Estimating Winter Design Temperatures from Daily Minimum Temperatures

NOLAN J. DOESKEN AND THOMAS B. MCKEE

Department of Atmospheric Science, Colorado State University, Fort Collins, CO 80523

(Manuscript received 23 September 1982, in final form 14 July 1983)

ABSTRACT

A methodology has been developed to estimate winter design temperatures (temperatures exceeded a specific number of hours during the December through February winter season—an important design parameter in building construction) from synthetic distributions of hourly temperatures for locations where only daily maximum and minimum temperatures are observed. Cumulative distributions of hourly temperatures and daily minimum temperatures were examined at seven different locations in Colorado having 10 or more consecutive years of complete hourly data. A consistent relationship between the two distributions was found for these stations by representing the lower half of each distribution with a best-fit power curve and relating the fitting coefficients. From these relationships an equation was derived that generated the shape of the lower half of the cumulative distribution of hourly temperatures. The only required input parameters are the regression coefficients resulting from the power curve fitting of the observed distribution of daily minimum temperatures.

The method was tested in Colorado stations having both hourly and daily temperature data. Excellent results were obtained for Colorado. Synthesized temperatures at probabilities of up to 0.50 were generally within 0.7°C of the observed values. The method has now been employed to calculate winter design temperatures for dozens of Colorado cities where such information has previously been unavailable.

1. Introduction

Hourly temperature information has an assortment of important uses and applications. Design temperatures—temperatures which are exceeded climatically a specific percentage of the time (a specific number of hours)—are used extensively to assist in building design, in furnace sizing, and to answer other energy-related design and construction questions. Many other design questions, heating and cooling load projections, and evapotranspiration rate calculations may make use of climatic distributions of hourly temperatures.

The need for and the use of hourly temperature data has increased rapidly. However, the availability of representative data often falls short of the needs. For example, in Colorado and other areas having complex terrain, most locations are climatically different from the few sites where hourly temperatures are recorded. To make matters worse, from 1964 to about 1980 only data at 3 h intervals have been digitized routinely by the National Climatic Data Center (NCDC), and the number of stations for which data are being digitized has been reduced. Few useful summaries of hourly data are published, and most users must have substantial computer resources in order to work with raw unsummarized data.

Fortunately, there is a vast resource of readily available daily maximum and minimum temperature data from thousands of locations across the United States. Previous studies have related certain aspects of daily maximum and minimum temperature distributions with hourly distributions and extremes. A graphical method was developed to estimate design temperatures by relating counts of hourly temperatures above or below specific probability thresholds to distributions of the annual extreme maximum and minimum temperature (Crow, 1963). Monthly mean temperatures, together with corresponding mean daily maximum and minimum temperatures, have been used to estimate probabilities of surface temperature extremes (Tattelman and Kantor, 1977).

This paper presents a new methodology for estimating winter design temperatures (normally defined as the temperature at a site that is not exceeded some small percentage of hours—generally 1%—during the three-month December–February winter season). A method is developed by defining the relationship between the bottom half of the cumulative distribution of daily minimum temperatures and the corresponding distribution of hourly temperatures.

2. Data

For the purposes of this investigation, all stations in Colorado having 15 years (1950–64) of digitized hourly temperature data 24 observations per day, were assembled. Only five stations met this requirement. Two additional stations, Pueblo and Trinidad, were also included although only 10 years of hourly data were available. A working data set was constructed by merging daily maximum and minimum temperatures from the National Climate Data Center (NCDC) TD-9727 daily data set with the corresponding 24 hourly temperature values extracted from the TD-1440 data set. Checks were performed to make sure that no hourly data fell outside the limits defined by the daily maximum and minimum.

A map (Fig. 1) shows by boxes the locations and approximate elevations of these seven stations. Significant geographic and climatic differences exist between some of the sites. For example, Table 1 shows the mean monthly temperatures and diurnal ranges, calculated from calendar-day maximum and minimum temperatures for July and January, each varied by 7°C or more among the seven stations.

TABLE 1. Station comparison (based on 1950–64 data).

