely in excess of Fisher's suggested "limit of significant deviation" of 0.05 *(6)*. Hence, there is, apparently, no reason for discarding the hypothesis that occurrence of a drought of more than 20 days is random, in accordance with the assumptions originally stated.

Emphasis may here be placed on the approximations resorted to in obtaining the expected distributions. Assumption of linear change in drought probability is certainly not strictly accurate, and this inaccuracy, together with the approximations made in deriving the formula for the expected distributions, has probably tended to decrease the representativeness of the theoretical distributions. It is altogether likely that if curvilinear drought variation were assumed and more accurate formulas applied the theoretical distributions would vary even less from the observed data. However, it is entirely questionable whether the data warrant such detailed techniques. The method employed has enabled calculation of theoretical distributions that are known to have fairly good correspondence with observed fact, and hence the method has served its purpose and can be put to valuable use. To summarize, it can be fairly stated that when assumption of linear change in drought probability within a season is not obviously erroneous, the theoretical distribution of drought sequences of more than 20 days, based upon the assumption of random occurrence, adequately represents the observed data.

[25] That is, a χ^2 test has been applied to the distribution of the individual P values, based upon the fact that the sum of a number of values of χ^2 is itself distributed in the χ^2 distribution and that $\log_e P = -\frac{1}{2}\chi^2$ for 2 degrees of freedom. For further description of this method, see R. A. Fisher *(6, pp. 103-106)*.

THEORETICAL DISTRIBUTION OF DROUGHT FOR ANNUAL OR GENERAL TIME INTERVALS

In the two preceding sections, a rough test has been made of the hypothesis that occurrence of drought is random with respect to a variable probability. Because of the difficulty of dealing directly with a curvilinear variable, simple approximations of drought probability have been devised for time intervals limited to one season or less. The expected distributions calculated by using these approximations have been shown to represent observed drought occurrence adequately for droughts in excess of 20 days. Since it is known that the assumption of a drought probability constant over a month or varying linearly over a season cannot be exactly correct, the adequacies of the expected distributions calculated from these assumptions will be taken as evidence of the adequacy of our original hypothesis of random occurrence with respect to a drought probability that varies in some complex and as yet unspecified manner. In this section annual curvilinear variation in drought probability is defined as a mathematical function, and the use of the annual curve in the analysis of drought occurrence for any time interval is illustrated.

In estimating frequency of occurrence of drought, one is seldom interested in a time interval exactly a month or a season in length. More often, a part or all of the growing season of a particular crop will be of special interest. When such time intervals are short, it may be sufficient to apply the methods developed on the preceding pages. Frequently, however, time intervals of 4 months or more will be under consideration, and it will therefore be necessary to develop some new method of obtaining theoretical distributions for these extended periods.

One method that immediately suggests itself is to consider the entire period as composed of two or three shorter, overlapping time intervals. Thus, for example, if it is desired to know the number of occurrences of a drought of 30 or more days expected between March 1 and September 30, it would be possible to divide the total time interval into three parts: March 1 to May 31, May 1 to July 31, and July 1 to September 30. By assuming linear variation in drought probability, the number of expected droughts of 30 or more days could be computed for each of these time intervals and the total number of such droughts expected from March 1 to September 30 could then be obtained by addition, after introducing a simple correction to allow for a drought of 30 or more days across two of the time intervals.

Several objections may be raised against this method, however. In considering data from a single station there is a distinct possibility that an assumption of linear variation in drought probability may be unwarranted. It has been shown that when the linear fit is poor, the expected distributions certainly do not adequately represent observed occurrence. Hence precautions should be taken against indiscriminate use of the assumption of linearity. In addition, it will be found that application of the suggested method would have the effect of making every calculation of drought occurrence a completely new problem. Thus, in the example cited above, reduction of the desired drought length to 25 or more days would necessitate a new selection of the component time intervals. New linear fits would have to be obtained, new calculations made, new corrections introduced. Rapid solution of a problem would be impossible, and hence extensive investigation of drought occurrence would be prohibitive. In devising a method of calculating expected drought occurrence for longer time intervals, it is essential that we attempt to avoid these difficulties.

The method followed herein will be to fit an annual drought probability curve, $f(x)$, to the parameters $p_1, p_2, \ldots, p_{12}$ (the monthly drought probabilities). This curve will then be utilized in obtaining the required results. The solution will be graphic rather than algebraic. While a general algebraic solution is possible and has, indeed, been derived, the formulas are too lengthy and unwieldy for practical use, and are not presented here.

Calculation of the Annual Probability Curve

In fitting the annual drought probability curve it will be sufficiently accurate to consider the year as divided into 12 equal months of 30 days each. Using the day as the unit for the abscissa, the problem will be to express drought probability, y, in terms of time, x, where the independent variable, x, varies from 1 to 360 (the number of days in the year). Also, it must be specified that f (360) equals $f(0)$ that is, drought probability must be the same at the end of 1 year and the beginning of the next.

A periodic function will be the obvious choice for $f(x)$. Consider the equation,

$$f(x) = a_0 + \sum_{i=1}^{r}\left(a_i \sin \frac{2\pi}{360} ix + b_i \cos \frac{2\pi}{360} ix\right).$$

If we allow each monthly drought probability to be centered within its month, the constants of this formula, when fitted to the data by least squares, will be

$$a_0 = \frac{1}{12}\sum_{j=1}^{12} p_j,$$

$$a_i = \frac{2}{12}\sum_{j=1}^{12} p_j \sin \frac{2\pi(2j-1)}{24} i,$$

$$b_i = \frac{2}{12}\sum_{j=1}^{12} p_j \cos \frac{2\pi(2j-1)}{24} i.$$

For purposes of analysis, it will be easier to rewrite the annual drought probability curve

$$f(x) = a_0 + \sum_{i=1}^{r} c_i \sin\left(\frac{2\pi}{360} ix + \alpha_i\right),$$

where

$$c_i = \sqrt{a_i^2 + b_i^2},$$

and

$$\alpha_i = \tan^{-1}\frac{b_i}{a_i}.$$

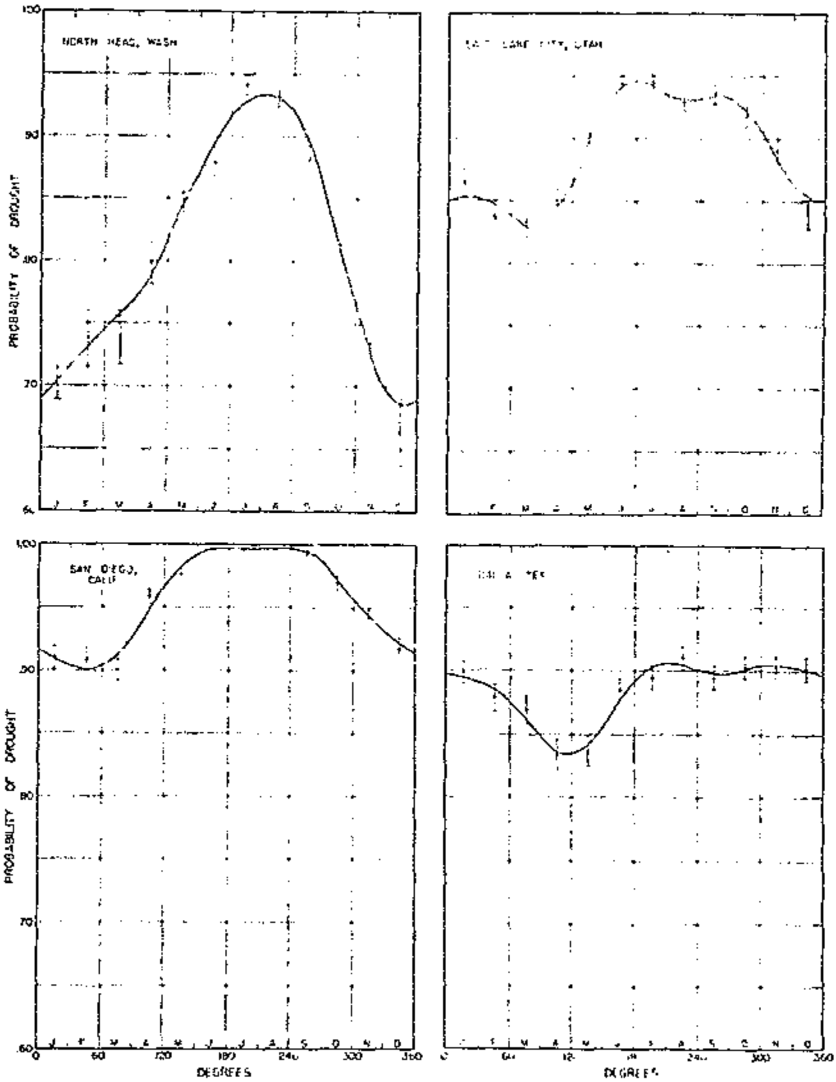


FIGURE 4. Annual drought-probability curves for eight selected stations fitted by least squares to observed monthly probabilities. North Head, Wash., $y=0.808+0.118\sin(x+253°)+0.024\sin(2x+349°)+0.008\sin(3x+358°)$; Salt Lake City, Utah, $y=0.889+0.056\sin(x+228°)+0.009\sin(2x+96°)+0.015\sin(3x+335°)$; San Diego, Calif., $y=0.956+0.050\sin(x+241°)+0.007\sin(2x+181°)+0.006\sin(3x+42°)$; Dallas, Tex., $y=0.884+0.029\sin(x+161°)+0.016\sin(2x+45°)+0.007\sin(3x+255°)$; Le Mars, Iowa, $y=0.879+0.073\sin(x+94°)+0.011\sin(2x+323°)+0.012\sin(3x+214°)$; Lake Placid, N. Y., $y=0.736+0.033\sin(x+230°)+0.023\sin(2x+251°)+0.005\sin(3x+135°)$; Kingston, Ky., $y=0.793+0.046\sin(x+183°)+0.027\sin(2x+230°)+0.017\sin(3x+352°)$; Key West, Fla., $y=0.858+0.084\sin(x+41°)+0.029\sin(2x+173°)+0.019\sin(3x+204°)$.

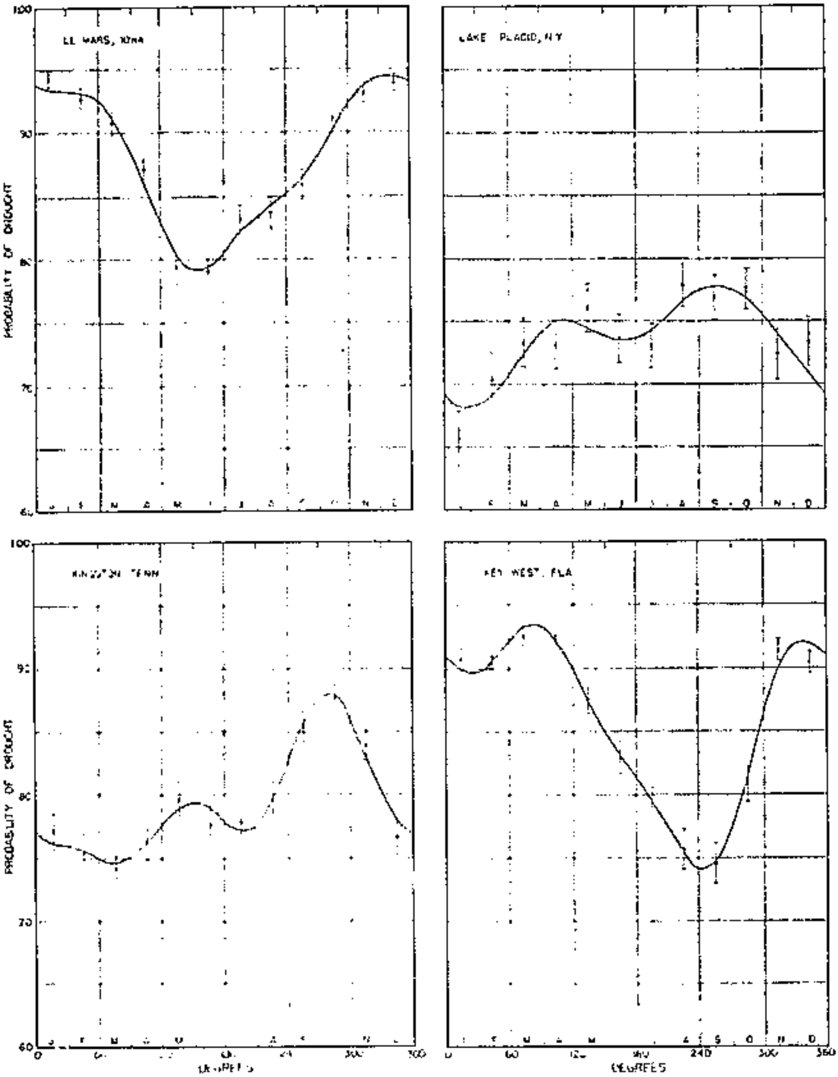


FIGURE 4. Continued
(See legend on opposite page)

Note that in this equation a_0 is the mean drought probability for the year, c_i is the amplitude of the sine wave of i cycles to the year, and α_i is the phase of the sine wave. In evaluating the curve, i will be taken successively as 1, 2, . . ., r, and r will be chosen as large as necessary to obtain adequate fit. In general, taking r as 3 is sufficient. Where $r=6$, the fit becomes exact. Figure 4 shows the annual drought probability curves for eight stations. The monthly probability values are indicated, and the standard errors of these values are shown as vertical lines about the observed points. As can be seen, the fit of the curves is good.

