DEPARTMENT OF COMMERCE
SINCLAIR WEEKS, Secretary

WEATHER BUREAU
F. W. REICHELDERFER, Chief

MONTHLY WEATHER REVIEW

Editor, JAMES E. CASKEY, JR.

Volume 82
Number 1

JANUARY 1954

Closed March 15, 1954
Issued April 15, 1954

THE RATIONAL RELATIONSHIP BETWEEN HEATING DEGREE DAYS AND TEMPERATURE [1]

H. C. S. THOM

Climatological Services Division, U. S. Weather Bureau, Washington 25, D. C.

[Manuscript received December 2, 1953]

ABSTRACT

The probability function of degree days below the base 65° F. is derived from the temperature probability function. Standard statistical analysis is applied to this function to obtain the relationship between mean degree days and mean temperature. This relationship is modified for use with available data and applied in the conversion of a monthly normal temperature for Detroit to the corresponding degree day normal.

INTRODUCTION

Almost from the time that heating degree days first came into use there has been a need for a rational relationship between temperature and degree-day statistics. The lack of such a relationship has always made it necessary to estimate degree-day means or normals from degree-day records which were often not available and tedious to compile. Temperature means, on the other hand, are already available for most stations and if not, are easy to compute from published data. Such a relationship makes degree-day statistics quickly available from any place with a temperature record. It also removes the difficulties associated with the lack of consistency between temperature and degree-day means which has been troublesome in the past. This has made it difficult to adjust degree-day means for a heterogeneous record. In the recent normals revision program of the Weather Bureau, for example, the usual arithmetical procedures could not be applied to obtain degree-day normals because of the numerous heterogeneities in the records at most stations. With a rational conversion formula available, properly adjusted temperature normals may be converted directly to degree-day normals with uniform consistency. More important than this use, perhaps, is the fact that the rational relationship is basic to the full development of the climatological analysis of degree-day data.

[1] Paper presented at 127th National Meeting of the American Meteorological Society, New York, N. Y., January 26, 1954.

289649—54——1

The study reported here is another phase [1] in the development of a general climatological analysis for degree days *below* a given base. With proper modification it may also be employed in the analysis of degree days *above* any base. The probability function of degree days derived here from the temperature distribution will form the basis for the later development of methods for obtaining degree-day probabilities.

THE TEMPERATURE FREQUENCY CURVE

In a previous paper [1] it was observed that the average temperatures of a particular day through a series of years have been found to have a normal probability or frequency function, or to be normally distributed. This probability function describes bell-shaped curves like those shown in figure 1 which are normal frequency curves on temperature scale t.

A normal probability function is known to be completely specified by its mean and standard deviation. The mean serves to locate the curve along the t axis while the standard deviation σ determines its scale, or how widely it is spread along the t axis. In figure 1 it is seen that both frequency curves are located by a mean temperature of 60° F. but have different scales or standard deviations. The curve with a standard deviation of 5.0 is spread out widely along the t axis while the curve with a standard deviation of 2.5 is more closely concentrated about the mean.

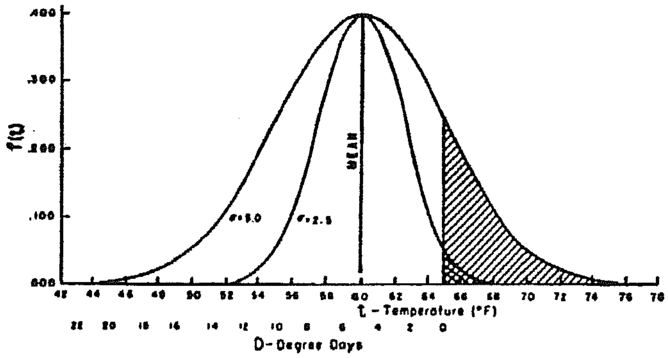

FIGURE 1.—Two examples of normal frequency curves, one for $\sigma=5.0$, the other for $\sigma=2.5$, both with mean temperatures of 60° F. Both the temperature scale t and its transformation by equation (1) to the degree-day scale D, are shown. As t is transformed to D, the distribution of t is transformed into the degree-day distribution. This is the unshaded portion under the temperature frequency curve distributed over the D scale together with an area of probability corresponding to the shaded portion concentrated at zero degree days. The entire distribution of degree days may be represented by the frequency curves shown in figure 2.

It is clear that as a result of these properties two changes may occur in the normal curve and hence in the distribution of temperature: (1) The mean may shift and move the curve to the left or right along the t axis, giving a location at a different value of t. (2) The scale or standard deviation may change causing the curve to spread out or become thinner. These changes are not statistically independent of each other but may be considered as separate component properties. An example of the first type of change is to move the curve $\sigma=2.5$ to the left two degrees of temperature, giving it a new mean of 58° but leaving the scale σ unchanged. The second type of change is represented in figure 1 by a change in scale from $\sigma=2.5$ to $\sigma=5.0$. This spreads the frequency curve without change in its location or mean. Also both types of change could occur together, giving a curve which is spread out as well as displaced along the t scale.

While the discussion of location and scale changes as climatic factors is a subject in itself, it will assist in our explanation of the degree-day distribution to have some understanding of climatic location and scale changes in the temperature distribution. The general principle observed over a wide range of climatic conditions is that the location of the temperature distribution increases as the scale decreases and conversely. This is in contrast to bounded elements such as precipitation where the location, as measured by the mean, varies directly as the scale. Since the location of the temperature distribution varies seasonally, as well as climatically, such variations are reflected in the seasonal march at a given station as well as from station to station for the same season.

The location and scale of the temperature distribution are best measured by the mean and standard deviation of the distribution. These parameters can therefore be related through the general principle. Although the variation of mean temperature with geographic position is not precise, there is, of course, a very marked tendency for it to decrease with increasing distance from the equator. Since the mean and standard deviation vary inversely, the standard deviation increases with increasing distance from the equator. In general then, the mean decreases with latitude while the standard deviation increases with latitude. Similarly in seasonal variation the mean is higher in summer and lower in winter and hence the standard deviation is lower in summer and higher in winter.