		Temperature characteristics (°C)			
Station	Elevation (m)	July mean	July average diurnal range	January mean	January average diurnal range
Colorado Springs	1873	21.6	16.0	−1.4	15.7
Denver	1610	22.9	16.1	−1.3	14.8
Eagle	1980	18.7	23.4	−8.1	18.3
Grand Junction	1480	25.8	16.2	−2.9	11.3
La Junta	1279	26.0	16.9	−0.9	16.5
Pueblo	1414	24.5	17.1	−1.0	17.2
Trinidad	1751	23.2	17.0	+0.2	16.3

3. Development of methodology

The winter design temperature is by definition a point on the cumulative distribution function (CDF) for hourly temperature. An example of the empirical CDF for daily maximum, minimum and hourly temperatures for Eagle, Colorado is given in Fig. 2. In this section, a procedure is developed to estimate a point on the low-probability portion of the hourly curve from a knowledge of the cumulative curve of daily minimum temperatures.

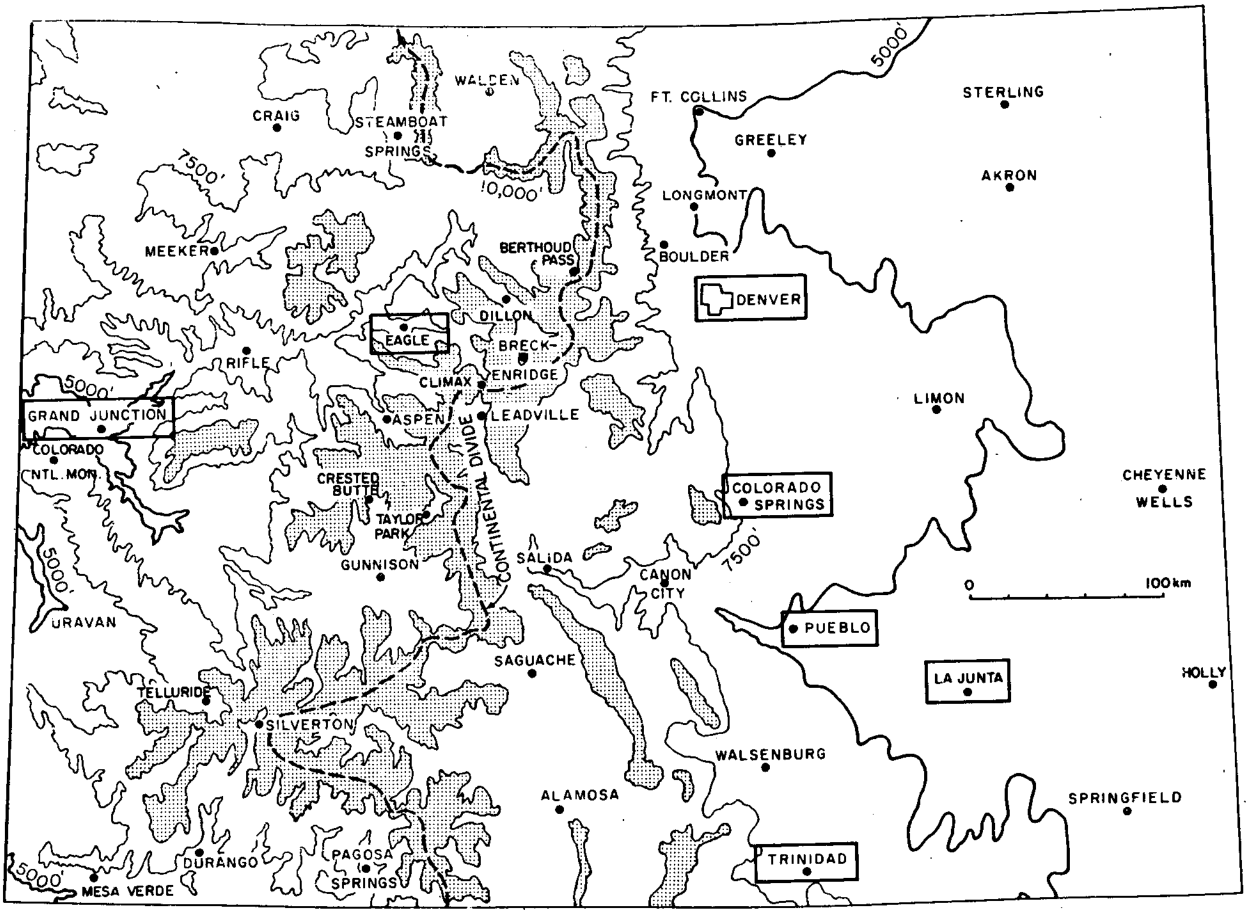

FIG. 1. Colorado weather stations (enclosed in boxes) used in the development of hourly-daily temperature distribution relationships. Elevation contours at 2500 ft (762 m) intervals.

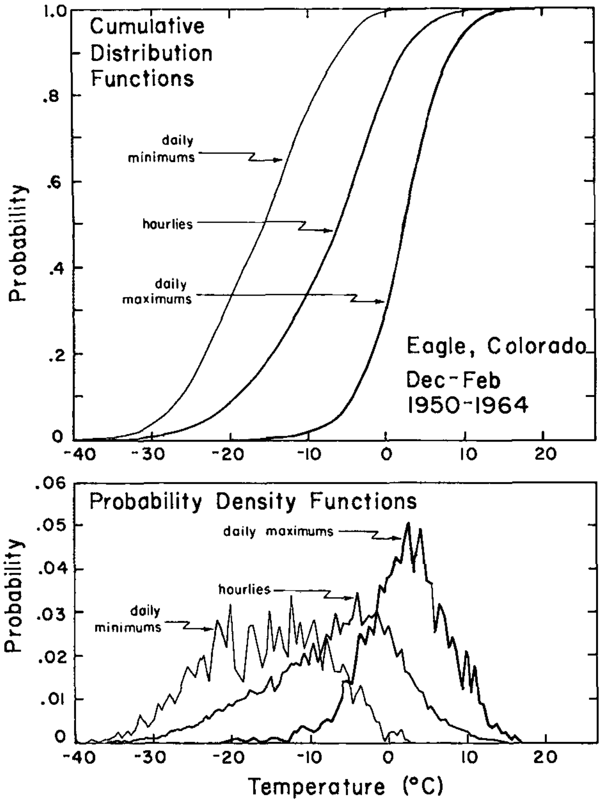

FIG. 2. Winter (December–February) probability density functions and the corresponding cumulative distribution functions for daily minimum, hourly and daily maximum temperatures at Eagle, Colorado.

Three obvious characteristics of the cumulative distributions of hourly and daily minimum temperatures form the physical basis for this analysis. First, the two distributions share approximately the same origin (T_0) for each station; that is, the coldest hourly temperatures always approach the coldest daily minimum temperature in any sample. Second, since the daily minimum temperature is always equal to or colder than any hourly temperature that day, the slope of the CDF of daily minimum temperatures is greater than the slope of the hourly temperature CDF throughout the lower half of the distributions as is seen in Fig. 2. Third, the shape of each empirical CDF is very similar, suggesting the feasibility for quantitatively comparing the bottom half of each distribution by comparing simple mathematical representations of each curve.

a. Procedure

The following experimental procedure was developed to generate winter design temperatures from distributions of daily minimum temperatures.

1) Divide the year into four 3-month seasons. Separate the 3-month December–February winter season.

2) For each station generate tabular and graphical histograms, empirical probability density functions (PDF's), and empirical cumulative distribution functions (CDF's) for daily minimum and hourly temperatures. Visually examine similarities and differences between curves.

3) Find a functional form which can be used to fit the lower half of each empirical CDF. Functional form should be as simple as possible and should fit especially well near the end of the distribution (for design temperature application).

4) Apply curve fitting techniques to determine fitting coefficients for each empirical CDF for all seven Colorado stations.

5) Determine the relationship between fitting coefficients of the daily minimum CDF and the hourly CDF curves.