For most purposes curves fitted in this manner will be sufficiently accurate. However, it may sometime be desirable to make more exact use of the parameters $p_1, p_2, \ldots, p_{12}$. Note that these values are actually the mean probabilities for each month. Hence, to be strictly correct, they cannot be considered as points centered within their respective months, but must be interpreted as mean values of the ordinate, $f(x)$, over a given area. To be specific, it must be considered that for $j=1, 2, \ldots, 12$

$$p_j=\frac{1}{30}\int_{30(j-1)}^{30j} f(x)dx.$$

Using these relationships, it can be shown that the constants in the curve

$$f(x)=a_0+\sum_{i=1}^{r}\left(a_i \sin\frac{2\pi}{360}ix+b_i\cos\frac{2\pi}{360}ix\right),$$

may be recalculated as

$$a_0=\frac{1}{12}\sum_{j=1}^{12}p_j$$

$$a_i=\frac{\pi}{12}\frac{i}{6}\frac{1}{\sin\frac{\pi}{12}i}\sum_{j=1}^{12}p_j\sin\left(2\pi\frac{2j-1}{24}i\right),$$

$$b_i=\frac{\pi}{12}\frac{i}{6}\frac{1}{\sin\frac{\pi}{12}i}\sum_{j=1}^{12}p_j\cos\left(2\pi\frac{2j-1}{24}i\right),$$

where, again, the equation is fitted by least squares.

Note that each of the constants, with the exception of a_0, is $[\pi i]/[12 \sin(\pi i/12)]$ times the corresponding constant in the curve fitted to the values p_j centered within their respective months. The new function can, of course, also be written in the form:

$$f(x)=a_0+\sum_{i=1}^{r}c_i\sin\left(\frac{2\pi}{360}ix+\alpha_i\right),$$

where c_i and α_i may be obtained from the new a_i and b_i.

The annual drought-probability curve for Dallas, Tex., has been recalculated in this manner and is shown in figure 5. By taking

r as 6, 12 constants are obtained, and the fit is made exact.[24] Note the close similarity between the curves fitted by the two methods. Valeus of the constants are almost the same for $i=1$, 2, and 3. The more detailed curve for Dallas is used as our sample in the work that follows.

Formula for Expected Occurrence of Drought Over a General Time Interval

Let $f(x)$ be the formula for the annual drought probability. Suppose it is desired to know the number of expected occurrences, (O_N), of a drought of n or more days between the time intervals a and b.

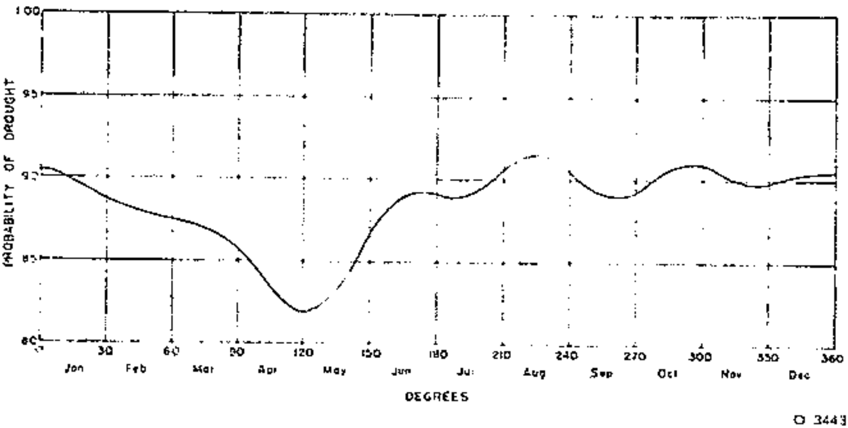


Figure 5. Annual drought-probability curve for Dallas, Tex., fitted by least squares to areas under the curve. $y=0.8842+0.0293\sin(x+161°)+0.0167\sin(2x+45°)+0.0079\sin(3x+255°)+0.0045\sin(4x+149°)+0.0089\sin(5x+36°)+0.0041\sin(6x+180°)$.

Let the ratio M_e/M be considered as a variable function, $F(x)$. This function will represent the chance of any given day being an independent day of record. The value of $F(x)$ can be estimated for any x by inspection of the monthly values of M_e/M (table 11). Then it will be possible to write

$$O_N = MF(a)\prod_{i=0}^{n-1} f(a+i) + M\sum_{x=a+1}^{b-n+1} F(x)[1-f(x-1)]\prod_{i=0}^{n-1} f(x+i).$$

If the time interval a to b is large, little error will be introduced by rewriting

$$O_N = M\int_a^{b-n+1} F(x)[1-f(x)]\prod_{i=0}^{n-1} f(x+i)dx,$$

and if we approximate, as in previous pages,

$$\prod_{i=0}^{n-1} f(x+i) = \left[\frac{1}{n}\int_x^{x+n} f(z)dz\right]^n,$$

[24] Apparently, there are 13 constants in the equation; but note that where i is 6,

$$b_6 = \frac{\pi}{12}\,\frac{1}{\sin\frac{\pi}{2}}\sum_{j=1}^{12} y_j \cos\left[\frac{\pi}{2}(2j-1)\right].$$

Since $\cos[\frac{\pi}{2}(2j-1)]=0$ for all integral values of j, it follows $b_6=0$, and there are actually only 12 constants in the final equation.

the equation for O_N becomes

$$O_N = M\int_a^{b-n+1} F(x)[1-f(x)]\left[\int_x^{x+n} f(z)dz\right]^n dx$$

Let

$$\phi(x) = F(x)[1-f(x)]\left[\int_x^{x+n} f(z)dz\right]^n.$$

Then

$$O_N = M\int_a^{b-n+1} \phi(x)dx$$

But $\phi(x)$ can be graphically determined with a high degree of accuracy by obtaining the value of the function for a number of values of x and connecting the points with a smooth curve. The value of O_N for M years of observation will then be determined as M times the area under $\phi(x)$ between the limits $x=a$ and $x=b-n+1$.

A family of curves can be prepared in this manner for various values of n and can be presented graphically. Figure 6 shows the curves $\phi(x)$ for Dallas, Tex., for n equal to 1, 5, 10, 15, 20, 25, 30, 35, and 40. These curves can be conveniently used in obtaining expected distributions for any time interval. The area under the curve can be measured with sufficient accuracy by any number of methods of approximation. This graphic method of evaluating O_N is subject to estimation errors. However, if the work is conducted with reasonable care, such errors can be limited to a maximum of 1 or 2 percent for the higher values of n. This is less than the likely sampling error and hence will be ignored.

Estimating the Standard Error of O_N

For shorter drought lengths, where the average probability of occurrence of a drought of n or more days is greater than 0.03, the standard error of O_n can be estimated in the same manner as in previous sections. Where p is the mean probability of a single day of drought and M_e is the mean number of independent years of record, we may estimate

$$(T_s)_n = M_e + M_e q(N-n),$$

and where

$$p_s = \frac{O_n}{(T_s)_n}$$

$$\text{S. E. of } O_n = \sqrt{O_n(1-p_s)}.$$

When p_s is less than 0.03, the distribution of the sampling error will be considered Poisson.

The Annual Expected Drought Distribution

Estimating the areas under the curves in figure 6 enables us to set up a cumulative distribution for drought occurrence at Dallas, Tex., for an annual time interval.[25] By taking first differences, a noncumulative grouping may be obtained. Droughts of 1 to 19 days will again be grouped together. Table 9 gives the theoretical dis-

[25] The areas were estimated by averaging the values of $\phi(x)$ for 15 degree intervals on the axis of abscissas. The 21 values thus summed were those used in plotting the graphs. The mean value of the 21 ordinates was taken as the mean ordinate for the entire curve and was multiplied by 365 (the number of days in the year) and then by 39.9 (the average number of years of record) to obtain O_n for each value of n.

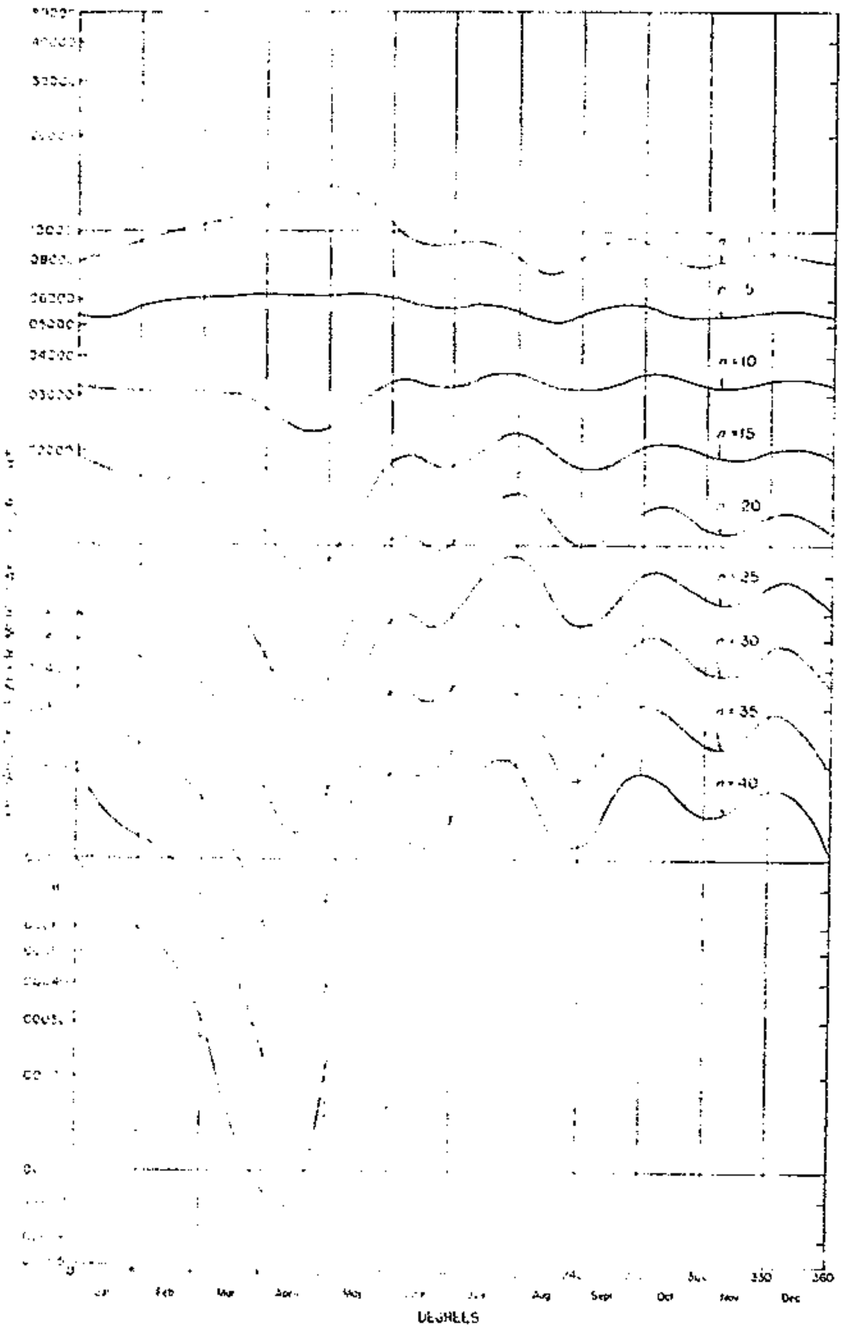


FIGURE 6. Curves for Dallas, Tex., showing the probability of each day of the year being a day of precipitation followed by n or more days of drought.

tribution thus obtained and the observed distribution for approximately 40 years of record.

TABLE 9.—*Distribution of the observed and the expected occurrence of annual drought for Dallas, Tex., 1898 to 1937*

Length of drought (days)	Expected distribution	Observed distribution	Length of drought (days)	Expected distribution	Observed distribution
1–19	1,251.27	1,256	30–34	19.02	23
20–24	63.88	59	35–39	10.89	6
25–29	34.01	30	40 and over	17.18	22

Apparently, the hypothesis of random occurrence of drought is adequate. The difference between the two distributions is great in only the two groups 35–39 and 40 and over. It might seem from a cursory examination of table 9 that there may be a significant excess of very long droughts in the observed distribution. Further investigation reveals, however, that this is not true. Actually, in the observed distribution, 12 of the 22 droughts of 40 or more days occur in the 40–44 group. The number of expected and observed droughts of 45 or more days would, therefore, compare closely. Apparently, the one major difference between the two distributions is that in the 40–44 day interval 4 or 5 more droughts were observed than were expected and in the 35–39 interval 4 or 5 less. It would be difficult to believe that this could be caused by some meteorologic factor rather than by sampling error. In addition, testing the two distributions by the χ^2 test yields a χ^2 of 6.68 for 5 degrees of freedom. The probability of obtaining a deviation of this magnitude by chance is 0.245. It may be said, then, that the theoretical distribution calculated from the hypothesis of random occurrence adequately represents the observed data.

A similar distribution could be worked out from the same set of curves for any time interval whatsoever. Hence the family of curves is a clear and concise way of presenting distribution of drought. It is not difficult to prepare graphs of this character, and in the future it will be possible to present in similar form probability data for a number of stations throughout the United States. As will be shown in the following section, these graphs may be utilized advantageously in obtaining likelihood of future drought as well as in describing past occurrences. Hence they will be of value in estimating drought hazard.

CALCULATION OF PROBABILITY OF FUTURE DROUGHT OCCURRENCE

In preceding sections theoretical drought distributions have been presented that may be considered representative of observed drought data for the years 1898 to 1937. The accuracy with which the techniques developed may be used in estimating future likelihood of drought will now be considered. It must be remembered that any theorem based upon an observed sample, no matter how large, cannot be used with absolute certainty as a measure of future probability. Yet it is consistent with the scientific approach to assume that such a theorem is correct until future events prove it otherwise. In the analysis of drought data once the hypothesis of random occurrence of long droughts has been granted, any future drought probability can be readily computed.