Large bodies of water have a great effect on the relation between location and scale of the temperature distribution. The pronounced effects of decreasing the rate of change of mean temperature with latitude and the narrowing of the range between summer and winter are well known. The effect on the standard deviation is even more pronounced. As a consequence, standard deviations are stabilized over extended areas along seacoasts and through the seasons in such areas. For example, the standard deviation for January along the east coast of the United States is almost uniform from Maine to Florida while in the interior it is three times larger in Minnesota than in Louisiana. Seasonal variation in the standard deviation is also smaller along the coasts, some stations having nearly the same standard deviation the year around. This occurs particularly along the west coast where the effect is more pronounced because of the prevailing winds off the ocean.

THE DEGREE-DAY FREQUENCY CURVE

These location and scale changes in the temperature frequency distribution produce corresponding changes in the associated degree-day distribution. They may be illustrated by transforming temperature to degree days by the well-known relationship

$$D=65-t, \quad D\geq 0 \tag{1}$$

where D is the degree-day value for a day and t is the day's average temperature in ° F. The inequality on the right is especially to be noted for it is an essential feature of the transformation which converts the t scale to the D scale of figure 1. As t is transformed to D, the distribution of t is transformed into the degree-day distribution. This is the unshaded portion under the temperature frequency curve distributed over the D scale together with an area or probability corresponding to the shaded portion concentrated at zero degree-days. Thus the probability of having degree days greater than zero on a particular day is equal to the unshaded portion below the temperature frequency curve and the probability of having zero degree days is the shaded portion. The manner in which these shaded and unshaded areas vary with the temperature distribution is clearly the key to the relation between temperature and degree-day statistics. Such variations may be interpreted in terms of the location and scale changes discussed above.

Since the degree-day base is fixed at 65° F., all location and scale changes occur in relation to it. With fixed scale or standard deviation, shifts in the mean produce

2

important changes in the size of the shaded area. As the mean temperature increases, the temperature frequency curve moves toward the right and the shaded area of the curve is increased while the unshaded area is decreased. This produces an increase in the probability of zero degree days and both a decrease in probability of degree days and an increased concentration of the probability at the lower degree-day values. The overall effect is to decrease the mean degree days. For a decrease in mean temperature the shaded portion of the curve decreases while the unshaded portion increases. This produces a decrease in the probability of zero degree days and an increased concentration of probability at higher degree days with a consequent increase in degree days. As the temperature mean moves to low values on the left, the amount of shaded area becomes negligible and the degree-day mean approaches $65-E(t)$ where $E(t)$ is the mean temperature. Thus, as has long been known, the degree-day mean increases as the temperature decreases and at low values is a function of the mean temperature alone. At higher values of mean temperature the shaded area becomes important and must be accounted for through use of both the mean and standard deviation since the size of the shaded area is a function of both parameters.

Variations in the degree-day mean produced by varying the temperature scale or standard deviation are not as easily depicted as those resulting from variation in the mean. With a fixed mean temperature, an increase in standard deviation increases the probability of zero degree days but also spreads the distribution to higher degree days. These changes have opposite effects on the degree-day mean so the effect of scale change is not a simple one and must be accounted for by an analytical relationship. Nevertheless, it is clear that changes in the temperature scale produce marked changes in the degree-day mean and hence must be accounted for in any relationship between degree days and temperature. As will be seen later, the scale or standard deviation is an important variable in the rational relationship.

THE PROBABILITY FUNCTION OF DEGREE DAYS

From the previous discussion it appears that the probability or frequency function of degree days consists of the portion of the temperature frequency curve below 65° and a probability concentrated at zero degree days equal to the probability of temperatures being above 65°. The former is the unshaded portion of the temperature frequency while the latter is equal to the shaded portion of the curve but concentrated at zero degree days. The unshaded portions of the frequency curves are truncated normal distributions which, when compounded with the probability densities at zero degree days, form mixed distributions which are the degree-day distributions. In sampling from such a distribution for a day on which zero degree days may occur, that day will have degree days greater than zero with a probability equal to the

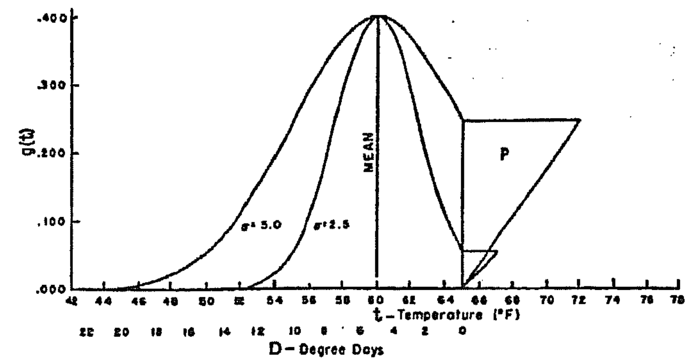

FIGURE 2.—Representation of the entire distribution of degree days, D, for two examples, corresponding to the two temperature frequency curves of figure 1 for which the standard deviations of temperature are 5.0 and 2.5, respectively, and mean temperature is 60° F. Note the area of probability, p, concentrated at zero degree days.

unshaded area of the frequency curve and zero degree days with a probability equal to the shaded portion of the curve. When degree-day values are greater than zero they will be further distributed according to the truncated probability function, the unshaded portion of the curve. They are not further distributed in the shaded portion of the curve, for here they always take the value zero.

The truncated normal distribution has been thoroughly investigated by several statisticians and most of the results we need have been reported in the literature (see [2, 3, 4, and 5]). There remains only to adapt the theory to cover the mixed distribution described above.

Let $F(t)$ be the normal distribution function of the average temperature for a day defined by

$$F(t)=\int_{-\infty}^{t} f(x)dx \tag{2}$$

where $f(x)$ is the normal probability function as shown in figure 1. Evidently $F(t)$ is the probability that an average temperature is less than t, and hence the probability that the average temperature is above the degree-day base is $p=1-F(65)$ and below the base is $q=F(65)$. Performing the transformation to degree days by equation (1), the distribution of degree days is

$$G(D|D\geq 0)=p+qF(65-t|t\leq 65) \tag{3}$$

where G gives the probability of less than D degree days and F is the normal distribution truncated at 65° (c. f. [2]). It will be noted that $G(0)=p$ which is the probability of the average temperature being 65° or greater, and hence is the probability of zero degree days. When $D\geq 0$, G is equal to p plus the probability of temperature being between 65° and some assigned lower value.

