CLIMATOLOGICAL ANALYSIS OF SNOWFALL THRESHOLDS

by

H.C.S. THOM

Advisory Committee on Weather Control
Washington, D.C.

CLIMATOLOGICAL ANALYSIS OF SNOWFALL THRESHOLDS

by

H.C.S. THOM

Advisory Committee on Weather Control
Washington, D.C.

Abstract

The climatological analysis for the first occurrence of 24-hour snowfalls of various depths is developed. This gives the probability function of snowfall thresholds which in turn makes any quantiles or probability values readily available. The probability function together with a special convention on years with no threshold value leads also to a definition of mean or expected threshold date. Examples of applications to several thresholds for New England data are presented.

Introduction

We define a meteorological time-threshold variable to be a variable expressed in time units, each value of which is associated with the passing of some critical meteorological value. The most common of this type of variable employed is the last date of t^o - freeze in spring and first date of t^o - freeze in fall. Here date is the time variable and t^o is threshold which is considered to be passed when the temperature is equal to or less than t^o. In general the passing of the threshold is in the direction of accentuating some effect, e.g., if a plant freezes at 32^o it also freezes at any lower temperature.

In this discussion we consider a time-threshold variable applied to 24-hour snowfall. This variable is the date of the first day in fall on which d-inches or more snowfall occurs. Such a variable has already had application in the motor vehicle tire industry and with varying values of d might well be applied to problems in snow removal and winter sports. The problems associated with these activities could of course be either operational or planning in nature. We shall consider only the planning problem which naturally involves climatological analysis.

For planning problems in general the main need is for quantiles or values of the variable associated with critical probabilities. Thus the tire manufacturer may wish to know on what date he may be assured with a 0.10 probability that no one inch 24-hour snowfall has occurred before. The winter sports enthusiast or sports club operator may wish to know the date on which the probability is a given value that a 24-hour snowfall of suitable amount will have fallen. A highway department may wish to know the date before which the chances of use of their snow removal equipment is small and the date by which heavier equipment must be available for use. The central objective of our discussion will therefore be to develop the probability theory for snowfall thresholds of a specific kind and apply this theory to the stations shown in Figure 1.