To calculate drought probabilities at a given station for any time interval, it is necessary to know only 24 constants: The 12 monthly drought probabilities and the 12 values that indicate the persistence of rainfall. The latter group may be evaluated as $(M_e/M)_j$ —that is, the ratio of the number of independent years of record to the actual number of years of record for each of 12 months ($j=1, 2, \ldots, 12$). This ratio represents the mean chance that a day within the specified month will be an independent day of record.

In calculating drought probabilities, it will be necessary to consider the effect of error in the parameters p_j and $(M_e/M)_j$. It is not possible to determine the true values of these parameters from a limited sample. However, the most likely values can be determined, as well as the standard errors of these values. It has been shown on preceding pages that

$$p=\frac{D-\sqrt{D^2-4M(D-S)}}{2M}\pm\sqrt{\frac{p(1-p)}{D}},$$

and it can also be demonstrated that

$$\frac{M_e}{M}=\frac{D}{pT}\pm\frac{M_e}{M}\sqrt{\frac{1-p}{pD}}.$$

The effect of considering these standard errors will be to introduce an uncertainty about the exactness of the calculated values of O_N (expected number of occurrences of a drought of n or more days within any defined time interval). It will be correct to write the expected number of occurrences as $O_N\pm\sigma_e$, where σ_e is the standard error of O_N and varies as the magnitude of the standard errors of p_j and $(M_e/M)_j$.

When O_N is small (as it will usually be, since we will be primarily interested in the likelihood of long droughts) the probability of observing 0, 1, 2, . . . occurrences of a drought of n or more days approximates a Poisson distribution. The more accurately the true O_N can be determined, [that is, the smaller the error factor in the parameters p_j and $(M_e/M)_j$] the closer the approximation will become. If $\sigma_e=0$, the expected value of O_N would exactly equal the true value, m, and the probability of observing 0, 1, 2, . . ., r, . . . occurrences of a drought of n or more days would come very close to being the successive terms of the series

$$e^{-m},\ me^{-m},\ \frac{m^2}{2!}e^{-m},\ \ldots,\ \frac{m^r}{r!}e^{-m},\ \ldots$$

Where $\sigma_e\neq 0$, it will be desirable to determine whether substitution of O_N for m in this series will yield the best measure of the successive terms.

Assuming that observed O_N will be normally distributed about the true value, m, with standard deviation of σ_e, it follows that

$$\text{Mean value of } \frac{O_N^r e^{-O_N}}{r!} = \int_{-\infty}^{+\infty} \frac{O_N^r e^{-O_N}}{r!} \frac{1}{\sigma_\epsilon\sqrt{2\pi}} e^{-\frac{(O_N-m)^2}{2\sigma_\epsilon^2}} d(O_N).$$

Integrating by infinite series and developing, this reduces to

$$\text{Mean value of } \frac{O_N^r e^{-O_N}}{r!} = \frac{e^{-\left(m-\frac{\sigma_\epsilon^2}{2}\right)}}{r!}\Big[(m-\sigma_\epsilon^2)^r + \frac{r(r-1)}{2}\sigma_\epsilon^2(m-\sigma_\epsilon^2)^{r-2} + \frac{r(r-1)(r-2)(r-3)}{2\cdot 4}\sigma_\epsilon^4(m-\sigma_\epsilon^2)^{r-4} + \ldots\Big],$$

where $r!=1$ for $r=0$ and the series within the brackets continues for $(r+1)/2$ terms where r is odd, and for $(r+2)/2$ terms where r is even. Thus:

$$\text{Mean value of } e^{-O_N} = e^{-\left(m-\frac{\sigma_\epsilon^2}{2}\right)};$$

$$\text{Mean value of } O_N e^{-O_N} = e^{-\left(m-\frac{\sigma_\epsilon^2}{2}\right)}(m-\sigma_\epsilon^2);$$

$$\text{Mean value of } \frac{O_N^2 e^{-O_N}}{2!} = \frac{e^{-\left(m-\frac{\sigma_\epsilon^2}{2}\right)}}{2!}[(m-\sigma_\epsilon^2)^2-\sigma_\epsilon^2].$$

The terms continue in a similar manner. It is obvious from these expressions that an adjustment should be made in the value O_N in evaluating the terms of the Poisson series. It can be readily shown from these relationships that the best measures of the first three terms will be:

$$e^{-m} = e^{-\left(O_N+\frac{\sigma_\epsilon^2}{2}\right)};$$

$$me^{-m} = (O_N+\sigma_\epsilon^2)e^{-\left(O_N+\frac{\sigma_\epsilon^2}{2}\right)};$$

$$\frac{m^2e^{-m}}{2!} = [(O_N+\sigma_\epsilon^2)^2+\sigma_\epsilon^2]\frac{e^{-\left(O_N+\frac{\sigma_\epsilon^2}{2}\right)}}{2!}, \text{ approximately.}$$

In deriving these formulas it has been implicitly assumed that a large number of observed occurrences of a drought of n or more days is possible. Thus, if 0.20 occurrences of a drought of 40 days are to be expected between June 1 and September 15, it has been assumed that actually 0, 1, 2, 3, . . . such droughts could occur. In reality, it is impossible for more than two 40-day droughts to fall within this time interval, and, in addition, the occurrence of a single drought of more than 40 days will greatly prejudice the chance of a second drought being observed. Hence it is certainly not strictly correct to assume the distribution to be Poisson. It is likely that a distribution similar to one developed by Greenwood and Yule would apply here (*8*). However, these distributions approach the same limit as O_N approaches 0. Usually primary interest is centered in investigating probable occurrence of long droughts, and hence O_N for any one year will usually be considerably less than 1.00. The error in considering the series to be Poisson will be negligible in evaluating the probabilities for $r=0$ and $r=1$. Error in the probabilities for $r\geqq 2$ will be considerable when expressed as a percentage, but the actual magnitude will be small. Thus, in the example of the 40-day drought, where $O_N=0.20$, the chance of observing two or more occurrences in a Poisson distribution is less than 0.02. Evaluated more accurately, this figure would be nearer 0.01, but it would definitely not be worth while to introduce additional complexities to correct for such a small difference.

It is now possible to calculate sample probabilities. The values and standard errors of the parameters p_j and $(M_e/M)_j$ for 40 selected stations are shown in tables 10 and 11 respectively. These tables are referred to in the examples below. Three examples have been chosen: (1) For a short time interval, wherein probability of drought has been considered constant; (2) for a slightly longer time interval, wherein probability of drought has been considered to vary linearly; and (3) for an annual time interval, wherein probability of drought has been considered to vary curvilinearly.

Drought Probabilities for Short Time Intervals

When the time invertal under consideration is short, drought probability may be considered constant within the period, and for one calendar year the following formulas can be used:

$$O_X=\frac{M_e}{M}p^n[(N-n+1)-(N-n)p],$$

$$\sigma_e=\sqrt{\left[\sigma_p\frac{\partial O_X}{\partial p}\right]^2+\left[\sigma_{M_e/M}\frac{\partial O_X}{\partial (M_e/M)}\right]^2},$$

where, as before, σ_e is the standard error of O_X, resulting from errors in the parameters; where σ_p is the standard error of the constant drought probability, and $\sigma_{M_e/M}$ is the standard error of the ratio M_e/M. The values of O_X and σ_e can be used to determine the probability of 0, 1, . . . occurrences of a drought of the required length.

Let us consider the problem of drought hazard with respect to the growing of spring wheat in North Dakota. Davis and Pallesen have shown that for this crop adequate rainfall is most necessary during the 30 to 40 days preceding July 10, the average date of heading (*5*). A drought of 20 days at this time would certainly have a detrimental effect on the harvest, particularly since we have defined drought as being terminated by a minimum of 0.10 inch of precipitation within 48 hours. It is desired, therefore, to determine the probability of the occurrence of a drought of 20 days or more between June 1 and July 10. Bottineau, N. Dak., has been selected as a representative station.

Referring to table 10, it can be seen that the probability of drought for June is 0.7861 ± 0.0147, and for July it is 0.8282 ± 0.0130. By weighting these values and combining, the constant probability for the time interval June 1 to July 10 is determined to be 0.7966 ± 0.0165. The standard error over the combined interval is only slightly larger than that for either month, since we do not know whether the signs of the monthly standard errors agree. Similarly, referring to table 11, the ratio (M_e/M) for June is found to be 0.9008 ± 0.0167 and for July, 0.9356 ± 0.0148. By interpolating, and combining weighted standard errors, the value for the time interval June 1 to July 10 is determined to be 0.9095 ± 0.0187.

TABLE 10.—*Mean monthly probability of a day of drought and associated standard errors for 40 selected stations*

Station	January	February	March	April	May	June	July	August	September	October	November	December
Ann Arbor, Mich.:												
Probability	0.8303	0.8297	0.8138	0.7896	0.7911	0.8111	0.8105	0.8371	0.8135	0.8338	0.8314	0.8105
Standard error	.0121	.0120	.0131	.0147	.0140	.0135	.0118	.0123	.0133	.0122	.0125	.0136
Billings, Mont.:												
Probability	.9336	.9563	[illegible]	.8571	.8367	.8750	.9020	.9171	.8071	.9111	.9263	.9368
Standard error	.0080	.0068	.0093	.0105	.0133	.0125	.0098	.0091	.0105	.0096	.0091	.0078
Bisbee, Ariz.:												
Probability	.9553	.9131	.9115	.9706	.9831	.9631	.7761	.7628	.9088	.9592	.9385	.9269
Standard error	.0083	.0101	.0075	.0076	.0040	.0063	.0167	.0166	.0101	.0065	.0083	.0089
Boise, Idaho:												
Probability	.8468	.8605	.8697	.8691	.9206	.9471	.9800	.9779	.9546	.9311	.8921	.8679
Standard error	.0121	.0125	.0110	.0114	.0085	.0077	.0040	.0045	.0064	.0077	.0116	.0114
Bottineau, N. Dak.:												
Probability	.9502	.9593	.9388	.9267	.8776	.7861	.8282	.8600	.8881	.9232	.9380	.9400
Standard error	.0067	.0062	.0070	.0080	.0101	.0147	.0130	.0118	.0110	.0085	.0077	.0073
Brownsville, Tex.:												
Probability	.9131	.9376	.9461	.9362	.9006	.8653	.9219	.9264	.8250	.9035	.9072	.9024
Standard error	.0101	.0089	.0078	.0084	.0105	[illegible]	.0093	.0097	.0150	.0104	.0103	.0106
Chico, Calif.:												
Probability	.8447	.8402	.8813	.9177	.9423	.9720	.9962	.9902	.9785	.9189	.8844	.8559
Standard error	.0128	.0133	.0101	.0087	.0077	.0048	.0009	.0009	.0043	.0065	.0102	.0117
Colorado Springs, Colo.:												
Probability	.9775	.9619	.9148	.9090	.8696	.8756	.8737	.8489	.9293	.9448	.9657	.9776
Standard error	.0045	.0060	.0073	.0083	.0117	.0109	.0116	.0128	.0086	.0073	.0054	.0043
Dalhart, Tex.:												
Probability	.9787	.9782	.9450	.9127	.8796	.8649	.8799	.8672	.9169	.9425	.9588	.9598
Standard error	.0048	.0071	.0071	.0109	.0117	.0126	.0112	.0123	.0101	.0098	.0073	.0098
Dallas, Tex.:												
Probability	.8985	.8791	.8698	.8337	.8564	.8847	.8911	.9101	.8950	.9026	.9031	.9019
Standard error	.0094	.0108	.0107	.0126	.0122	.0102	.0096	.0089	.0096	.0080	.0092	.0092
Duluth, Minn.:												
Probability	.9078	.9120	.8662	.8652	.8178	.7836	.7961	.8142	.8136	.8612	.8956	.9080
Standard error	.0090	.0093	.0108	.0112	.0131	.0140	.0134	.0128	.0133	.0112	.0089	.0092
Escanaba, Mich.:												
Probability	.8609	.8744	.8426	.8117	.8030	.7980	.7814	.8143	.8053	.8403	.8558	.8677
Standard error	.0112	.0111	.0118	.0124	.0130	.0138	.0131	.0128	.0137	.0121	.0115	.0108
Eureka, Calif.:												
Probability	.7868	.7761	.8138	.8430	.8942	.9182	.9629	.9891	.9413	.9076	.8290	.7949
Standard error	.0165	.0172	.0117	.0126	.0099	.0069	.0022	.0030	.0072	.0093	.0138	.0158
Franklin, La.:												
Probability	.8309	.8218	.8461	.8700	.8651	.7911	.6619	.6855	.7911	.8873	.8721	.7833
Standard error	.0121	.0132	.0122	.0107	.0113	.0142	.0189	.0184	.0146	.0102	.0108	.0141
Fresno, Calif.:												
Probability	.8939	.8898	.8851	.9513	.9641	.9933	.9991	.9992	.9925	.9709	.9159	.9092
Standard error	.0097	.0105	.0102	.0065	.0050	.0022	.0012	.0007	.0027	.0050	.0097	.0089
Greensboro, Ala.:												
Probability	.8050	.7653	.7998	.8198	.8410	.8284	.7727	.8075	.8838	.8936	.8563	.7729
Standard error	.0129	.0146	.0129	.0126	.0122	.0124	.0142	.0129	.0101	.0094	.0112	.0137
Hatteras, N. C.:												
Probability	.7912	.7615	.7811	.8074	.8211	.8217	.7888	.7939	.8153	.8658	.8555	.7871
Standard error	.0143	.0163	.0141	.0132	.0127	.0129	.0144	.0143	.0119	.0112	.0117	.0148
Hays City, Kans.:												
Probability	.9712	.9281	.9119	.8817	.8213	.8211	.8680	.8363	.8997	.9313	.9390	.9621
Standard error	.0053	.0088	.0095	.0109	.0136	.0139	.0111	.0128	.0100	.0081	.0080	.0060
Independence, Calif.:												
Probability	.9770	.9541	.9644	.9867	.9857	.9919	.9904	.9802	.9883	.9778	.9605	.9631
Standard error	.0052	.0068	.0066	.0033	.0037	.0027	.0030	.0042	.0033	.0043	.0040	.0057
Key West, Fla.:												
Probability	.9075	.9400	.9261	.9267	.8761	.8400	.7937	.7573	.7153	.8080	.9146	.9049
Standard error	.0092	.0092	.0080	.0079	.0136	.0130	.0140	.0153	.0168	.0140	.0088	.0090
Kingston, Tenn.:												
Probability	.7718	.7546	.7121	.7627	.7903	.7759	.7787	.7967	.8537	.8765	.8397	.7660
Standard error	.0145	.0157	.0153	.0146	.0136	.0142	.0141	.0156	.0141	.0103	.0120	.0146
Lake Placid, N. Y.:												
Probability	.6574	.7949	.7390	.7311	.7663	.7966	.7303	.7791	.7689	.7765	.7240	.7310
Standard error	.0219	.0214	.0191	.0192	.0182	.0192	.0184	.0170	.0171	.0168	.0201	.0194
Le Mars, Iowa:												
Probability	.9438	.9270	.9080	.8718	.7912	.7901	.8303	.8462	.8588	.9114	.9306	.9391
Standard error	.0067	.0082	.0089	.0111	.0148	.0147	.0120	.0123	.0117	.0090	.0080	.0070
Lexington, Ky.:												
Probability	.7657	.7716	.7567	.7678	.8077	.7826	.8119	.8039	.8175	.8584	.8179	.7848
Standard error	.0154	.0152	.0157	.0151	.0135	.0144	.0131	.0131	.0121	.0112	.0130	.0144
Marengo, Ill.:												
Probability	.8892	.8718	.8531	.8289	.8084	.7792	.8116	.8400	.8193	.8513	.8697	.8784
Standard error	.0101	.0113	.0116	.0125	.0134	.0147	.0118	.0090	.0127	.0116	.0110	.0101
Muskogee, Okla.:												
Probability	.8877	.8886	.8503	.8019	.7965	.8195	.8823	.8766	.8771	.8710	.8866	.9053
Standard error	.0101	.0105	.0121	.0131	.0156	.0121	.0163	.0106	.0108	.0108	.0102	.0096