The probability function for degree days is the derivative of (3) which is

$$g(D|D\geq 0)=qf(65-t|t\leq 65). \tag{4}$$

This function is the equation of the unshaded portions of the curves in figure 1 referred to the D scale and is required in obtaining the mean value of D. The entire distribution may be represented by the frequency curves shown in figure 2.

THE RATIONAL RELATIONSHIP

The expected or mean value of degree days is defined in the usual manner by

$$E(D)=\int_0^{\infty} Dg(D)dD. \tag{5}$$

Applying this operation to the right hand side of equation (4) it is found that [2, 3]

$$E(D)=q[65-E(t)+\lambda\sigma]. \tag{6}$$

Here $E(t)$ is the mean temperature, σ is the standard deviation [1], and $\lambda=f(65)/F(65)$. Tables of the reciprocal of this function have been prepared by Pearson [4]. Assuming that t is normally distributed, (6) is the exact relationship between mean temperature and mean degree days. Since $E(t)$ and σ completely define the normal distribution which in turn determines q and λ, the mean value of D is easily found when $E(t)$ and σ are known. Values of F and f are given as functions of the argument $(t-E(t))/\sigma$ in any table of the normal probability function. q and λ are evaluated at $t=65$ and for convenience we designate $(65-E(t))/\sigma$ as h.

For $\sigma=5.0$ and $E(t)=60$ as shown in figure 1, it is seen that the base 65 is one standard deviation above the mean so, from tables of the normal distribution, $q=0.841$ and $\lambda=0.242/0.841=0.288$. Hence the degree-day mean for a day with $\sigma=5.0$ and $E(t)=60$ is

$$E(D)=.841[65-60+(0.288)5]=5.4.$$

APPLICATION OF THE RATIONAL RELATIONSHIP

The rational relationship applies to the means of daily degree days. However our interest is primarily in monthly means so the relationship will be adjusted to give these di-

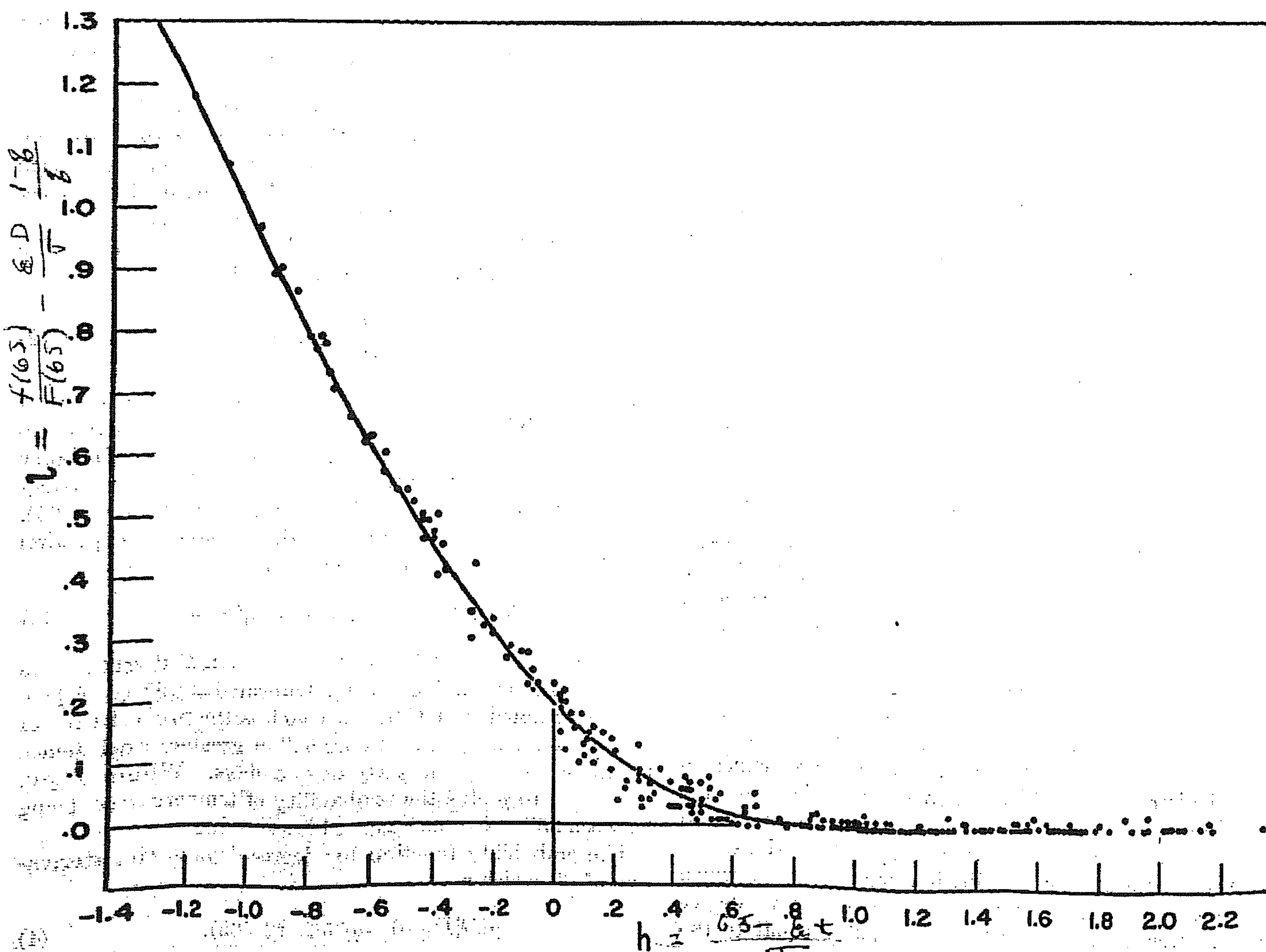

FIGURE 2.—The empirical relationship of l vs. h. The dots are observed values of l against h.

rectly. A simple way of doing this is to determine the relationship for a hypothetical average day of the month and multiply the resulting degree days by the number of days in the month. This average day is not a real day on which any particular average occurs, but a hypothetical day whose mean and standard deviation are such that when the conversion is made to degree days and the result multiplied by the number of days in the month the result is the mean degree days for the month.