6) Express equation for bottom half of hourly CDF in terms of fitting coefficients for the daily minimum CDF curves.

7) Calculate "synthesized" hourly CDF's and compare results to hourly empirical CDF's based on actual data.

The year was divided into four 3-month seasons. Only the December–February winter season was selected for study here. This was done for two reasons: 1) to match the 3-month period from which winter design temperatures have traditionally been calculated (ASHRAE, 1972), and 2) to examine the season having the greatest geographic, day to day and year to year variability. Hence, it should be a "worst case" example for studying temperature distribution characteristics.

Distributions of hourly temperatures and daily minima were generated for each of the seven Colorado stations. Graphs of the empirical PDF's and the associated empirical CDF's were prepared for each station similar to Fig. 2.

b. Curve fitting

Several functional forms were explored which resembled the shape of the lower half of the empirical CDF. Attempts to fit CDF's of hourly and minimum temperatures using polynomial and exponential functions yielded encouraging results. The best fit, particularly on the low end of the distribution, was obtained using a power function of the form

$$\left.\begin{aligned} P &= a(T - T_0)^b \\ \text{or} \quad \ln P &= \ln a + b \ln(T - T_0) \end{aligned}\right\} \quad \text{for} \quad T > T_0, \quad (1)$$

where a and b are coefficients, P is the probability of nonexceedance, T the temperature at that probability level, and T_0 the reference origin temperature. According to this expression, P could equal 0 if T were allowed to drop to T_0.

The initial curve fitting was done using a least-squares fitting routine for power curves. The curve fitting equations were

$$b = \frac{\Sigma(\ln x_i)(\ln y_i) - \dfrac{(\Sigma \ln x_i)(\Sigma \ln y_i)}{n}}{\Sigma(\ln x_i)^2 - \dfrac{(\Sigma \ln x_i)^2}{n}}, \tag{2}$$

$$a = \exp\left[\frac{\Sigma \ln y_i}{n} - b\frac{\Sigma \ln x_i}{n}\right], \tag{3}$$

$$r^2 = \frac{\left[\Sigma(\ln x_i)(\ln y_i) - \dfrac{(\Sigma \ln x_i)(\Sigma \ln y_i)}{n}\right]^2}{\left[\Sigma(\ln x_i)^2 - \dfrac{(\Sigma \ln x_i)^2}{n}\right]\left[\Sigma(\ln y_i)^2 - \dfrac{(\Sigma \ln y_i)^2}{n}\right]}, \tag{4}$$

where $x = T - T_0$ and $y = P$. The regression coefficients are a and b. The coefficient of determination r^2 indicates the quality of fit achieved by the regression.

c. Sensitivity tests

A variety of tests were performed to obtain the best possible fits to the empirical CDF curves and to test the sensitivity of these fits to T_0 and to the range of points used in the fit. Values of T_0 were chosen ranging from much below the coldest observed temperatures in the sample to T_0 corresponding to a nonzero probability of about 0.005. Varying numbers of temperature—probability data pairs along the CDF between probability levels of 0.001 and 0.50 were tried, and the quality of the resulting fits were compared. Starting points with temperatures corresponding to probability levels from near 0 to 0.01 and ending points with temperatures corresponding to probability levels from 0.30 to 0.50 were used.

The results of curve fitting experiments conducted on all seven Colorado stations showed that the selection of a T_0 from 3 to 8°C colder than the coldest temperature in the sample gave the best fit over the full probability range up to 0.50 both for minimum temperature CDF's and for hourly CDF's. However, the overall quality of the fit was not terribly sensitive to T_0 and very good fits could be obtained over a wide range of T_0's. Fits to the hourly CDF's were consistently better than to minimum CDF's because the hourly curves were based on 24 times more data points and, hence, were much smoother. In general, all of the experimental fitted curves matched the shape of the actual CDF curves very well.