TABLE 10. *Mean monthly probability of a day of drought and associated standard errors for 40 selected stations*—Continued

Station	January	February	March	April	May	June	July	August	September	October	November	December
New Haven, Conn.												
Probability	0.7565	0.7718	0.7623	0.7756	0.8060	0.7922	0.7798	0.8078	0.8293	0.8429	0.8148	0.7084
Standard error	.0150	.0152	.0148	.0148	.0134	.0138	.0141	.0132	.0121	.0118	.0133	.0144
North Head, Wash.												
Probability	.7143	.7397	.7306	.7987	.8547	.8792	.9110	.9324	.8807	.8129	.7311	.6645
Standard error	.0216	.0226	.0214	.0173	.0156	.0125	.0078	.0084	.0120	.0160	.0215	.0273
North Platte, Nebr.												
Probability	.9658	.9478	.9307	.8857	.8587	.8273	.8412	.8532	.9177	.9372	.9597	.9590
Standard error	.0056	.0077	.0077	.0106	.0123	.0127	.0118	.0116	.0085	.0074	.0059	.0062
Omaha, Nebr.												
Probability	.9531	.9160	.9048	.8501	.7968	.8159	.8529	.8154	.8128	.9046	.9243	.9358
Standard error	.0075	.0096	.0092	.0127	.0136	.0131	.0129	.0118	.0121	.0092	.0082	.0071
Pierre, S. Dak.												
Probability	.9608	.9429	.9336	.8702	.8542	.8271	.8766	.8820	.9132	.9317	.9502	.9472
Standard error	.0060	.0077	.0077	.0110	.0117	.0126	.0114	.0104	.0088	.0077	.0062	.0068
Portland, Maine												
Probability	.7390	.7462	.7672	.7868	.8219	.7757	.8062	.8142	.8175	.8029	.8025	.7683
Standard error	.0137	.0160	.0146	.0142	.0127	.0142	.0131	.0128	.0131	.0132	.0136	.0144
Port Oxford, Oreg.												
Probability	.7476	.7317	.7729	.8478	.8742	.9148	[illegible]	.9726	.8817	.8807	.8120	.7179
Standard error	.0261	.0252	.0260	.0176	.0151	.0117	.0062	.0062	.0148	.0146	.0210	.0276
Rockingham, N. C.												
Probability	.8266	[illegible]	.8056	.8215	.8257	.7850	.7727	.7472	.8640	.8920	.8927	.8248
Standard error	.0132	[illegible]	.0136	.0136	.0126	.0136	.0154	.0161	.0115	.0097	.0097	.0132
St. Augustine, Fla.												
Probability	.8586	[illegible]	.8565	.8870	.8806	.8468	.7784	.7820	.8148	.8142	.9058	.8460
Standard error	.0117	.0129	.0124	.0106	.0101	[illegible]	.0154	.0144	.0143	.0124	.0094	.0148
Salt Lake City, Utah												
Probability	.8617	.8370	.8550	.8171	.8452	.9428	.9148	.9289	.9352	.9176	.8896	.8895
Standard error	.0112	.0128	[illegible]	.0121	.0095	.0072	.0070	.0078	.0078	.0087	.0100	.0116
San Diego, Calif.												
Probability	.9404	.9087	.9004	[illegible]	.9764	[illegible]	.9966	.9966	.9937	.9698	.9491	.9177
Standard error	.0085	.0096	.0094	.0064	.0044	.0016	.0020	.0016	.0024	.0050	.0067	.0087
Tucumcari, N. Mex.												
Probability	.9758	.9655	[illegible]	[illegible]	.9164	[illegible]	[illegible]	[illegible]	.9498	.9379	[illegible]	.9588
Standard error	.0050	.0062	.0062	[illegible]	.0094	.0110	[illegible]	.0134	.0094	.0081	.0068	.0071
Winnemucca, Nev.												
Probability	.9851	.8057	[illegible]	.9282	[illegible]	[illegible]	.9814	.9822	[illegible]	[illegible]	.9257	.9148
Standard error	.0062	.0100	.0083	.0080	.0072	.0064	[illegible]	[illegible]	.0076	.0062	.0082	.0087
Yuma, Ariz.												
Probability	.9818	.9686	.9747	.9828	.9984	.9993	.9841	.9880	.9785	.9860	.9804	.9703
Standard error	.0056	.0054	.0046	[illegible]	.0032	[illegible]	.0036	.0051	.0035	.0033	.0040	.0051

TABLE 11. *Mean monthly measure of persistence of rainfall (M, M) and associated standard errors for 40 selected stations*

Station	January	February	March	April	May	June	July	August	September	October	November	December
Ann Arbor, Mich.												
Persistence	0.8975	0.9257	0.8840	0.8421	0.8570	0.9074	0.9248	0.9276	0.9023	0.9218	0.8943	0.8063
Standard error	.0135	.0130	.0140	[illegible]	.0150	[illegible]	.0129	.0136	.0148	.0137	.0135	.0132
Billings, Mont.												
Persistence	.9350	.9774	.9572	.9312	.8755	.8790	.9529	.9403	.9404	.9567	.9231	.9041
Standard error	.0081	.0074	.0098	.0106	.0138	.0128	[illegible]	.0096	.0108	.0099	.0091	.0051
Bisbee, Ariz.												
Persistence	.9197	.9172	.9500	.9797	[illegible]	.9559	.7870	.8418	.9247	.9666	.9587	.9119
Standard error	.0086	.0107	.0076	.0057	.0059	.0062	.0168	.0187	.0103	.0067	.0084	.0089
Boise, Idaho												
Persistence	.8257	.8556	.8874	.9062	.9272	.9511	.9844	.9872	.9552	.9468	.8600	.8703
Standard error	.0139	.0123	.0124	.0114	.0085	.0075	.0042	.0042	.0062	.0077	.0110	.0115
Bottineau, N. Dak.												
Persistence	.9526	.9887	.9795	.9533	.9168	.9008	.9356	.9346	.9115	.9522	.9614	.9648
Standard error	.0068	.0064	.0075	.0082	.0108	.0267	.0248	.0274	.0112	.0086	.0081	.0075
Brownsville, Tex.												
Persistence	.9100	.9500	.9517	.9174	.9303	.9633	.9700	.9368	.8265	.9042	.9597	.8948
Standard error	.0098	.0092	.0076	.0087	.0104	.0115	.0093	.0105	.0153	.0109	.0111	.0109
Chico, Calif.												
Persistence	.8264	.8282	.8785	.9518	.9568	.9833	.9975	.9977	.9843	.9535	.9199	.8679
Standard error	.0128	.0128	.0103	.0084	.0077	.0050	.0007	.0007	.0042	.0065	.0107	.0119
Colorado Springs, Colo.												
Persistence	.9857	.9671	.9285	.9187	.8755	.9166	.8898	.8957	.9097	.9680	.9760	.9779
Standard error	.0045	.0059	.0076	.0095	.0118	.0112	.0120	.0131	.0086	.0075	.0053	.0042

TABLE 11. — *Mean monthly measure of persistence of rainfall (M_c/M) and associated standard errors for 40 selected stations* — Continued

Station	January	February	March	April	May	June	July	August	September	October	November	December
Dalhart, Tex.:												
Persistence	0.9933	0.9950	0.9670	0.9939	0.9113	0.9184	0.9355	0.9097	0.9150	0.9076	0.9737	0.9596
Standard error	.0049	.0070	.0071	.0110	.0122	.0134	.0117	.0129	.0098	.0095	.0073	.0100
Dallas, Tex.:												
Persistence	.9265	.9433	.9685	.9508	.9015	.9288	.9475	.9383	.9488	.9415	.9373	.9335
Standard error	.0099	.0115	.0118	.0138	.0130	.0107	.0102	.0090	.0102	.0094	.0098	.0094
Duluth, Minn.:												
Persistence	.9160	.9308	.9255	.9003	.8020	.8480	.8908	.9033	.8695	.8851	.8800	.8667
Standard error	.0090	.0096	.0116	.0116	.0139	.0159	.0149	.0131	.0131	.0117	.0099	.0080
Escanaba, Mich.:												
Persistence	.9048	.9058	.9160	.8672	.8728	.8048	.8855	.9293	.8828	.8893	.8950	.8988
Standard error	.0117	.0116	.0129	.0128	.0146	.0141	.0150	.0145	.0149	.0129	.0120	.0111
Eureka, Calif.:												
Persistence	.6263	.6778	.7105	.8108	.8095	.9310	.9680	.9938	.9138	.8565	.7380	.6615
Standard error	.0131	.0151	.0128	.0121	.0099	.0069	.0022	.0030	.0073	.0080	.0122	.0127
Franklin, La.:												
Persistence	.9128	.9285	[illegible]	.9774	.9026	.8723	.7787	.7803	.8820	.9174	.9318	.9018
Standard error	.0138	.0147	.0113	.0118	.0123	.0155	.0220	.0211	.0153	.0104	.0114	.0162
Fresno, Calif.:												
Persistence	.9103	.9192	.8870	.9510	.9753	.9987	.9992	.9997	.9565	.9715	.9856	.9120
Standard error	.0086	.0101	.0090	.0061	.0050	.0022	.0012	.0007	.0027	.0050	.0067	.0082
Greensboro, Ala.:												
Persistence	.9405	.9860	.9730	.9725	.9287	.9237	.9003	.9390	.9178	.9585	.9680	.9840
Standard error	.0151	.0186	.0158	.0153	.0138	.0138	.0169	.0138	.0146	.0162	.0127	.0174
Hatteras, N. C.:												
Persistence	.9081	.8811	.8882	.9103	.9500	.8865	.8526	.8315	.8950	.8885	.9106	.8619
Standard error	.0167	.0183	.0163	.0149	.0142	.0139	.0151	.0150	.0129	.0115	.0127	.0161
Hays City, Kans.:												
Persistence	1.0491	.9637	.8876	.9230	.8806	.9015	.9565	.9257	.9231	.9090	.9682	.9771
Standard error	.0053	.0092	.0066	.0114	.0115	.0151	.0119	.0111	.0103	.0083	.0084	.0062
Independence, Calif.:												
Persistence	.9565	.9578	.9851	.9887	.9756	.9914	.9922	.9916	.9868	.9838	1.0154	.9775
Standard error	.0062	.0068	.0057	.0035	.0030	.0027	.0028	.0030	.0033	.0042	.0042	.0057
Key West, Fla.:												
Persistence	.9108	.9690	.9714	[illegible]	[illegible]	.8608	.8813	.8348	.7500	.8926	.9230	.9395
Standard error	.0095	.0069	.0079	.0085	[illegible]	.0133	.0157	.0170	.0168	.0140	.0080	.0084
Kingston, Tenn.:												
Persistence	.8798	.9013	.8845	.9288	.8830	.8955	.9020	.8890	.9380	.9258	.9210	.8843
Standard error	.0165	.0188	.0186	.0178	.0156	.0168	.0164	.0158	.0125	.0111	.0138	.0169
Lake Placid, N. Y.:												
Persistence	.8501	.8207	.8808	.8661	[illegible]	.8867	.9186	.9215	.9071	.9036	.8270	.8261
Standard error	.0285	.0273	.0231	.0229	[illegible]	.0230	.0229	.0260	.0205	.0193	.0214	.0216
Le Mars, Iowa:												
Persistence	1.0060	.9779	.9682	.9037	.8977	.8858	.9170	.9365	.8763	.9100	.9518	.9745
Standard error	.0073	.0099	.0096	.0122	.0153	.0162	.0138	.0139	.0120	.0095	.0081	.0073
Lexington, Ky.:												
Persistence	.8293	.8788	.7975	.8778	.8458	.8615	.8740	.8809	.8923	.8980	.8935	.8375
Standard error	.0166	.0174	.0167	.0160	.0140	.0154	.0146	.0149	.0127	.0117	.0126	.0156
Marengo, Ill.:												
Persistence	.9056	.9285	.9152	.8928	.8916	.9021	.9297	.9513	.9185	.9133	.9317	.9521
Standard error	.0113	.0117	.0121	.0138	.0148	.0169	.0129	.0106	.0144	.0124	.0118	.0111
Muskogee, Okla.:												
Persistence	.9284	.9592	.9147	.9311	.9149	.8882	.9277	.9250	.9123	.9087	.9380	.9069
Standard error	.0105	.0112	.0130	.0157	.0158	.0127	.0169	.0132	.0114	.0114	.0100	.0092
New Haven, Conn.:												
Persistence	.8805	.8863	.8863	.8520	.8710	.9003	.8815	.8875	.9083	.9105	.8695	.9030
Standard error	.0156	.0171	.0170	.0165	.0145	.0158	.0160	.0145	.0138	.0129	.0145	.0169
North Head, Wash.:												
Persistence	.4417	.5368	.5400	.6726	.7137	.7215	.9214	.9138	.7984	.6956	.5059	.4256
Standard error	.0151	.0153	.0157	.0146	.0150	.0111	.0076	.0084	.0111	.0136	.0157	.0176
North Platte, Nebr.:												
Persistence	.9698	.9617	.9393	.8903	.8685	.8865	.9098	.9085	.9418	.9570	.9705	.9515
Standard error	.0050	.0077	.0077	.0107	.0126	.0176	.0129	.0124	.0090	.0074	.0061	.0063
Omaha, Nebr.:												
Persistence	.9428	.9435	.9258	.9008	.8795	.8897	.9403	.9187	.8970	.9183	.9395	.9523
Standard error	.0077	.0093	.0094	.0125	.0150	.0144	.0135	.0127	.0129	.0094	.0085	.0077
Pierre, S. Dak.:												
Persistence	.9212	.9639	.9479	.9195	.8938	.8910	.9185	.9115	.9123	.9515	.9518	.9663
Standard error	.0084	.0076	.0080	.0113	.0125	.0139	.0125	.0105	.0087	.0078	.0062	.0069
Portland, Maine:												
Persistence	.9121	.9140	.8900	.8708	.8810	.8855	.8898	.9093	.8948	.9030	.8898	.8985
Standard error	.0170	.0194	.0166	.0150	.0135	.0169	.0149	.0145	.0144	.0162	.0162	.0169
Port Oxford, Oreg.:												
Persistence	.5501	.7701	.6677	.7211	.7873	.8718	.9808	.9758	.8865	.7576	.6557	.5414
Standard error	.0198	.0257	.0175	.0159	.0111	.0112	.0060	.0060	.0150	.0134	.0172	.0206
Rockingham, N. C.:												
Persistence	.9200	.9511	.9385	.9213	.9122	.8939	.8871	.9091	.9227	.9581	.9546	.9251
Standard error	.0145	.0187	.0152	.0154	.0145	.0167	.0179	.0193	.0124	.0104	.0105	.0149