In adjusting the relationship to obtain monthly statistics it was found convenient to use the standard deviation of monthly average temperature which is a function of the daily standard deviation and much easier to obtain. If σ is the standard deviation for the average day as above, σ_m the standard deviation of the monthly average, and r the mean correlation between all days for a month of N days, it may be shown [5] that

$$\sigma=\frac{\sqrt{N}\sigma_m}{\sqrt{1+(N-1)r}} = \sigma_m\sqrt{\frac{1}{\frac{1}{N}-\frac{r}{N}}}$$

Since the factor $(1+Nr)$ is not known but does not seem to vary greatly from station to station, we let it be accounted for in the overall proportional adjustment to the rational relationship by assuming

$$\sigma=\sqrt{N}\sigma_m. \tag{7}$$

Since σ for a single day is known only proportionally, q, which is a function of h, will also be known only proportionally. The approximation we need may be obtained by rearranging the rational relationship (6) in the form

$$\lambda-\frac{E(D)}{\sigma}\frac{(1-q)}{q}=\frac{E(D)-65+E(t)}{\sigma} \tag{8}$$

Substituting $\sqrt{N}\sigma_m$ for σ and l for the term on the left, we find

$$l=\frac{E(D)-65+E(t)}{\sqrt{N}\sigma_m} \tag{9}$$

Since all of the variables in (8) are functions of h, l will also be a function of h. Solving (9) for NE(D), the mean monthly degree days, gives

$$NE(D)=N(65-E(t)+l\sqrt{N}\sigma_m). \tag{10}$$

Next, l can be established as a function of h by plotting observed values of l against h. These values were computed from 30-year records at 30 stations representing all climatic conditions in the United States. The data which are for all of the 12 months are shown plotted on figure 3 together with the empirical l vs. h relationship. It is to be noted that the relationship is independent of climate and season and is only dependent on the parameters of the temperature frequency distribution. In this respect the l-function is general, like the λ-function, in that it is also dependent only on h. It is also similar in shape to the

TABLE 1.—*The factors h and l, for use in computing degree days from equation (10)*

h	l	h	l	h	l	h	l
−0.70	0.70	−0.32	0.39	0.05	0.17	0.42	0.05
−.69	.70	−.31	.38	.06	.17	.43	.05
−.68	.69	−.30	.38	.07	.16	.44	.04
−.67	.68	−.29	.37	.08	.16	.45	.04
−.66	.67	−.28	.36	.09	.15	.46	.04
−.65	.66	−.27	.36	.10	.15	.47	.04
−.64	.65	−.26	.35	.11	.14	.48	.04
−.63	.64	−.25	.34	.12	.14	.49	.03
−.62	.63	−.24	.34	.13	.13	.50	.03
−.61	.62	−.23	.33	.14	.13	.51	.03
−.60	.61	−.22	.32	.15	.13	.52	.03
−.59	.60	−.21	.32	.16	.12	.53	.03
−.58	.59	−.20	.31	.17	.12	.54	.03
−.57	.58	−.19	.30	.18	.11	.55	.03
−.56	.58	−.18	.30	.19	.11	.56	.02
−.55	.57	−.17	.29	.20	.11	.57	.02
−.54	.56	−.16	.29	.21	.10	.58	.02
−.53	.55	−.15	.28	.22	.10	.59	.02
−.52	.54	−.14	.27	.23	.10	.60	.02
−.51	.53	−.13	.27	.24	.09	.61	.02
−.50	.53	−.12	.26	.25	.09	.62	.02
−.49	.52	−.11	.25	.26	.09	.63	.02
−.48	.51	−.10	.25	.27	.08	.64	.02
−.47	.50	−.09	.24	.28	.08	.65	.01
−.46	.50	−.08	.24	.29	.08	.66	.01
−.45	.49	−.07	.23	.30	.07	.67	.01
−.44	.48	−.06	.23	.31	.07	.68	.01
−.43	.47	−.05	.22	.32	.07	.69	.01
−.42	.47	−.04	.22	.33	.07	.70	.01
−.41	.46	−.03	.21	.34	.06	.71	.01
−.40	.45	−.02	.20	.35	.06	.72	.01
−.39	.44	−.01	.20	.36	.06	.73	.01
−.38	.44	.00	.19	.37	.06	.74	.01
−.37	.43	.01	.19	.38	.06	.75	.01
−.36	.42	.02	.18	.39	.05	.76	.01
−.35	.41	.03	.18	.40	.05	.77	.01
−.34	.41	.04	.17	.41	.05	.78	.00
−.33	.40						

For $h \geq 0.78$, $l=0$
For $h \leq -0.70$, $l=-h$

λ-function and has analogous limiting properties, e. g., $l=-h$ for large values of $-h$, and $l=0$ for $h\geq 0.78$. Values read from figure 3 have been entered in table 1 for convenience in use.

In order to use (10) to compute normal monthly degree days, a set of manuscript charts has been prepared showing isolines of monthly standard deviations, s_m. Using the appropriate value of s_m and the normal value of the temperature, $\bar{t}$, as estimates of σ_m and $E(t)$, h may be readily calculated. Entering the table or graph with this value of h one finds the proper value of l. Substituting this together with $\bar{t}$ and s_m in (10) and multiplying by N, the number of days in the month, gives the degree-day normal $N\bar{D}$ a statistical estimate of $NE(D)$.

As an example, for September at Detroit we find the normal temperature $\bar{t}=64.3$ and the standard deviation $s_m=2.7$. Then h is easily found to be $(65-64.3)/(5.48)(2.7)=0.047$. For this value of h table 1 gives $l=0.17$ and hence $\sqrt{N}ls_m=(5.48)(2.7)(0.17)=2.51$. Substituting in (10) gives

$$N\bar{D}=30(65-64.3+2.51)=96$$

This is Detroit's degree-day normal for September.