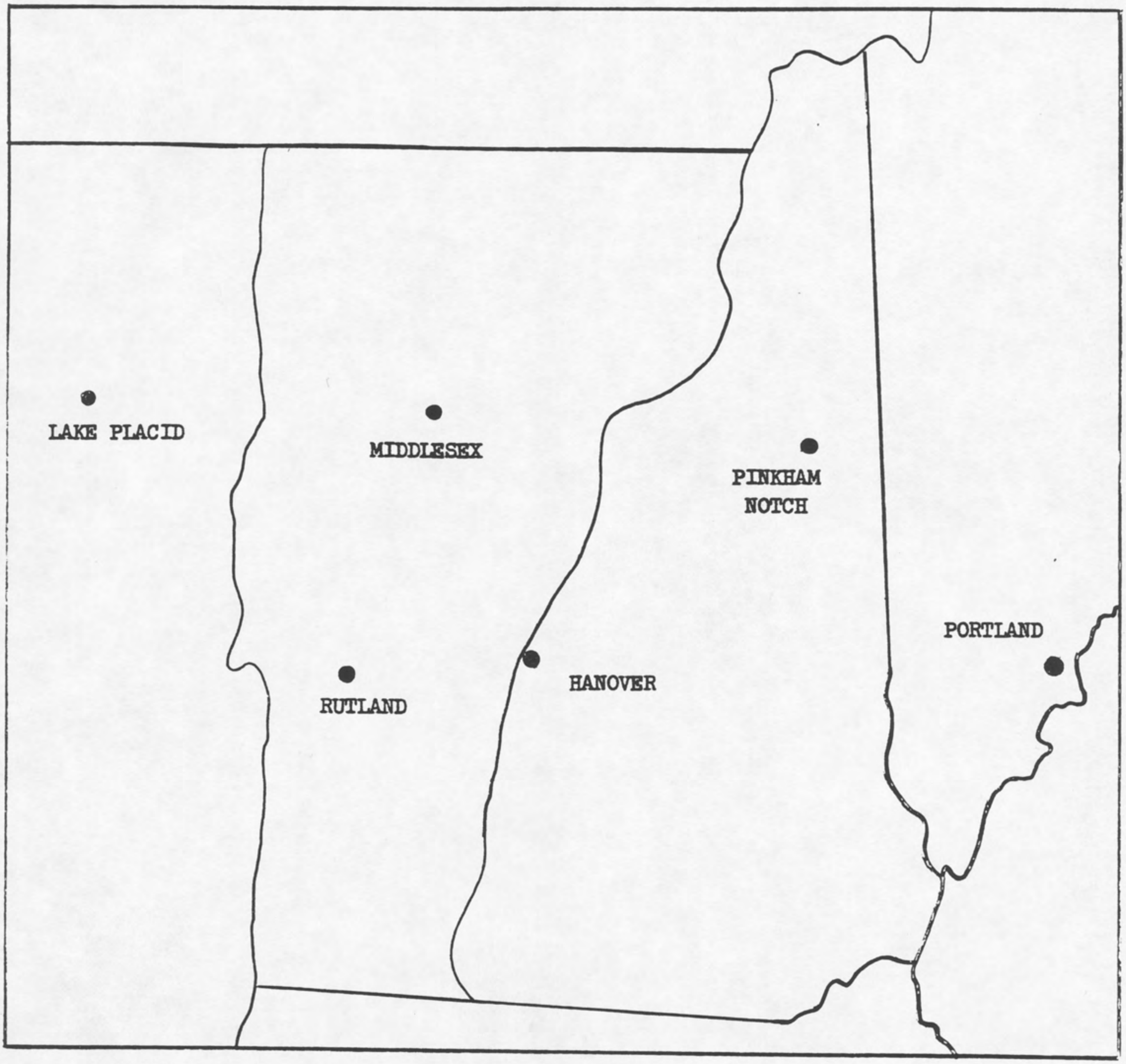
LAKE PLACID
MIDDLESEX
PINKHAM
NOTCH
PORTLAND
RUTLAND
HANOVER

FIGURE 1

The Special Threshold Distribution

Since the date of first occurrence of d-inches or more of snowfall each year forms a random series through years, our concern will be to find the frequency distributions of such series. There are two situations which arise: (a) The special case in which the threshold value occurs in all years or the probability of it occurring is one. (b) The general case in which the threshold value occurs with a probability p which can vary from zero to one. In the latter situation it will be noted that the threshold value will not in general occur in all years; so there will be years in the series without a day having a snowfall equal to or greater than d-inches. We treat the special case first.

From other work particularly on freeze thresholds (Reference 1 & 2) where threshold dates were found to be approximately normally distributed, it was conjectured that snowfall threshold dates might also be so distributed. Normal distributions were fitted to complete data series for those stations of Table 1 which showed probabilities $p = 1$ and to incomplete series consisting of only the years which actually experienced thresholds and for which p is accordingly less than one. In respect to year order these incomplete series are random samples from the complete series.

It will be noted in Table 1 that for all stations $p = 1$ when the threshold $T = 1, 2,$ or 4 inches. For Pinkham Notch and Lake Placid $p = 1$ also for $T = 6$ inches. The present statistical analysis was performed on coded dates or day numbers representing the threshold dates. These day numbers were obtained by numbering days consecutively to 365 beginning with July 1. The usual method of fitting the normal distribution was employed; i.e., by estimating the mean and standard deviation. These estimates were obtained from the formulas.

$$\bar{t} = \frac{\sum x}{n} \quad \text{and} \quad s = \left(\frac{\sum (t - \bar{t})^2}{n - 1} \right)^{1/2}$$

where t is date and n is the number of years of record. In Table 1 $\bar{t}$ is expressed for convenience in date while s is left in number of days as of course it must be since it is a scale value.