TABLE 11.—*Mean monthly measure of persistence of rainfall (M_r:M) and associated standard errors for 40 selected stations*—Continued

Station	January	February	March	April	May	June	July	August	September	October	November	December
St. Augustine, Fla.:												
Persistence	0.9257	0.9218	0.9161	0.9295	0.8808	0.8073	0.8065	0.8641	0.8095	0.8493	0.9582	0.9600
Standard error	.0125	.0141	.0122	.0116	.0104	.0132	.0160	.0159	.0143	.0121	.0098	.0134
Salt Lake City, Utah:												
Persistence	.8887	.9066	.8859	.8910	.8880	.9410	.9605	.9138	.9379	.9940	.9105	.9445
Standard error	.0115	.0138	.0133	.0127	.0097	.0070	.0070	.0078	.0077	.0087	.0102	.0134
San Diego, Calif.:												
Persistence	.8885	.8608	.9095	.9495	.9770	.9880	.9905	.9988	.9938	.9680	.9533	.8893
Standard error	.0080	.0093	.0094	.0061	.0045	.0019	.0015	.0015	.0022	.0050	.0068	.0085
Tucumcari, N. Mex.:												
Persistence	1.0365	1.0186	.9603	.9558	.9530	.8864	.9321	.9243	.9600	.9527	.9430	.9683
Standard error	.0053	.0068	.0061	.0079	.0091	.0115	.0120	.0115	.0091	.0080	.0068	.0071
Winnemucca, Nev.:												
Persistence	.9035	.9160	.9005	.9115	.9522	.9560	.9845	.9770	.9560	.9137	.9300	.9510
Standard error	.0094	.0102	.0083	.0082	.0073	.0065	.0037	.0037	.0061	.0061	.0082	.0085
Yuma, Ariz.:												
Persistence	1.0368	.9749	.9964	.9855	1.0290	.9903	.9875	.9830	.9939	1.0208	.9888	.9748
Standard error	.0080	.0054	.0047	.0034	.0042	.0007	.0037	.0053	.0043	.0034	.0012	.0051

These values may now be used in obtaining O_X and σ_e. Substituting in the indicated equations and solving, we find $O_X = 0.049$ and $\sigma_e = 0.017$. The probability of observing no occurrence of a 20-day drought between June 1 and July 10 at Bottineau, N. Dak., is:

$$e^{-\left[0.049+\frac{0.017^2}{2}\right]}.$$

Solving, we find the probability of no occurrence to be very nearly 0.95. Hence there is about a 1 in 20 chance that a drought of 20 days will be observed.

It is to be noted that the introduction of the term σ_e has not changed the final probability figure to any significant degree. This is the usual condition when the parameters p, and $(M_r$:$M)_r$ are based upon an adequate record. However, it is well to calculate σ_e so as to be able to obtain an idea of the probable accuracy of the calculated O_X. In the given example a drought of 20 or more days would be expected to occur at Bottineau between June 1 and July 10 once every 20 years, and it can further be said, with a fair degree of confidence, that the true mean frequency lies between once every 15 and once every 31 years.

DROUGHT PROBABILITIES FOR INTERMEDIATE TIME INTERVALS

When the time interval under consideration is 3 months or less, drought probability can frequently be considered to vary linearly. A straight line, $f(x)$, is fitted to the monthly probabilities, and the constants p (mean drought probability) and m (slope of the fitted line) are obtained. Then, where

$$Y = p - \frac{m}{2}(N - n)$$

and

$$Z = p + \frac{m}{2}(N - n),$$

it can be shown that for one calendar year

$$O_X = \frac{M_0}{M}\left[Y^{n+1}\left(\frac{1}{Y} - \frac{1}{m(n+1)} + \frac{Y}{m(n+2)}\right) - Z^{n+1}\left(\frac{1}{m(n+1)} - \frac{Z}{m(n+2)}\right)\right].$$

Also, the standard errors of p and m can be evaluated from the standard errors of the monthly probabilities. For example, where 3 full months are included,

$$\sigma_p = \sqrt{\frac{\sigma_{p_1}^2 + \sigma_{p_2}^2 + \sigma_{p_3}^2}{3}},$$

and

$$\sigma_m = \sqrt{\frac{\sigma_n \quad \sigma_s^2}{60}}.$$

From which it follows that

$$\sigma_O = \frac{M_0}{M}\sqrt{\left(\sigma_p \frac{\partial O_X}{\partial p}\right)^2 + \left(\sigma_m \frac{\partial O_X}{\partial m}\right)^2}.$$

The standard error of (M_0/M) has been ignored, since experience has demonstrated that this error is always negligible in relation to the standard errors of p and m.

Taking partial derivatives with respect to p and m, and simplifying:

$$\sigma_O = \frac{M_0}{M}\sqrt{\sigma_p^2\left[\frac{Z^{n+1} \quad Z}{m} \quad \frac{Y^{n+1} \quad Y}{m} + nY^{n-1}\right]^2 + \sigma_m^2\left[\frac{Z^{n+1} \quad Z - Y^{n+1} \quad Y}{m} \quad \frac{N}{2} \quad n \quad \frac{O_X \quad Y^n}{m} \quad nY^n \quad \frac{N}{2} \quad n\right]^2}.$$

This rather complex form can be reduced with a minimum of difficulty, since most of the constants appearing in the formula are repeated from the original expression of O_X.

Let us take as an example the drought hazard during the growing season of corn in Illinois. Davis and Pallesen have shown that adequate rainfall is most necessary to the crop during the 40 days preceding and the 20 days following silking (5). In the vicinity of Marengo, Ill., planting time is during the first week in May (1). Silking occurs near July 25. A drought between June 15 and August 15 would, therefore, be particularly detrimental to the harvest, and 20 consecutive days with no precipitation of 0.10 inch or more within 48 hours would probably materially affect the crop. Let us calculate the probability of such a drought.

The monthly drought probabilities for June, July, and August for Marengo, Ill., are 0.7702 ± 0.0147, 0.8446 ± 0.0118, and 0.8300 ±

0.0090, respectively.[26] Consider these points centered at $x=-30$, $x=0$, and $x=30$, and weight them 1:2:1 in proportion to the number of days from each month in the required time interval. Then $p=0.8208\pm0.0085$ and $m=0.00100\pm0.00029$. The mean value of (M_r/M) is found to be 0.9281. These values in the given formulas yield the solutions $O_N=0.142$ and $\sigma_r=0.023$. From these data it can readily be ascertained that the probability of a 20-day drought at Marengo, Ill., between June 15 and August 15 is very close to 0.133.

It should be pointed out that calculation of drought probabilities in individual problems is necessarily a rather complex and lengthy task if linear variation in drought probability is assumed. It has been found, however, that a good part of the work lies in setting up the problem, fitting a straight line to the monthly probability figures, and obtaining p and m. Thus, if probabilities are desired for a large number of values of n for the same time interval, the problem is not greatly complicated. Forms have been set up that enable computation of probabilities by workers who possess no more than a knowledge of the elementary use of logarithms and simple arithmetic.

Drought Probabilities for Annual or General Time Intervals

In considering the likelihood of drought within long time intervals or within shorter intervals for which it is not preferable to apply one of the methods previously outlined, it will be necessary to calculate a family of annual drought probability curves for a station in the area under consideration. A family of such curves for Dallas, Tex., has been shown in figure 6.

Let us take an example of the use of these curves in calculating likelihood of drought. Suppose that one desires to know the probability of observing a drought in excess of 40 days at Dallas, Tex., between February 1 and December 10. To obtain O_N it is only necessary to obtain the mean ordinate on the curve for $n=40$ between February 1 and November 1 (40 days previous to December 10) and multiply by 273, the number of days in the shorter time interval. Using the data from which the curve for $n=40$ was plotted, the mean ordinate can be estimated as very nearly 0.00098, and $O_N=0.268$.

The value of σ_r could be obtained by calculating and combining the standard errors in the plotted curve for a number of evenly spaced values of x. However, this involved process does not seem necessary in view of the fact, shown in preceding examples, that consideration of σ_r does not materially change the probability. It will be accurate enough in general to consider the probability of lack of appearance of a drought of 40 days to be $e^{-0.268}$, or about 0.76. Hence the chance of such a drought appearing between February 1 and December 10 will be about 1 in 4.

It can be seen that the family of annual drought probability curves can be readily used in obtaining likelihood of drought. This method is particularly applicable in that it requires a minimum of calculation and can be employed by an untrained worker. Families of curves are in preparation for a large number of stations in the United States, and these curves will be available to climatologists for estimating probabilities of drought.

[26] Actually, a linear fit is not very close. It would probably be as accurate, and less involved, to obtain the probability from annual drought probability curves.

USES AND POSSIBLE EXTENSIONS OF THE PROBABILITY ANALYSIS OF DROUGHT

In appraising the applicability of probability theory to drought occurrence, it should be realized that the method of analysis can be utilized in many ways in addition to the calculation of drought probability at a single station as indicated on previous pages. Thus, by taking the mean drought probability for an area, drought hazard within that area can be measured for any drought length or time interval desired. Also, as will be later shown, the drought probability figures can be reordered in such a manner that they will be of value in a variety of fields related to climatology and meteorology.

Figure 8 presents the mean monthly drought probability for 216 selected stations (fig. 7) in the United States. Values from these maps may be used in approximating likelihood of drought within an area. The mean monthly drought probability for any designated month can be roughly obtained by inspecting the map for that month and taking the average of the drought probabilities for the stations within the desired area. This drought probability can be directly used in estimating drought hazard, since the likelihood of any number of days of drought will be approximately equal to the mean drought probability raised to that power. It is to be noted that where topographic relief is not great, the monthly probability values exhibit only slight variation from station to station. The amount of variation is proportional, of course, to the variability of drought hazard within the area.