REFERENCES

1. H. C. S. Thom, "Seasonal Degree-day Statistics for the United States", *Monthly Weather Review*, vol. 80, No. 9, Sept. 1952, pp. 143–149.

2. Harold Cramér, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, 1946, pp. 247–248.
3. R. A. Fisher, "Sampling Error of Estimated Deviates", etc., British Association, *Mathematical Tables*, vol. 1, 1931, p. xxxiii.
4. Karl Pearson (Editor), *Tables for Statisticians and Biometricians*, Part II, Cambridge University Press, London, 1931, pp. xxx and 11.
5. John F. Kenney, *Mathematics of Statistics*, Part II, D. Van Nostrand Co., New York, N. Y. 1939, p. 101.

6

MONTHLY WEATHER REVIEW

VOLUME 94 NUMBER 7
JULY 1966

CONTENTS

U.S. DEPARTMENT OF COMMERCE
ENVIRONMENTAL SCIENCE SERVICES ADMINISTRATION
WEATHER BUREAU

NORMAL DEGREE DAYS ABOVE ANY BASE BY THE UNIVERSAL TRUNCATION COEFFICIENT

H. C. S. THOM
Environmental Data Service, Environmental Science Services Administration, Washington, D.C.

ABSTRACT

Equations are developed for obtaining mean monthly degree days above any base from mean monthly temperature and standard deviation of monthly average temperature. By the use of data for all months for twelve widely scattered stations and four bases it is shown that the truncation coefficient for degree days below any base with proper modification of the argument also applies for degree days above any base. This is also proved analytically, which leads to some further aspects of the universality of the truncation coefficient. Two formulas for the coefficient are also developed.

1. INTRODUCTION

The study reported here is the third phase in the development of the general climatological analysis of degree days [1], [2], [3]. Reference [2] established the rational relationship between mean monthly degree days below 65° F. and temperature and gave a table of the truncation coefficient l. Reference [3] showed that the rational relationship also applied to mean monthly degree days below any base using the same table of l. Later it was noted [4] that a slightly modified relationship using the same table of l also gave mean monthly degree days above any base. Thus the table of the truncation coefficient proved to be universal, applying to mean monthly degree days below or above any base. The evidence for the final steps in establishing universality has never been given although the method has been used extensively in this country and Canada [5], [6]. It is the purpose of this paper to give this evidence as well as an analytical form for the universal truncation coefficient useful in computer applications.

Degree days above particular bases, although not as yet used as extensively as degree days below a base, are of growing importance in horticulture and in air conditioning requirements and power consumption estimations. Horticulturists use bases ordinarily between 40° and 50° F. in systems for estimating growth progress and harvest dates. The literature on this application is extensive, of which reference [5] is a good example. Application to air conditioning requirement and power consumption has been much less extensive and even less has been published. The key paper is the one by Marston [6]. Indications are that there will be an increase in the use of degree days in this area.

2. DISTRIBUTION FUNCTION AND EXPECTED VALUE

It was shown previously that the degree day distribution describes a mixed population of degree day values equal to zero and greater than zero. This arises from the definition of the degree day; a particular value of which is the number of degrees of temperature above (or below) a fixed base temperature. Thus the temperature distribution truncated at the base temperature transformed to degree days, the continuous part of the distribution, and the truncated portion, the probability of zero degree days, form the mixed distribution of degree days. For degree days above a given base b the transformation from temperature to degree days is

$$D=t-b;\quad (D\geqq 0). \tag{1}$$

where t is ordinarily the average temperature for a day.

The truncated probability density function for temperature may be expressed by

$$f(t|b\leq t)=\frac{f(t)}{\int_b^\infty f(t)dt}=\frac{f(t)}{1-F(b)}, \tag{2}$$

where F is the distribution function of t, and the probability density function has the value given by (2) on the interval $b\leqq t<\infty$, and zero elsewhere. If the transformation (1) is applied to equation (2) in the usual fashion, the result is the probability density function of degree days

$$g(D|D\geq 0)=\frac{f(b+D)}{1-F(b)}\frac{dt}{dD}=\frac{f(b+D)}{1-F(b)}. \tag{3}$$

Integrating this over the open interval $0<D<\infty$ gives the distribution function of degree days greater than zero

$$G(D|D>0)=\frac{1}{1-F(b)}\int_{0+}^{D} f(b+D)dD=\frac{F(b+D)-F(b)}{1-F(b)}. \tag{4}$$

Multiplying by $1-F(b)$ and adding $F(b)$ gives the desired distribution function on the closed interval $0\leqq D<\infty$, i.e., including the zero values of degree days.

8

As in the first work on the rational relationship between degree days and temperature the analysis is performed on a hypothetical middle day of a month. The average temperature on this day is assumed to have a normal distribution whose mean and standard deviation are such that when the conversion is made to degree days, and the result multiplied by the number of days in the month, the result is the mean degree days for the month [2]. The normal probability density function is designated by φ and the distribution function by Φ.

Let the standardized variable of temperature be $z=[t-E(t)]/\sigma$ where $E(t)$ and σ are the population mean and standard deviation and the truncation point is z_0. Then the probability density function for the truncated normal distribution according to equation (2) may be expressed as

$$\varphi(z|z_0\leq z)=\frac{1}{\sqrt{2\pi}}e^{-z^2/2}/[1-\Phi(z_0)]. \tag{5}$$

Hence the mean of this distribution is given by

$$E(z|z_0<z)=\frac{1}{\sqrt{2\pi}[1-\Phi(z_0)]}\int_{z_0}^{\infty} ze^{-z^2/2}dz. \tag{6}$$