TABLE 1

T	$\bar{t}$	$\bar{t}'$	s	p	g_1	a	Quantiles			
							.10	.30	.70	.90
				Portland, 31 years						
1	12/1	12/1	16.2	1.00	-.07	.82	11/10	11/22	12/9	12/21
2	12/8	12/8	18.2	1.00	-.01	.81	11/14	11/28	12/17	12/31
4	12/28	12/28	24.2	1.00	1.13	.70	11/27	12/16	1/10	1/28
6	1/5	1/11	28.6	.97	.62	.76	11/29	12/22	1/22	2/16
8	1/11	3/19	26.4	.61	.18	.80	12/17	1/11	-----	-----
				Hanover, 29 years						
1	11/22	11/22	20.3	1.00	-.22	.75	10/27	11/12	12/3	12/18
2	11/30	11/30	21.0	1.00	-.07	.74	11/3	11/19	12/11	12/27
4	12/26	12/26	25.5	1.00	-.30	.77	11/22	12/12	1/7	1/26
6	1/5	1/22	33.8	.90	.05	.71	11/24	12/21	1/31	-----
8	1/25	3/27	44.1	.61	.13	.81	12/13	1/24	-----	-----
				Pinkham Notch, 20 years						
1	11/1	11/1	15.5	1.00	.06	.85	10/12	10/24	11/9	11/21
2	11/13	11/13	17.9	1.00	-.28	.71	10/21	11/4	11/22	12/6
4	11/27	11/27	23.2	1.00	.01	.82	10/28	11/15	12/9	12/27
6	12/9	12/9	32.0	1.00	.47	.80	10/29	11/23	12/26	1/19
8	12/24	1/2	45.7	.95	.28	.84	10/30	12/3	1/22	-----
				Rutland, 31 years						
1	11/26	11/26	20.3	1.00	-.15	.83	10/31	11/15	12/6	12/22
2	12/4	12/4	25.8	1.00	.02	.74	11/1	11/21	12/18	1/6
4	12/22	12/22	29.4	1.00	.27	.77	11/14	12/6	1/6	1/28
6	1/21	2/20	40.0	.81	.67	.86	12/6	1/8	3/6	-----
8	1/14	4/24	39.3	.40	.06	.73	12/19	2/9	-----	-----
				Lake Placid, 30 years						
1	11/3	11/3	15.5	1.00	-.49	.85	10/15	10/26	11/11	11/23
2	11/10	11/10	18.4	1.00	-.42	.79	10/18	11/1	11/20	12/4
4	11/26	11/26	23.4	1.00	-.25	.75	10/27	11/14	12/8	12/26
6	11/29	12/29	38.6	1.00	.82	.78	11/9	12/9	1/18	2/16
8	1/7	1/20	42.7	.93	.11	.81	11/16	12/18	2/5	3/29

Measures of skewness g_1 and kurtosis $\underline{a}$ were also computed to test whether the threshold date t could be considered $\underline{to}$ be nearly normally distributed for all values of T. g_1 is the standardized third moment and $\underline{a}$ is the standardized first absolute moment. Tables of the distribution of these statistics for making tests for normality have been given by Geary (Reference 4). As has been found previously in certain meteorological data (Reference 3) skewness and kurtosis vary above and below the values expected for normality with some values significant. This is partly due to random sampling but a part is also due to the dependence between the thresholds causing a greater variability among the individual g_1's and a's shown in Table 1.

If we assume that the five sets of six g_1's and a's in the table are independent and that they all are based on 30 year records, the mean (g_1) and mean (a) will have distributions equivalent to sampling with n = 30 x 30 = 900. The means of g_1 and $\underline{a}$, 0.111 and .784, computed from Table 1, are well within the 10% limits for n = 900. Since the individual data comprising the g_1's and a's are not independent but positively correlated, the significance limits for the means would actually be broader than for independent data as assumed in the tables (Reference 4). Hence the hypothesis that the g_1's and a's are not normally distributed may be rejected with a higher probability than indicated and the suspected departures from normality are all the less significant. Further statistical analysis will therefore be made on the assumption that the threshold dates are normally distributed.

The General Threshold Distribution

As mentioned above the general threshold distribution will be applicable to thresholds having both complete and partial series or for series where the threshold is passed either every year or only in any proportion p of the years. Thus the general case also includes the special case which occurs when p = 1.

The distribution model for this case is obtained by analogy with the freeze threshold problem (Reference 3). Thus the general threshold series may be thought of as comprised of a mixture of two series: one comprising the series of years in which there were no threshold values, the other comprising the series where there were threshold values and these were further distributed according to date. The distribution associated with this mixed series is a mixed distribution, i.e., it is a distribution involving a mixture of the discrete series of years without thresholds and the series of relatively continuous threshold dates. If the distribution function F(t) represents the chance that a threshold T occurs before date t when only years with threshold dates are considered, the general distribution function of thresholds will be