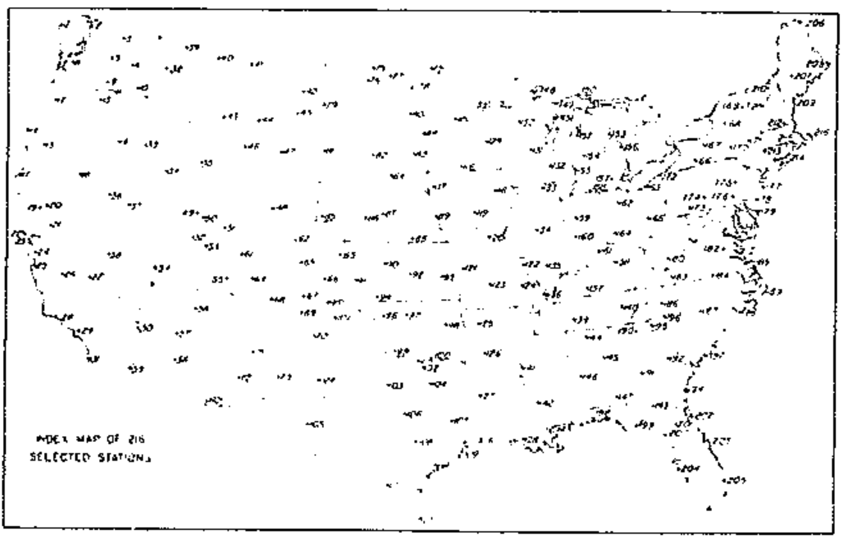
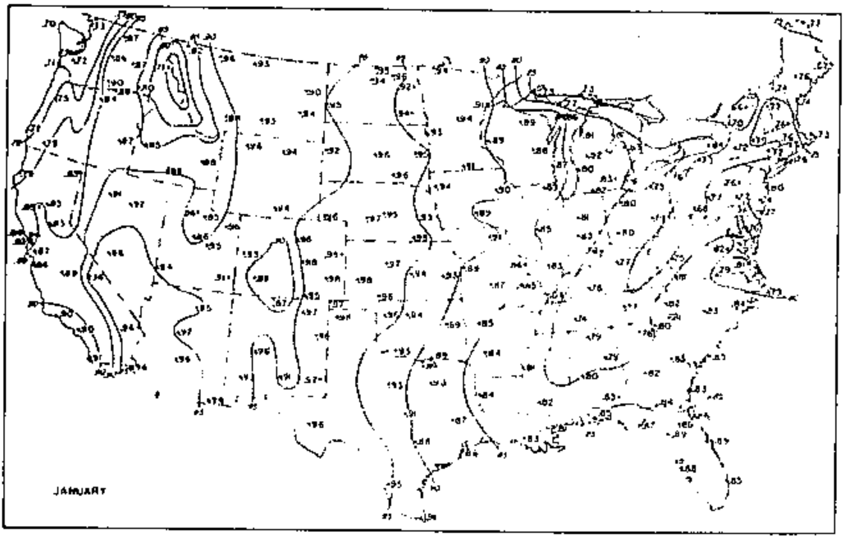


FIGURE 7.—Index list of 216 selected stations: Washington: 1, Tatoosh Island; 2, Clearbrook; 3, Conconully; 4, Olympia; 5, Wenatchee; 6, Odessa; 7, North Head; 8, Centralia; 9, Sunnyside; 10, Dayton. Oregon: 11, Umatilla; 12, Corvallis; 13, Condon; 14, Port Orford; 15, Jacksonville; 16, Riverside. California: 17, Eureka; 18, Cedarville; 19, Willows; 20, Chico; 21, Rocklin; 22, Point Reyes; 23, San Francisco; 24, San Jose; 25, Salinas; 26, Fresno; 27, Independence; 28, Santa Barbara; 29, Los Angeles; 30, Needles; 31, San Diego. Idaho: 32, Kellogg; 33, Boise; 34, Twin Falls; 35, Blackfoot. Nevada: 36, Winnemucca; 37, Beowawe; 38, Tonopah. Montana: 39, Fortine; 40, Cut Bank; 41, Havre; 42, Circle; 43, Bozeman; 44, Billings; 45, Miles City. Wyoming: 46, Shoshone Canyon Dam; 47, Buffalo; 48, Saratoga. Utah: 49, Salt Lake City; 50, Park City; 51, Fort Duchesne; 52, Levan; 53, Castle Dale; 54, Modena; 55, Blanding. Arizona: 56, Tuba City; 57, Jerome; 58, Phoenix; 59, Yuma; 60, Bisbee. Colorado: 61, Grand Junction; 62, Boulder; 63, Burlington; 64, Silverton; 65, Colorado Springs; 66, Las Animas; 67, Trinidad. New Mexico: 68, Chama; 69, Springer; 70, Tucumcari; 71, Magdalena; 72, Fort Bayard; 73, Cloudcroft; 74, Lovington. North Dakota: 75, Bottineau; 76, Granville; 77, Cando; 78, Larimore; 79, Beach; 80, Lisbon. South Dakota: 81, Spearfish; 82, Pierre; 83, Desmet; 84, Academy. Nebraska: 85, Kimball; 86, North Platte; 87, Broken Bow; 88, Hebron; 89, Omaha. Kansas: 90, Hays City; 91, Lakin; 92, McPherson; 93, Burlington; 94, Ashland. Oklahoma: 95, Kenton; 96, Mutual; 97, Waukomis; 98, Muskogee; 99, Chattanooga; 100, Durant. Texas; 101, Dalhart; 102, Gainesville; 103, Albany; 104, Dallas; 105, Fort Davis; 106, Lampasas; 107, Huntsville; 108, Luling; 109, Galveston; 110, Fort McIntosh; 111, Brownsville. Minnesota: 112, Roseau; 113, Duluth; 114, Beardsley; 115, Fort Ripley; 116, Winnebago. Iowa: 117, Le Mars; 118, Independence; 119, Indianola. Missouri: 120, Kirksville; 121, Harrisonville; 122, St. Louis; 123, Lebanon; 124, Arcadia. Arkansas: 125, Dutton; 126, Amity. Louisiana: 127, Grand Cane; 128, Franklin. Wisconsin: 129, Downing; 130, Minocqua; 131, Waupaca; 132, Port Washington. Illinois: 133, Marengo; 134, Lincoln; 135, Mount Vernon. Kentucky: 136, Blandville; 137, Bowling Green; 138, Lexington. Tennessee: 139, Hohenwald; 140, Kingston. Mississippi: 141, Greenville; 142, Brookhaven; 143, Pearlington. Alabama: 144, Decatur; 145, Talladega; 146, Greensboro; 147, Ozark. Michigan: 148, Calumet; 149, Munising; 150, Whitefish Point; 151, Escanaba; 152, Traverse City; 153, Harrisville; 154, Big Rapids; 155, Grand Haven; 156, Harbor Beach; 157, Ann Arbor; 158, Centerville. Indiana: 159, Delphi; 160, Greencastle; 161, Madison. Ohio: 162, Findlay; 163, Cleveland; 164, Waynesville; 165, Philo. New York: 166, Angelica; 167, Shortsville; 168, Lowville; 169, Lake Placid; 170, Norwich; 171, West Berne. Pennsylvania: 172, Erie; 173, Somerset; 174, Huntington; 175, Mauch Chunk; 176, Ephrata. New Jersey: 177, Asbury Park; 178, Indian Mills; 179, Cape May City. West Virginia: 180, Beckley. Maryland: 181, Cambridge. Virginia: 182, Ashland; 183, Wytheville; 184, Clarksville; 185, Norfolk. North Carolina: 186, Marion; 187, Rockingham; 188, Sloan; 189, Hatteras. Georgia: 190, Dahlonega; 191, Marshallville; 192, Louisville; 193, Quitman; 194, Brunswick. South Carolina: 195, Clemson College; 196, Spartanburg; 197, Charleston. Florida: 198, Pensacola; 199, Carrabelle; 200, Cedar Keys; 201, Gainesville; 202, St. Augustine; 203, Titusville; 204, Bradenton; 205, Hypoluxo. Maine: 206, Van Buren; 207, Madison; 208, Eastport; 209, Portland. Vermont: 210, Bloomfield; 211, Chelsea. New Hampshire: 212, Nashua. Connecticut: 213, Cream Hill; 214, New Haven. Rhode Island: 215, Providence. Massachusetts: 216, Provincetown.



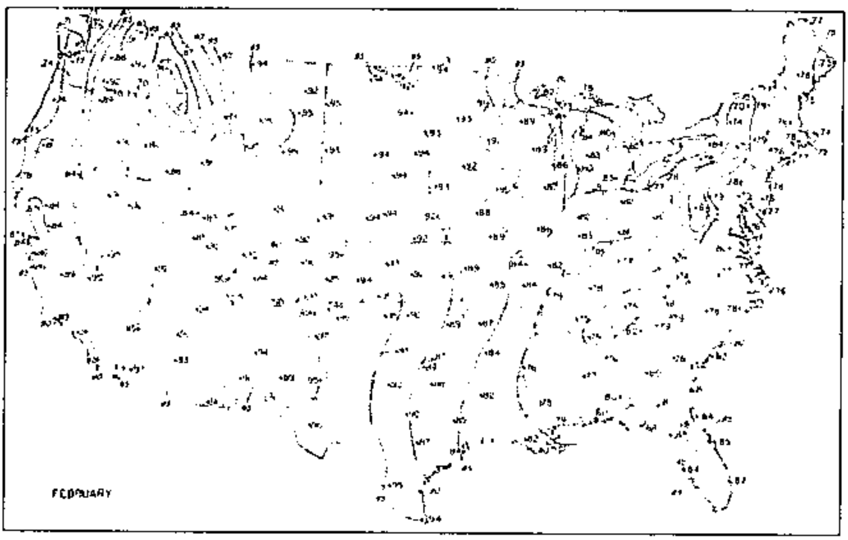


FIGURE 8.—Monthly drought-probability maps for 216 selected stations in the United States.

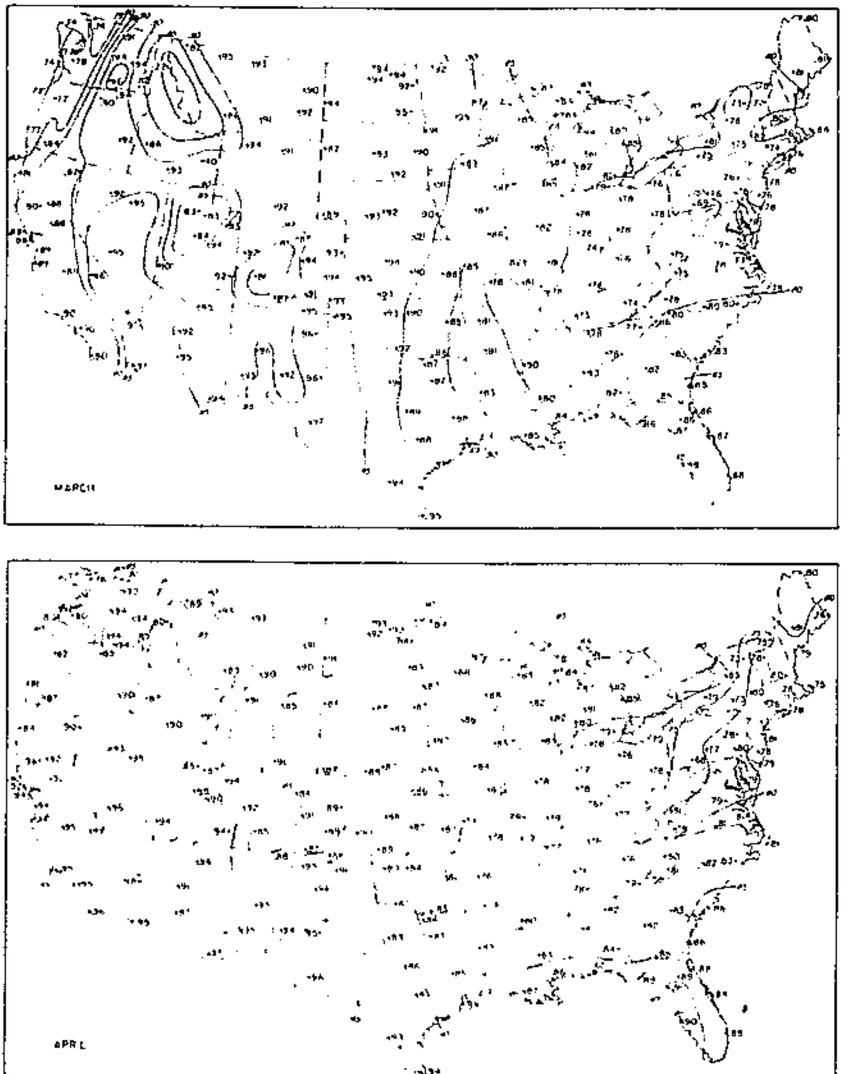
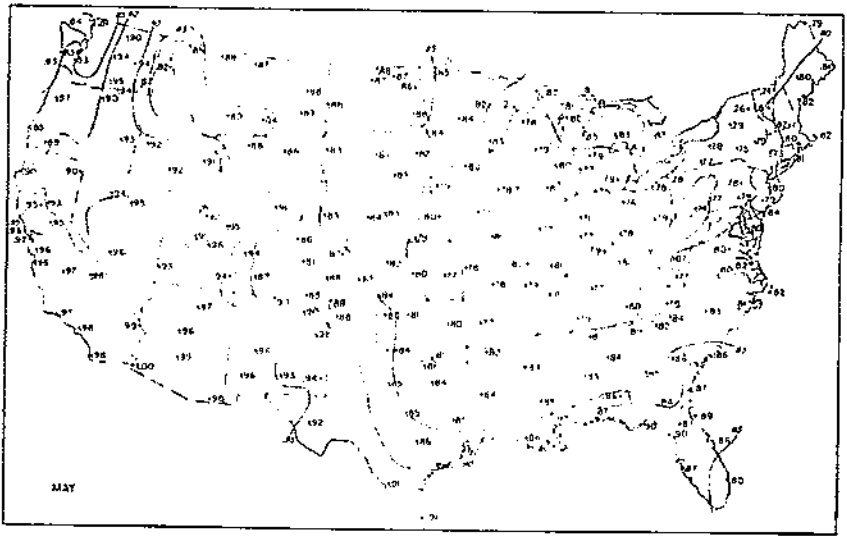


FIGURE 8.—Monthly drought-probability maps for 216 selected stations in the United States—Continued.