To evaluate the integral it is only necessary to make the substitutions $u=-z^2/2$ and $dz=-du/z$ whence

$$-\int e^u du=-e^u=-e^{-z^2/2}.$$

Substitution in (6) and evaluation of the integral between z_0 and ∞ yields the reciprocal Mill's ratio

$$E(z|z_0<z)=\frac{e^{-z_0^2/2}}{\sqrt{2\pi}[1-\Phi(z_0)]}=\frac{\varphi(z_0)}{1-\Phi(z_0)}=\lambda_*(z_0) \tag{7}$$

where the inferior star indicates truncation on the left of the distribution. To return to the variable t it is only necessary to take the mean of $t=z+E(t)$ over the truncated distribution giving

$$E(t|b<t)=\sigma E(z|z_0<z)+E(t) \tag{8}$$

which on substituting (7) yields

$$E(t|b<t)=\sigma\lambda_*(z_0)+E(t). \tag{9}$$

The mean number of degree days greater than zero is found by taking the expected value of $D=t-b$, giving $E(D|D<0)=E(t|b<t)-b$ which on substitution of equation (9) gives

$$E(D|D>0)=\sigma\lambda_*(z_0)+E(t)-b. \tag{10}$$

The mean of degree days for the mixed population of zero and non-zero degree days is the weighted mean of these components or

$$E(D|D\geq 0)=\Phi(z_0)\cdot 0+[1-\Phi(z_0)]E(D|D>0).$$

Substituting from (10) gives

$$E(D|D\geq 0)=[1-\Phi(z_0)][\sigma\lambda_*(z_0)+E(t)-b] \tag{11}$$

which is the theoretical relationship between mean temperature and degree days for a middle day with mean $E(t)$ and standard deviation σ. Unfortunately, estimates of σ and therefore of z_0 are not available; so, as for degree days below a base [2], an approximation of σ must be employed.

3. GENERAL DEGREE DAY FORMULA

With σ unavailable it is necessary to make some adjustment to equation (11) which makes computation possible. The most suitable procedure was found to be to follow the method used for degree days below a base, i.e., to solve as much as possible for $\lambda_*(z_0)$ and associated functions of z_0 which are not known.

Rearranging the terms in equation (11) and writing $E(D)$ for $E(D|D\geq 0)$ yields

$$\lambda_*(z_0)-\frac{E(D)}{\sigma}\left[\frac{\Phi(z_0)}{1-\Phi(z_0)}\right]=\frac{E(D)-[E(t)-b]}{\sigma}. \tag{12}$$

As with degree days below a base, the left hand side is set equal to a new truncation coefficient Λ_* after a modification of the right hand side to take care of the fact that no direct estimate is available for σ. Let σ_m be the standard deviation of monthly average temperature and $\bar{\rho}$ the mean correlation between all possible pairs of N days of a month; then

$$\sigma=\sqrt{N}\sigma_m/[1+(N-1)\bar{\rho}]^{1/2}. \tag{13}$$

The factor $[1+(N-1)\bar{\rho}]^{1/2}$ is unknown because $\bar{\rho}$ is unknown, but call it k nevertheless so (13) becomes

$$\sigma=\sqrt{N}\sigma_m/k. \tag{14}$$

In order to standardize the argument on which Λ_* is dependent, $E(t)$, b, and $\sqrt{N}\sigma_m$ are combined into a single term to make the standardized truncation point

$$-x_0=(\bar{t}-b)/(\sqrt{N}\sigma_m). \tag{15}$$

This was $-h$ of the previous paper [2]. Now since k is unknown, replace σ on the right of (12) by $\sqrt{N}\sigma_m$ and let the factor k divide the term on the left. Finally replace the left hand term by $\Lambda_*(x_0)$ so that

$$\Lambda_*(x_0)=\frac{E(D)-[E(t)-b]}{\sqrt{N}\sigma_m}. \tag{16}$$

This is the population value of the truncation coefficient for degree days above a base b. Solving for $E(D)$ and multiplying by N to get the monthly mean degree days above b gives