$$G(t) = pF(t) + q(365) \qquad \ldots(1)$$

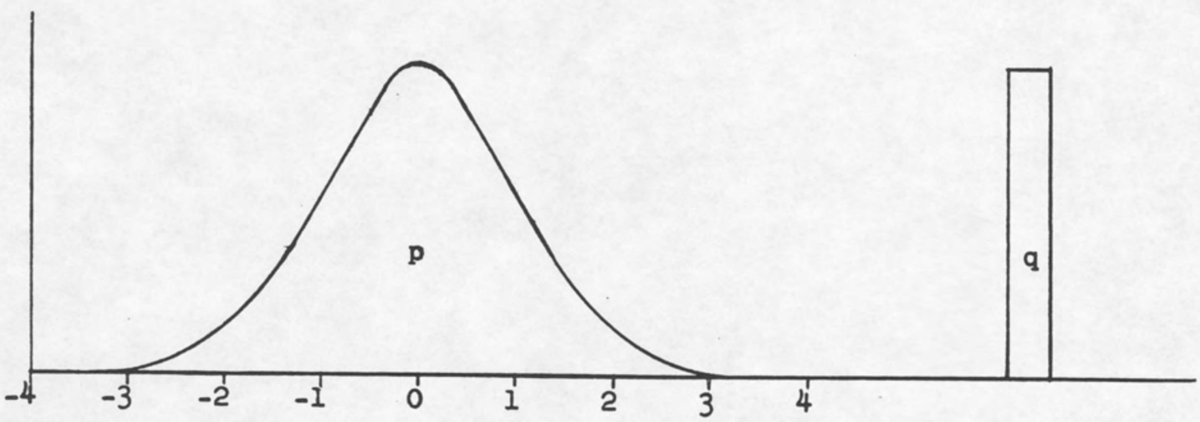

FIGURE 2

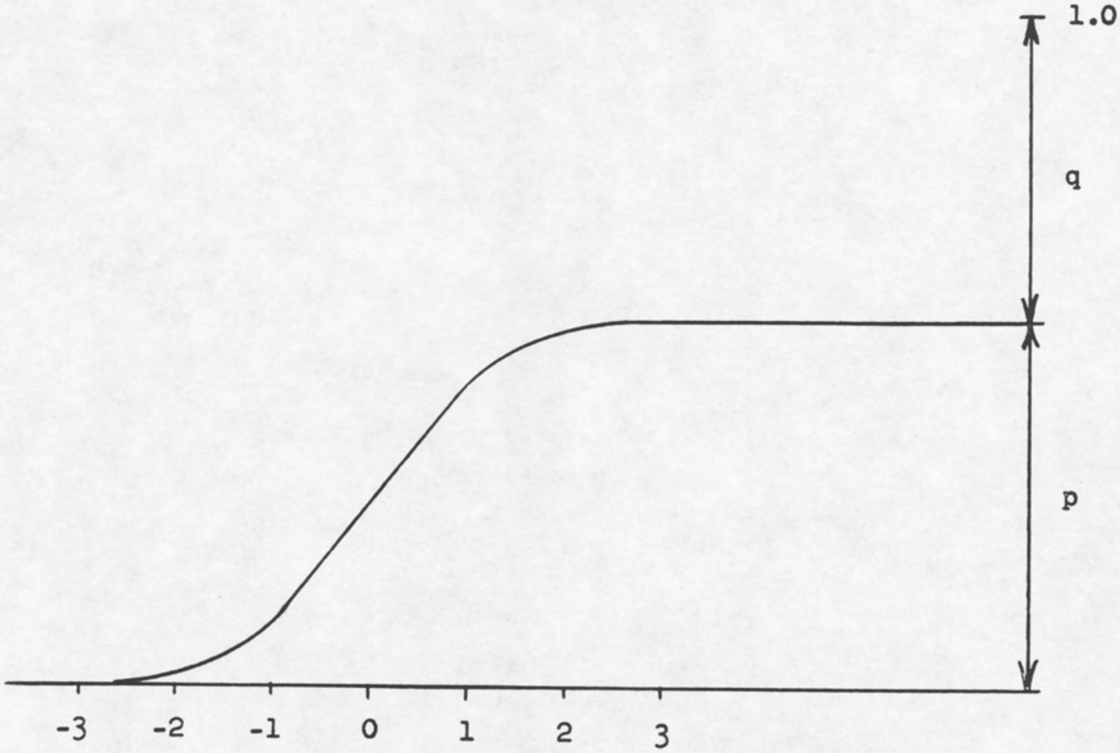

FIGURE 3

Here $G(t)$ is the probability of a threshold occurring before t, p is the probability of a year having a threshold, and $q(365) = 1 - p$ the probability of a year with no threshold. The latter is assumed to entail a series of thresholds occurring at the end of the season or in effect not occurring at all. Thus (1) meets the necessary condition for a distribution function that when t is near zero, $G(t)$ is also near zero while as t passes 365, $G(t) = 1$. The frequency distribution for this model is shown in Figure 1. Note that the probability q is concentrated far to the right after the distribution of threshold date on the left has ceased to have appreciable area under its right hand tail. The schematic representation of the integral of the frequency distribution or distribution function (1) is shown in Figure 2. Here it is to be noted that the ogive rises only to p instead of one and then rises vertically without respect to the abscissa the remaining distance q to one. Notice also that at this stage q need not be concentrated at a particular value as long as it is placed at a considerable time interval after all thresholds occur. This, it is clear, is tantamount to the non-occurrence of a threshold. Effectively we need to consider only the first term of (1) for the probability p of a threshold occurring is distributed over the season of threshold occurrence which is our main interest. It is seen that the first term of (1) is also a conditional distribution involving the probability that a threshold will occur and the distribution that probability over the relatively continuous variable, threshold date.