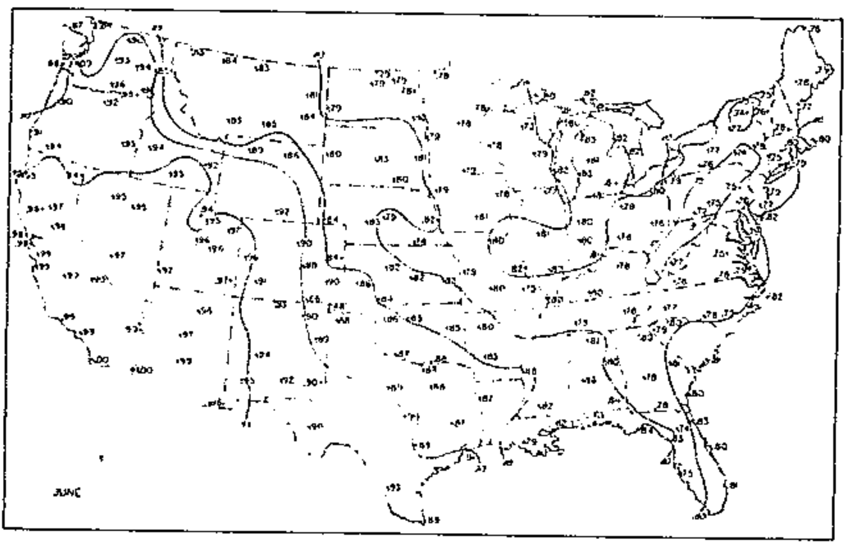


FIGURE 8.—Monthly drought-probability maps for 216 selected stations in the United States—Continued.

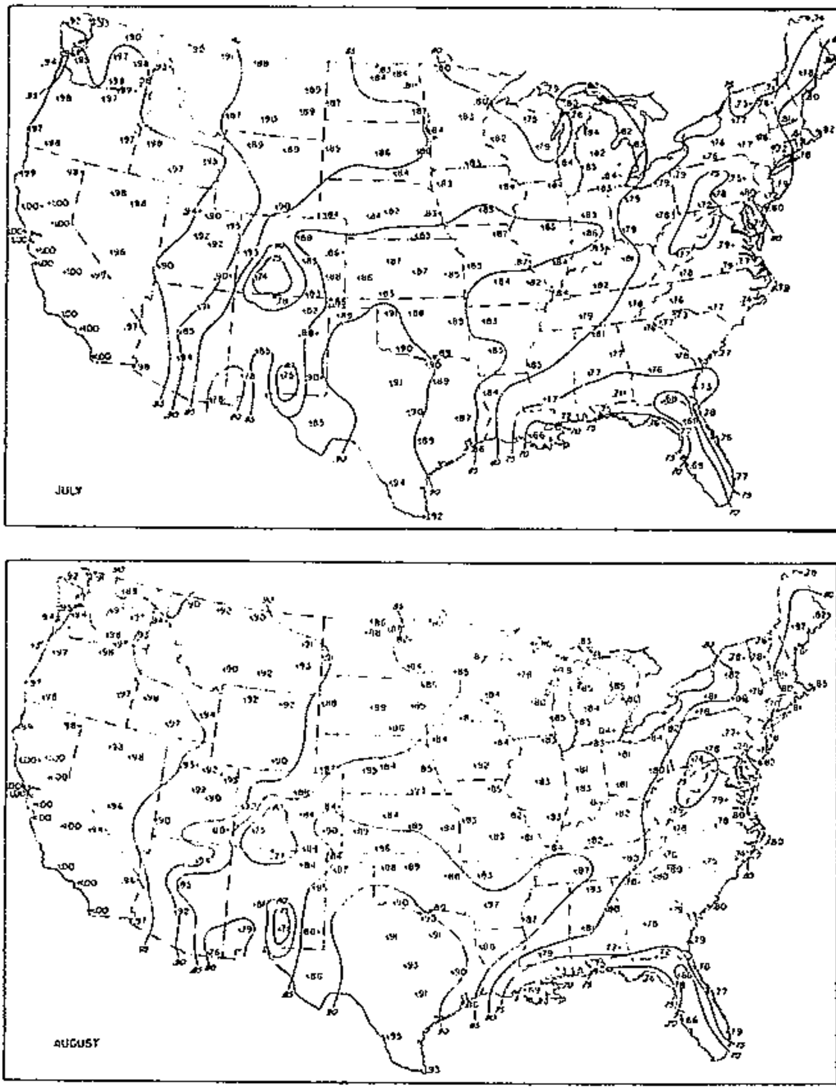
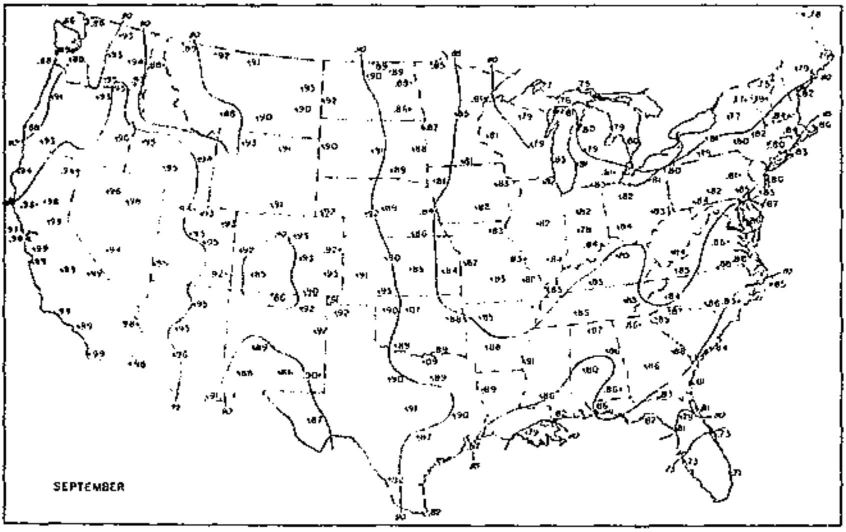


FIGURE 8.—Monthly drought-probability maps for 216 selected stations in the United States—Continued.



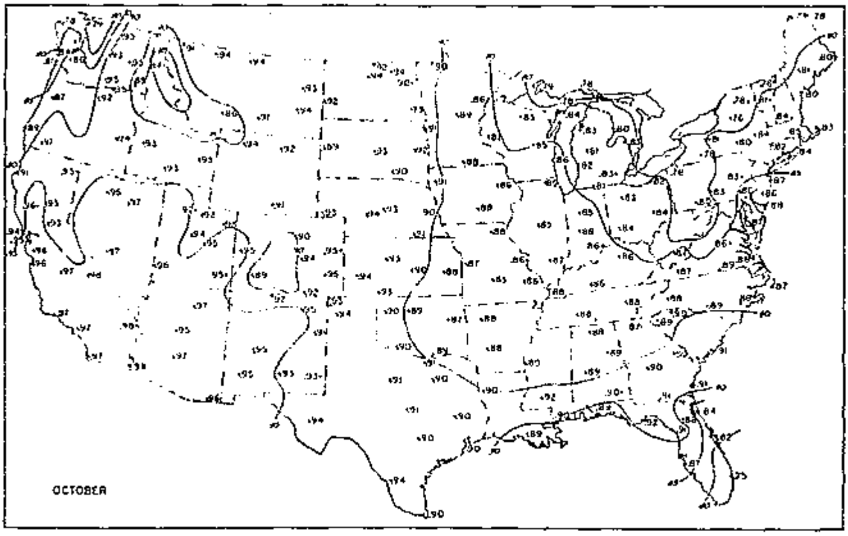
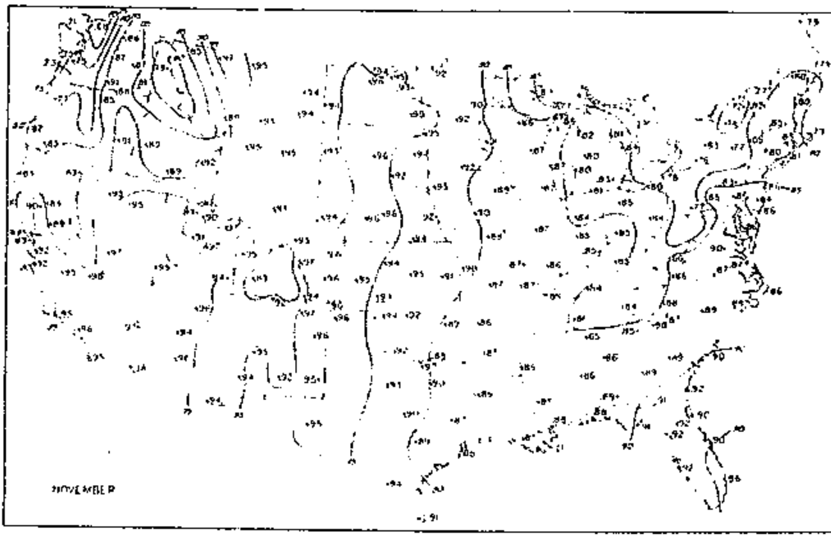


FIGURE 8.—Monthly drought-probability maps for 216 selected stations in the United States—Continued.



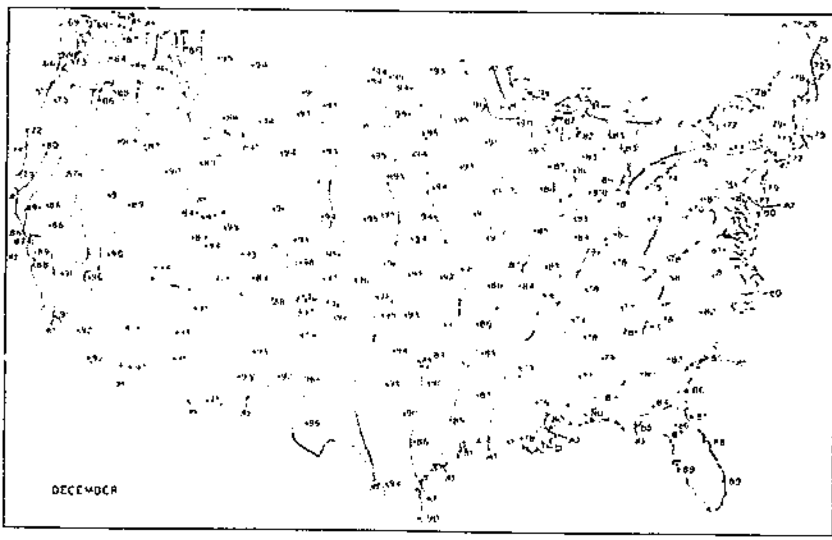


FIGURE 8. Monthly drought-probability maps for 216 selected stations in the United States—Continued.

The drought-probability figures can be used in another way. The family of annual drought-probability curves shown in figure 6 can be modified and put into a new form that will be at least equally valuable. In figure 6 the curves show the chance of a drought of n days or more beginning on any specified day. By estimating the area under the curve between any two time limits, the number of droughts of n days or more expected within the desired time interval can be obtained, and the probability of observing 0, 1, 2, . . . such droughts can be calculated. However, the curves are of limited value in that any single ordinate has little practical meaning. The ordinates do not furnish a measure of drought hazard since they specify rainfall immediately preceding the day plotted on the abscissa. Hence the formula for the curves includes the factor $1-p$, which approaches 0 as p approaches 1. If the curves are to be a useful measure of drought hazard, this factor must be eliminated.

This can be done by removing the restriction that the drought of n or more days must start on the day indicated on the abscissa. The curves will then represent the chance of n or more days of drought irrespective of weather conditions preceding the n days. Such a family of curves for Dallas, Tex., shown in figure 9, represents the variation of mean drought conditions with season. Curves of this type will be of value in making and verifying synoptic weather forecasts and will also be useful since they show in simple, utilizable form the time intervals within the year when sequences of dry or rain days are most likely to occur.

Suppose that on July 1 a drought is in progress at Dallas, and that synoptic forecasting indicates that precipitation is unlikely to occur within the next 2 or 3 days. It will be important to be able to estimate statistically the number of additional days of drought which may be expected. Examination of figure 9 shows that under mean conditions the drought would last 6 more days. The chances are 1 in 10 that it will last 20 more days, and about 1 in 30 that it will last 30 more days. Assuming that occurrence of drought is random, these odds will not be affected by the number of days of drought preceding the date of observation, July 1.[27] It is not suggested, of course, that statistical analysis can be substituted for synoptic forecasting; however, knowledge of the likelihood of climatic events, based solely upon past record, can be of value in day-to-day forecasting as well as in long-range climatic analyses.

The family of curves in figure 9 can also be used in other ways. Suppose that a conservation program has been outlined that requires the planting of extensive areas of grassland. If the grass is to become established it will be necessary that no lethal drought occur immediately after planting. By obtaining the abscissa for which the drought-probability curves are a minimum it will be possible to determine the approximate date following which a killing drought is least likely. At Dallas, a 10-day drought is least likely in the 10-day period beginning about April 25, and a 25-day drought is least likely in the period beginning about April 18. If the time required for seed germination and initial growth of the plant is known, the date most likely to be favorable for planting can be determined.

---

[27] Although occurrence of short drought has been shown to be not strictly random, the error in taking the curves as representative of mean conditions for small as well as large values of n will, in practice, be too small to be of significance.