$$NE(D)=N[\Lambda_*(x_0)\sqrt{N}\sigma_m+E(t)-b]. \tag{17}$$

9

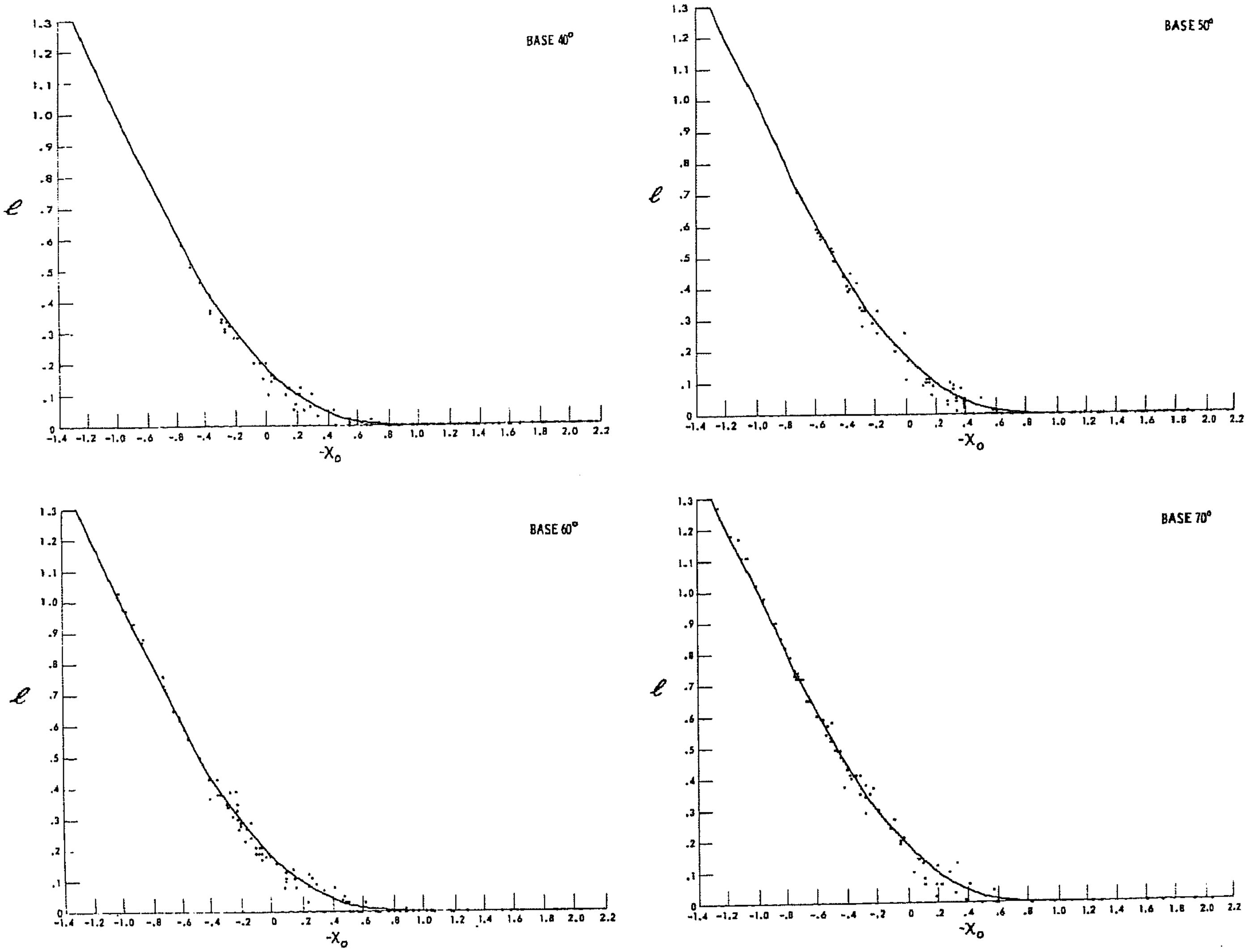

FIGURE 1.—l vs. x_0 data for various bases. The l-curve for degree days below 65° F. is superimposed on data for above various bases to show universality of the l-curve.

The sample estimates of these two equations are

$$l_*(x_0)=\frac{\overline{D}-(\bar{t}-b)}{\sqrt{N}s_m} \tag{18}$$

and

$$N\overline{D}=N[l_*(x_0)\sqrt{N}s_m+\bar{t}-b]. \tag{19}$$

It was now immediately conjectured that l plotted against $-x_0$ or $-h$ would produce an l curve independent of the base and identical with the l curve previously established for degree days below any base [3].

To show the universality of the truncation coefficient l_*, values were computed using equation (17) on degree day means above four bases for all months having degree days at 12 widely scattered stations. These are shown plotted against $-x_0$ in figure 1. The l-curve established for degree days below 65° F. [2] and found to hold for degree days below any base [3] was then superimposed on the data for the four bases in figure 1. The fit is equally as good as found for degree days below a base. This completes the empirical demonstration of the universality of the truncation coefficient for mean monthly degree days above or below any base. The demonstration will be made analytically in the next section.

It is clear from the above that if the previous h is set equal to x_0, representing the truncation coefficient for degrees below a base as $l^*(x_0)$ then empirically at least

$$l_*(x_0)=l^*(-x_0). \tag{20}$$

Substituting in (19) gives

$$N\overline{D}=N[l^*(-x_0)\sqrt{N}s_m+\bar{t}-b]. \tag{21}$$

Writing $\overline{D}^{+}$ for degree days above a base and $\overline{D}^{-}$ for degree days below a base generalizes (21) to

$$N\overline{D}^{\pm}=N[l^{*}(\mp x_0)\sqrt{N}s_m\pm(\bar{t}-b)] \tag{22}$$

which covers all situations.

4. UNIVERSAL TRUNCATION COEFFICIENT

In this section Λ^* and Λ_* are always a function of x_0 and λ^* and λ_* is still a function of z_0.

For degree days above a base there are two expressions for $E(D)$ given by equation (11) and an inversion of (16). This gives

$$\Lambda_*\sqrt{N}\sigma_m+(E(t)-b)=(1-\Phi)[\lambda_*\sigma+(E(t)-b)]. \tag{23}$$

Dividing by $\sqrt{N}\sigma_m$, substituting the value of k from equation (14), and rearranging terms yield finally

$$\Lambda_*=(1-\Phi)(\lambda^*/k)+\Phi x_0. \tag{24}$$

For degree days below a base [2] there are analogous expressions for $E(D)$ which when set equal give

$$\Lambda^*\sqrt{N}\sigma_m-(E(t)-b)=\Phi[\lambda^*\sigma-(E(t)-b)]. \tag{25}$$

Again dividing by $\sqrt{N}\sigma_m$, etc., as above, yields finally

$$\Lambda^*=\Phi(\lambda^*/k)-(1-\Phi)x_0. \tag{26}$$

If equations (23) and (25) are divided by σ instead of $\sqrt{N}\sigma_m$ using, of course, $z_0=[b-E(t)]/\sigma$ and manipulations similar to those above, there results

$$\Lambda_*=[(1-\Phi)\lambda_*+\Phi z_0]/k \tag{27}$$

and

$$\Lambda^*=[\Phi\lambda^*-(1-\Phi)z_0]/k. \tag{28}$$

Setting the value of Λ_* from (24) equal to that of (28) gives

$$z_0=kx_0 \tag{29}$$

which is also clear from the definitions of x_0 and z_0.

The basic equations (24) and (26) may be transformed by recalling that $\lambda^*=\varphi/\Phi$ and $\lambda_*=\varphi/(1-\Phi)$. Substituting yields

$$\Lambda^*=\varphi/k-(1-\Phi)x_0 \tag{30}$$

and

$$\Lambda_*=\varphi/k+\Phi x_0. \tag{31}$$

Since $\varphi(-x_0)=\varphi(x_0)$, φ is not affected by a change of sign of its argument. It is noted from equation (29) that k must also be an even function of x_0, thus the first terms of (30) and (31) are not affected by a change in sign of x_0. Returning for the moment to explicit expressions for $\Lambda^*(x_0)$ and $\Phi(x_0)$ and substituting $-x_0$ for x_0 in (31) yield

$$\Lambda^*(-x_0)=\varphi/k-[1-\Phi(-x_0)](-x_0). \tag{32}$$

Recalling that $\Phi(-x_0)=1-\Phi(x_0)$ and making this substitution in (32) give

$$\Lambda^*(-x_0)=\varphi/k+\Phi x_0. \tag{33}$$

But this is identical with (31); hence

$$\Lambda^*(-x_0)=\Lambda_*(x_0) \tag{34}$$

which demonstrates the universality of the truncation coefficient. Starting from equation (31) and following similar operations give

$$\Lambda_*(-x_0)=\Lambda^*(x_0). \tag{35}$$

There are a number of other symmetrical relations which are interesting: Subtracting equation (30) from equation (33) yields the relation

$$\Lambda^*(-x_0)-\Lambda^*(x_0)=x_0. \tag{36}$$

Likewise, following similar operations or simply substituting (34) and (35) in (36) gives

$$\Lambda_*(x_0)-\Lambda_*(-x_0)=x_0. \tag{37}$$

These relations indicate the fundamental properties of the truncation function which assist in establishing its analytical form.