The selection of the range of points used in fitting the CDF curves was found to be significant, especially for fitting the minimum temperature CDF curves. Use of the entire range from probability levels of near zero up to 0.50 did not yield the best results. The best fit over the bottom half of the curve was actually obtained when only the portion between probability levels of 0.006 and 0.40 was used in the fitting. The best fit over the probability range 0.005 to 0.100, the portion most critical for determination of minimum design temperatures, was obtained when fitting was performed using only the portion of the curve from 0.007 to 0.30. The improvement in the overall quality of the curve fits (particularly the fits to the minimum temperature CDF's), when the starting input points were changed from probability levels of near 0.0 up to 0.006 or 0.007, resulted from omitting those few points on the extreme lower tail of the distribution. Those few points often did not match the general shape of the distribution, especially when less than 15 years of data were used (see Fig. 3).

d. Application to Colorado

A consistent set of curve fits was generated for both the hourly and minimum temperature CDF's for all seven Colorado stations. Since the primary problem is to determine the temperature at the 0.01 probability level, the empirical CDF data for probabilities from 0.007 to 0.30 were input into the curve fitting equations. A constant T_0 of −37.2°C was chosen for all stations corresponding to the average value for Colorado of the 50-year return period extreme annual minimum temperature (Nicodemus and Guttman, 1980).

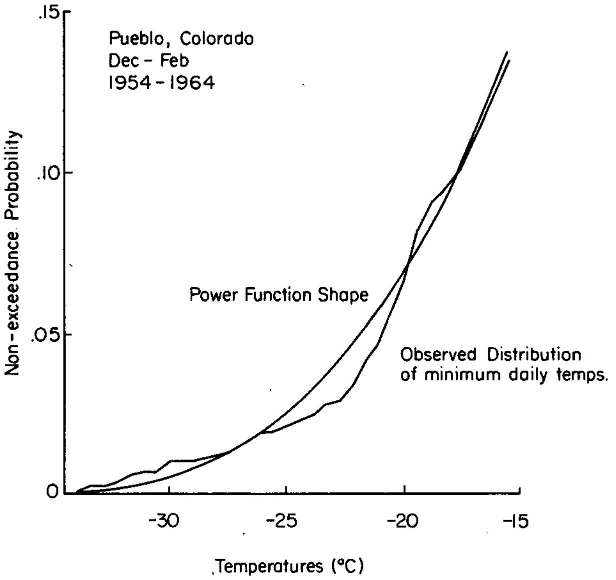

FIG. 3. Lower end of cumulative distribution function of observed minimum daily winter (December–February) temperatures at Pueblo, Colorado, 1954–64, compared with power function shape. Points near the extreme end of curve often do not fit basic power curve shape well, especially when less than 15 years of data are used.

TABLE 2. Power curve fitting coefficients [fit to cumulative distributions of winter (December–February) temperatures] using $T_0 = -37.2°\mathrm{C}$ and a fitting range of 0.007–0.30 probability level.

		Minimum temperature CDF				Hourly temperature CDF			
Station	Years	a_m	$\ln a_m$	b_m	r^2	a_h	$\ln a_h$	b_h	r^2
Colorado Springs	1950–64	0.168×10^{-7}	−17.9	4.28	0.991	0.251×10^{-9}	−22.1	5.12	0.999
Denver	1950–64	0.496×10^{-7}	−16.8	4.00	0.997	0.139×10^{-8}	−20.4	4.69	0.999
Eagle	1950–64	0.175×10^{-3}	−8.7	2.15	0.984	0.143×10^{-5}	−13.5	3.20	0.999
Grand Junction	1950–64	0.948×10^{-11}	−25.4	6.17	0.993	0.376×10^{-13}	−30.9	7.29	0.999
La Junta	1950–64	0.372×10^{-7}	−17.1	4.06	0.997	0.131×10^{-9}	−22.8	5.26	0.999
Pueblo	1954–64	0.299×10^{-5}	−12.7	2.93	0.967	0.315×10^{-8}	−19.6	4.51	0.998
Trinidad	1950–59	0.159×10^{-9}	−22.6	5.47	0.997	0.102×10^{-11}	−27.6	6.44	0.999

A set of constants, a and b, resulted for each power curve fit. The a's, b's, and the coefficients of determination r^2 are shown in Table 2.