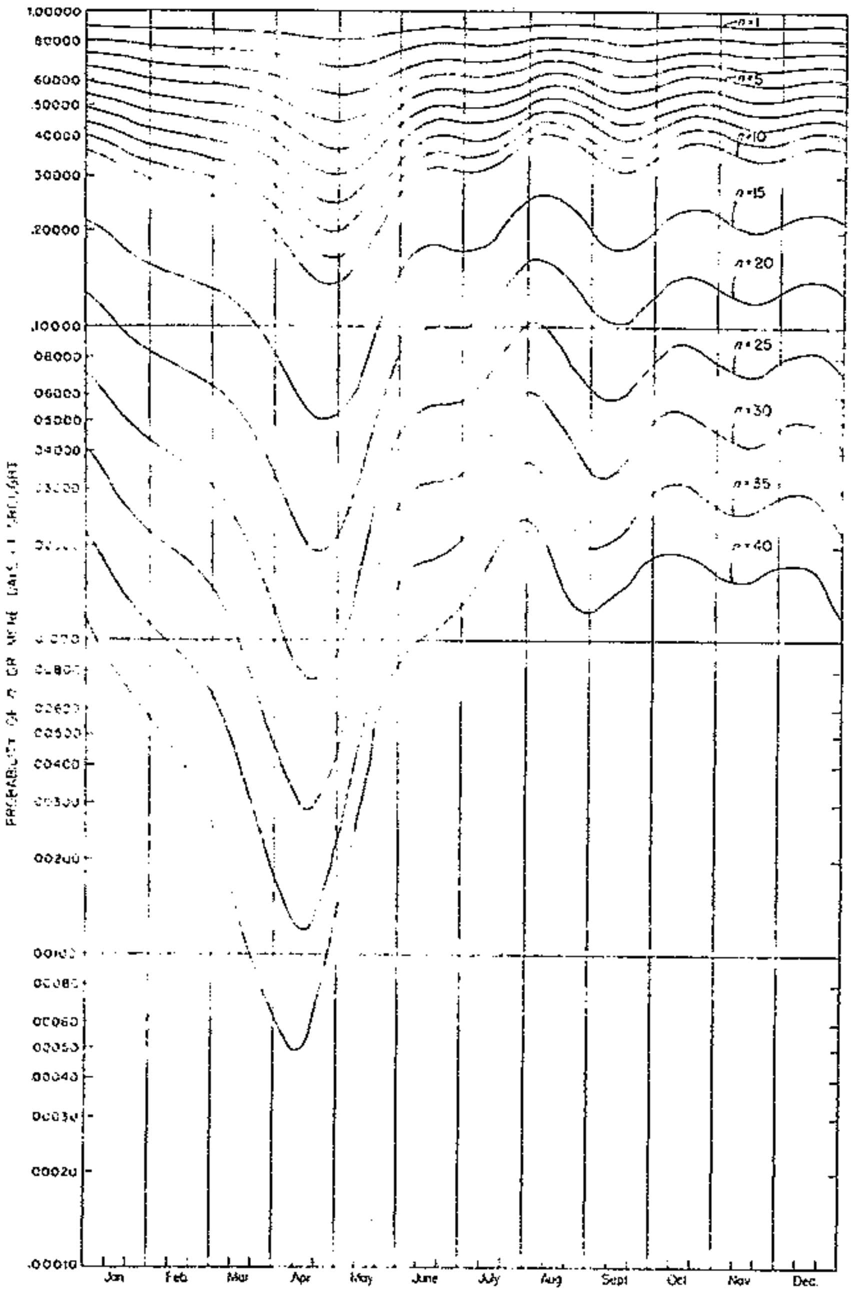


FIGURE 9. Curves for Dallas, Tex., showing the probability on each day of the year of n or more days of drought.

In this example it was required to know the date on which drought is least likely. The time at which drought is most likely may also be required and may be obtained from the curves. Assume that it is decided to undertake a program of terracing during a dry season to avoid extensive erosion in the construction of terraces. The drought-probability curves will indicate the time of year at which this danger is at a minimum. At Dallas, Tex., terracing work requiring 20 days for completion should be begun about August 5 in order to be least liable to interference by rain.

In all these examples, the term "drought" has been applied to sequences of days wherein no more than 0.10 inch of precipitation fell in any 2-day period. That has been the definition of drought adopted in making the probability analysis of drought occurrence. However, it will be found that for many practical purposes this definition of drought is inappropriate and that a change of definition would be desirable. There is no apparent meteorological reason why similar analyses based on other definitions should not yield similar results or why occurrence of drought should not again prove to be substantially random. In such studies, the monthly drought probability would, of course, increase with an increase in the amount of precipitation considered necessary to terminate a drought. In addition, it is possible that the minimum drought length above which random occurrence of drought may be assumed will change. If rainfall amounts of 1 or 2 inches are required to terminate a drought, it is possible that the occurrence of even the shortest droughts may be random, since the effect of precipitation from successive frontal passages will have been largely eliminated.

These questions are open to further study. Through probability analysis of drought occurrence for other definitions it may be possible to determine the change in drought probability with change in definition.[28] Drought occurrence could then be described as a function of three variables: Season of the year, minimum length of drought, and amount of precipitation necessary to terminate the drought. Although it may prove impractical to set up an algebraic solution of the general problem, a graphic solution would certainly be feasible. A series of families of annual drought probability curves for various definitions could be prepared, and could be used to solve many problems concerning the likelihood of occurrence of drought or rainfall.

The probability analysis as applied to drought occurrence in this study may well be of value in the examination of other climatic data, such as rainfall intensity and frequency of extreme temperatures. Many of the problems encountered in such analyses parallel problems in the study of drought. With some modifications and alterations the probability analysis developed here will prove applicable to a variety of climatic data. Through the continued application of this technique further progress can be made in understanding and statistically describing areal and temporal variation in climate.

---

[28] Preliminary investigations, conducted for 8 definitions of drought for Washington, D. C., show that a curve of the form

$$d = k_1 + k_2 \frac{\log(1-p)}{\log p} + k_3 \frac{1}{\log p}$$

fits the observed probabilities very closely. Here p is the mean monthly drought probability, d is the definition of drought, expressed in hundredths of an inch of precipitation necessary to terminate a drought, and k_1, k_2, and k_3 are constants. If this relationship holds, there would be a linear relationship between d and a minimum length of drought in days when O_N is held constant.

SUMMARY

This bulletin develops a statistical method for the description of occurrence of drought and shows how drought hazard can be expressed in terms of probability or chance. Drought occurrence was tabulated for a large number of stations in the United States for the interval 1898–1937. Length of each drought was expressed in days, and frequency distributions were prepared showing the number of occurrences of droughts of various length. For the purposes of this study, drought was considered terminated by 0.10 inch or more of precipitation in 48 hours or less.

The theory of runs was used to determine the annual distribution of drought to be expected from a random arrangement of dry and rainy days. If occurrence of a day of drought were random with respect to a constant probability, the frequency of drought would vary as an exponential function of length. Distributions approximating such exponential curves were observed. It was noted, however, that for different climatic areas the closeness with which the observed distributions approximated an exponential curve was, in general, proportional to the uniformity of drought hazard from season to season.

To apply the theory of runs adequately, it was necessary to assume that drought probability varied from day to day. It was assumed, further, that the variation in drought probability within the year could be expressed as a continuous function. Curvilinear variation in drought probability was then approximated by computing a mean probability of a day of drought for each month, assuming a constant probability throughout the month. A periodic function was used to fit the 12 mean drought probabilities thus obtained.

Analysis of drought by the theory of runs for monthly, seasonal, and annual time intervals indicated that observed drought differed from random occurrence in two ways:

(1) It was found that days of observed rainfall tended to cluster more than might be expected from a random arrangement of dry and rainy days. This may have been anticipated. Many types of storms tend to last longer than a single day. In addition, since 24-hour amounts were used in computing drought, even a short storm might frequently result in the tabulation of 2 days of rain, since the storm might easily occur across the arbitrary time limit dividing two days.

The clustering of days of rainfall does not preclude the possibility that once a dry period has begun additional days of drought occur at random. This concept was retained, and drought probability was defined in terms of number of droughts and total number of days of drought observed, without regard to the distribution of days of rainfall. By defining drought probability in this manner, the fit of the theoretical distributions was greatly improved.

(2) In further comparing the theoretical and the observed drought distributions, it was found that there was an excess of observed short "droughts" (1 to 5 or 6 days in length) and a deficiency of longer droughts. However, the deficiency rapidly disappeared with increase in drought length, and for droughts longer than 16 to 20 days the expected distributions adequately represented observed conditions.

These results may also be expected meteorologically. In many areas frontal precipitation occurs at irregular intervals, averaging 4 or

5 days apart. Together with prefrontal and postfrontal showers these recurrent frontal storms would tend to make short droughts more likely than random occurrence would indicate. Under such conditions, droughts of moderate length (5 or 6 to 16 to 20 days) would be observed less frequently than anticipated. However, in the case of longer droughts (over 16 to 20 days) the factor of recurrent frontal passage would be minimized, since a drought of this length would indicate either that some unusual meteorological condition caused a break-down in the usual pattern of recurrent frontal precipitation, or that the station was in an area in which precipitation was infrequent or not generally associated with frontal passage.

Since observed occurrence of long droughts was adequately represented by the theoretical distributions, these distributions were used in estimating future drought hazard. In estimating drought likelihood it was necessary, of course, to take into consideration the fact that the parameters used in constructing the drought probability curves were not known a priori, but were obtained from observed data.

Through the use of the techniques developed, it has been shown how drought hazard can be approximated where the minimum length of the required drought and the time interval within which it must occur were adjusted at will. Nevertheless, the bulletin is primarily a methodologic study. Only one definition of drought has been used, and for that definition the data have been presented in the most useful form for only a sample station. (See fig. 6.) However, there is no apparent reason why the results observed in this study should not be repeated with a change in definition. If the definition can be changed, a series of families of curves similar to figure 6 can be prepared for any station or area. These curves would make possible the estimation of drought hazard where the definition of drought also can be chosen to fit the problem at hand.

Construction of families of such curves for a large number of stations is now in progress. These curves will be of significant value to climatologists, soil conservationists, agronomists, and other scientists who are concerned with the causes or effects of drought.

LITERATURE CITED

(1) BAKER, OLIVER E., BROOKS, CHARLES F., COVERT, JAMES R., and HAINSWORTH, REGINALD G.
1922. SEEDTIME AND HARVEST: A GRAPHIC STUDY OF SEASONAL WORK ON FARM CROPS. U. S. Dept. Agr. Dept. Cir. 183, 53 pp., illus.

(2) BARTELS, J.
1935. RANDOM FLUCTUATIONS, PERSISTENCE, AND QUASI-PERSISTENCE IN GEOPHYSICAL AND COSMICAL PERIODICITIES. Jour. Terrestrial Magnetism and Atmos. Elect. 40: 1-60, illus.

(3) BERNARD, MERRILL M.
1932. FORMULAS FOR RAINFALL INTENSITIES. Amer. Soc. Civ. Engin. Trans. 96: [592]-606, illus.

(4) COCHRAN, W. G.
1938. AN EXTENSION OF GOLD'S METHOD OF EXAMINING THE APPARENT PERSISTENCE OF ONE TYPE OF WEATHER. Roy. Met. Soc. [London] Quart. Jour. 64: 631-634.

(5) DAVIS, FLOYD E., and PALLESEN, J. E.
1940. EFFECT OF THE AMOUNT AND DISTRIBUTION OF RAINFALL AND EVAPORATION DURING THE GROWING SEASON ON YIELDS OF CORN AND SPRING WHEAT. Jour. Agr. Res. 60: 1-23, illus.

(6) FISHER, R. A.
1938. STATISTICAL METHODS FOR RESEARCH WORKERS. Ed. 7, rev. and enl., 356 pp., illus. Edinburgh and London.

(7) GOLD, E.
1929. NOTE ON THE FREQUENCY OF OCCURRENCE OF SEQUENCES IN A SERIES OF EVENTS OF TWO TYPES. Roy. Met. Soc. [London] Quart. Jour. 55: 307-309.

(8) GREENWOOD, MAJOR, and YULE, G. UDNY.
1920. AN INQUIRY INTO THE NATURE OF FREQUENCY DISTRIBUTIONS REPRESENTATIVE OF MULTIPLE HAPPENINGS WITH PARTICULAR REFERENCE TO THE OCCURRENCE OF MULTIPLE ATTACKS OF DISEASE OR OF REPEATED ACCIDENTS. Roy. Statis. Soc. Jour. 83: [255]-279.

(9) HANN, JULIUS.
1908-11. HANDBUCH DER KLIMATOLOGIE. Umgearb. u. verm. Aufl. 3, 3 v., illus. Stuttgart. [Ed. 4, 1932, only v. 1 published.]

(10) MIAMI CONSERVANCY DISTRICT ENGINEERING STAFF.
1936. STORM RAINFALL OF EASTERN UNITED STATES. Rev. ed., 352 pp., illus. Dayton, Ohio. (Miami [Ohio] Conserv. Dist. Tech. Rpts., pt. 5.)

(11) MOOD, A. M.
1940. THE DISTRIBUTION THEORY OF RUNS. Ann. Math. Statis. 11: 367-392.

(12) PEARSON, KARL, ed.
1930. TABLES FOR STATISTICIANS AND BIOMETRICIANS. Ed. 3, pt. 1, 143 pp., illus. London.

(13) SHERMAN, CHARLES W.
1931. FREQUENCY AND INTENSITY OF EXCESSIVE RAINFALLS AT BOSTON, MASSACHUSETTS. Amer. Soc. Civ. Engin. Trans. 95: [951]-968, illus. [Discussion, pp. 960-968.]

(14) THORNTHWAITE, C. W.
1941. ATLAS OF CLIMATIC TYPES IN THE UNITED STATES, 1900-1939. U. S. Dept. Agr. Misc. Pub. 421, 7 pp., 96 maps.

(15) TOLLEY, HOWARD ROSS.
1916. FREQUENCY CURVES OF CLIMATIC PHENOMENA. U. S. Monthly Weather Rev. 44: 634-642, illus.

(16) YARNELL, DAVID L.
1935. RAINFALL INTENSITY-FREQUENCY DATA. U. S. Dept. Agr. Misc. Pub. 204, 68 pp., illus.

END