5. ANALYTICAL FORMS OF THE TRUNCATION CURVE

The truncation curve is not a very simple function as can be seen from the previous development. Since for practical applications it need only be known to two significant figures, it seemed reasonable to fit a curve to the l-table given in [2] taking into account the symmetry properties of the previous section.

None of the functional forms related to Mill's ratio proved to be of much help. Finally, it was found that the sum of two exponentials gave a very satisfactory result. Fitting to the original l-table gave the following pair of equations:

$$l^*(x_0)=0.34e^{-4.7x_0}-0.15e^{-7.8x_0} \tag{38}$$

and by (36)

$$l^*(-x_0)=l^*(x_0)+x_0.$$

These equations smoothed the l-table slightly. Departures from the unsmoothed table are not greater than 0.01.

It appeared to be of interest to relate the truncation function to Mill's ratio. It is necessary now to use x_0 as the independent variable for all functions. Solving equation (30) for k yields

$$k(x_0)=\frac{\varphi(x_0)}{\Lambda^*(x_0)+[1-\Phi(x_0)]x_0}. \tag{39}$$

Recalling that $\varphi(x_0)$ is an even function and substituting $-x_0$ for x_0 give

$$k(-x_0)=\frac{\varphi(x_0)}{\Lambda^*(x_0)+[1-\Phi(x_0)]x_0}. \tag{40}$$

11

Mill's ratio is defined for this purpose as

$$R(x_0)=[1-\Phi(x_0)]/\varphi(x_0). \qquad (41)$$

Solving for $[1-\Phi(x_0)]$ and substituting in (39) give

$$k(x_0)=\frac{\varphi(x_0)}{\Lambda^*(x_0)+\varphi(x_0)R(x_0)x_0}. \qquad (42)$$

Since it is required to fit $k(x_0)$ for both positive and negative values of x_0, it is necessary to have a formula for $k(-x_0)$. Solving (41) for $\Phi(x_0)$ and substituting in (40) yield

$$k(-x_0)=\frac{\varphi(x_0)}{\Lambda^*(-x_0)+\varphi(x_0)R(x_0)-1}. \qquad (43)$$

A series of values of $l^*(x_0)$ and $l^*(-x_0)$, $x_0=h$, for each tenth between 1.00 and -2.00 was obtained from the l-table of [2]. Values of $\varphi(x_0)$ and $R(x_0)$ were found in tables II and III in reference [7]. When these values are substituted in equations (42) and (43) a series of $k(x_0)$ values is obtained. Note that the positive value of x_0 is always used in $R(x_0)$.

Examination of equations (13) and (14) suggests that it might be more interesting to determine the equation for k^2 instead of k since $k^2=1+(N-1)\bar{\rho}$. The series of values obtained from equations (42) and (43) were therefore squared before being fitted as a function of x_0.

After a new series was formed by subtracting one from each k^2 a functional form $k^2-1=y=a\cos^n\theta$ was intuitively suggested. If $\tan\theta=x_0$, $\theta=\tan^{-1}x_0$, for $x_0=-2.0$, -1.0, and 0, $\theta=-1.11$, -0.785, and 0 radians, hence $\cos\theta=0.4474$, 0.7071, and 1. Since k^2-1 is about 3.410 at $\theta=0.785$ radians, $1.326=3.410(0.7071)^n$ and $n=2.73$. (This will incidentally be very close to the final value.) An approximation to the equation is then $y=3.410\cos^{2.73}\theta$.

On substituting $\theta=\tan^{-1}x_0$ and by simple trigonometry we find $y=3.410[1+x_0^2]^{-1.36}$. The general form of the equation is then

$$k^2-1=a[1+x_0^2]^{-m}. \qquad (44)$$

The logarithmic form of this was fitted by least squares giving finally

$$k^2(x_0)=1+3.44(1+x_0^2)^{-1.35}. \qquad (45)$$

The fit of this to the l-table was very good, for the correlation between the logarithms was $r^2=0.9897$, leaving only about 1 percent of the variance unexplained by equation (45). With the k-function in analytical form a second method of computing l^* using Mill's ratio is available.

REFERENCES

1. H. C. S. Thom, "Seasonal Degree-Day Statistics for the United States," *Monthly Weather Review*, vol. 80, No. 9, Sept. 1952, pp. 143–149.
2. H. C. S. Thom, "The Rational Relationship between Heating Degree Days and Temperature," *Monthly Weather Review*, vol. 82, No. 1, Jan. 1954, pp. 1–6.
3. H. C. S. Thom, "Normal Degree Days Below Any Base," *Monthly Weather Review*, vol. 82, No. 5, May 1954, pp. 111–115.
4. H. C. S. Thom, "Standard Deviation of Monthly Average Temperature," *U.S. National Atlas*, 1955, pp. 1–123.
5. R. M. Holmes and G. W. Robertson, "Heat Units and Crop Growth," *Publication 1042*, Canada Department of Agriculture, 1959, 31 pp.
6. A. D. Marston, "Degree Days for Summer Air Conditioning," Kansas City Power and Light Company, Report, 1937.
7. K. Pearson (editor), *Tables for Statisticians and Biometricians*, Part II, Cambridge University Press, London, 1931, pp. 2–10 and 11–15.

[*Received April 27, 1966*]

12