The Expected Threshold Date

By definition the mean or expected value of a mixed series is a sum of the mean values of the respective components weighted by their probabilities of occurrence. Before we can apply this to threshold data we must find the expected value of the series when no threshold occurs. Since there is no date associated with this occurrence, we must adopt a reasonable convention. If the year of threshold occurrence is assumed to begin on July 1, the last day or day number 365, seems to be a reasonable date to assign arbitrarily for all non-occurrences of thresholds. This seems reasonable, for occurrence at this date would for practical purposes be a non-occurrence since the probability of occurrence after this date has already decreased to a negligible value as shown on the continuous frequency curve of Figure 1. This convention gives

$$E(t) = p\,E\,(t/T) + 365q \qquad \ldots(2)$$

for the mean threshold date. Here $E(t)$ is the mean or expected date, $E(t/T)$ is the mean of threshold dates in the series, and p and q are as previously. Other conventions on q may be adopted for particular circumstances but the placing of q at 365 seems reasonable and illustrates the analysis easily.

Results of Analysis

The results of the entire analysis are shown in Table 1. As stated previously normal distributions were fitted to the coded dates of thresholds in all series. This gave $\bar{t}$ an estimate of the mean and s an estimate of the standard deviation. For p = 1 the threshold dates constitute the complete series and hence the quantiles may be obtained directly from a table of the normal distribution (Reference 6). These are shown in Table 1 for those T's where p = 1. For thresholds where p is less than one it is necessary to distribute only the probability p under the normal curve. Thus the quantiles shown in Table 1 must be computed from probabilities obtained by dividing the tabled probabilities by p. If these derived probabilities are greater than one, the threshold is beyond the date range of threshold occurrence; and hence is a non-occurrence. This accounts for the dashed spaces in Table 1 under the quantiles.

The mean value column $\bar{t}'$ is the result of applying equation (2) to the data in Table 1. Besides making it possible to have a mean for all thresholds, $\bar{t}'$ might have a practical interpretation where some variable in an application is a linear function of the date of occurrence of T. It is also the median or 0.50 probability value for the T's where p = 1.

The most important results of the analysis are of course the quantiles. These give the dates associated with the various probabilities of thresholds preceding these dates. Thus the probability of a day with 4 inches or more of snow occurring before November 14 at Rutland is 0.10 while at Pinkham Notch the date for the same probability is October 28, a date which is over two weeks earlier. Likewise for 0.90 probability the corresponding dates are January 26 for Rutland and December 27 for Pinkham Notch. Thus Pinkham Notch may be assured of a 4 inches or greater 24-hour snowfall with 0.90 probability a month earlier than Rutland. Numerous other similar interpretations may be made of the quantiles. Other quantiles may also be readily computed using the mean and standard deviation of Table 1 in conjunction with tables of the normal distribution.

Acknowledgments

The writer wishes to acknowledge the assistance of H. Landsberg in the selection of the stations studied and the help of R.G. Stone in the selection of the thresholds above one inch. For programming the computations on IBM equipment the writer is indebted to Daniel Harton of the National Weather Records Center.

References

(1) Reed, W.G. "The Probable Growing Season." Monthly Weather Review, Vol. 44, p. 509.

(2) Thom, H.C.S. and Shaw, R.H. "Freeze Hazard for Iowa." (Manuscript)

(3) Thom, H.C.S. "The Distributions of Freeze-date and Freeze-free Period When the Probability of Freeze is Less than One." Cornell University Biometrics Unit, BU-30-M, 1952.

(4) Geary, R.C. and Pearson, E.C. "Tests of Normality." Biometrika Office, London 1938.

(5) Thom, H.C.S. "Seasonal Degree-day Statistics for the United States." Monthly Weather Review, Vol. 80, p. 143.

(6) Hodgman, C.D., compiler, "Mathematical Tables from Handbook of Chemistry and Physics." 8th Ed. Chemical Rubber Pub. Co., Cleveland, 1947, p. 204.