The next step was to show quantitatively the form of the relationship between the lower halves of the CDF's of hourly and daily minimum temperatures. The fitted exponent b_m for the daily minimum distribution was plotted against the exponent b_h fit to the hourly distribution for each of the seven Colorado stations. The resulting graph (Fig. 4b) showed a strikingly linear relationship. The same was done with the natural logarithm of the coefficient a for both the hourly and minimum distributions a_h and a_m. A similar linear relationship was observed (Fig. 4a). This behavior results from the dependent interactions of the two coefficients and from the consistent relationship between corresponding pairs of hourly and minimum temperature CDF's. This linear relationship is obtained only when the same T_0 has been assigned to all distributions. Hence, a constant T_0 is a necessary condition for this analysis.

Linear regression, performed on the seven data points in each graph, resulted in the following relationships:

$$\ln a_h = 5.3 + 0.99 \ln a_m, \quad r^2 = 0.96, \tag{5}$$

$$b_h = 1.3 + 0.95 b_m, \quad r^2 = 0.96. \tag{6}$$

From these relationships, "synthesized" distributions of the bottom half of hourly temperature CDF's could be calculated. Given the fitting coefficients a_m and b_m obtained from the observed cumulative distribution of winter minimum temperatures at a station, a_h and b_h could be calculated using the above equations. These values then defined the lower half of the "synthesized" hourly temperature CDF by substitution into the equation:

$$\ln P = \ln a_h + b_h \ln(T - T_0). \tag{7}$$

Synthesized hourly temperature CDF's were generated for the seven Colorado stations, and these distributions were compared to the actual observed hourly distributions. Fig. 5 shows a comparison of synthesized and actual observed hourly CDF's for Colorado Springs. Distributions compared extremely well over most of the probability range up to 0.50 for all seven stations.

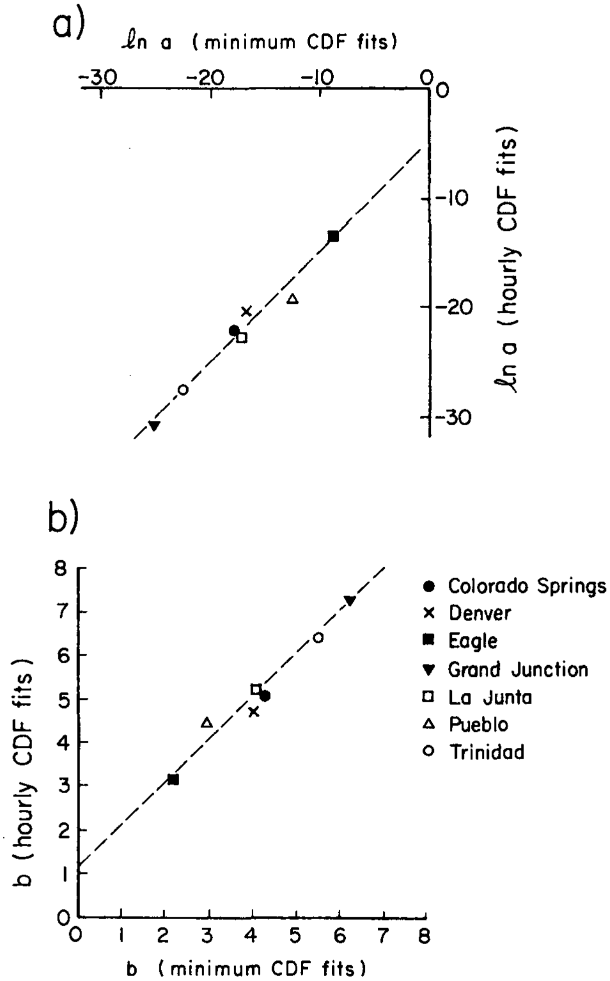

FIG. 4. Relationships between power curve function fitting coefficients, (a) $\ln a$ and (b) b, fit to cumulative distribution functions of hourly and daily minimum temperatures. Results are based on December–February data at seven Colorado stations.

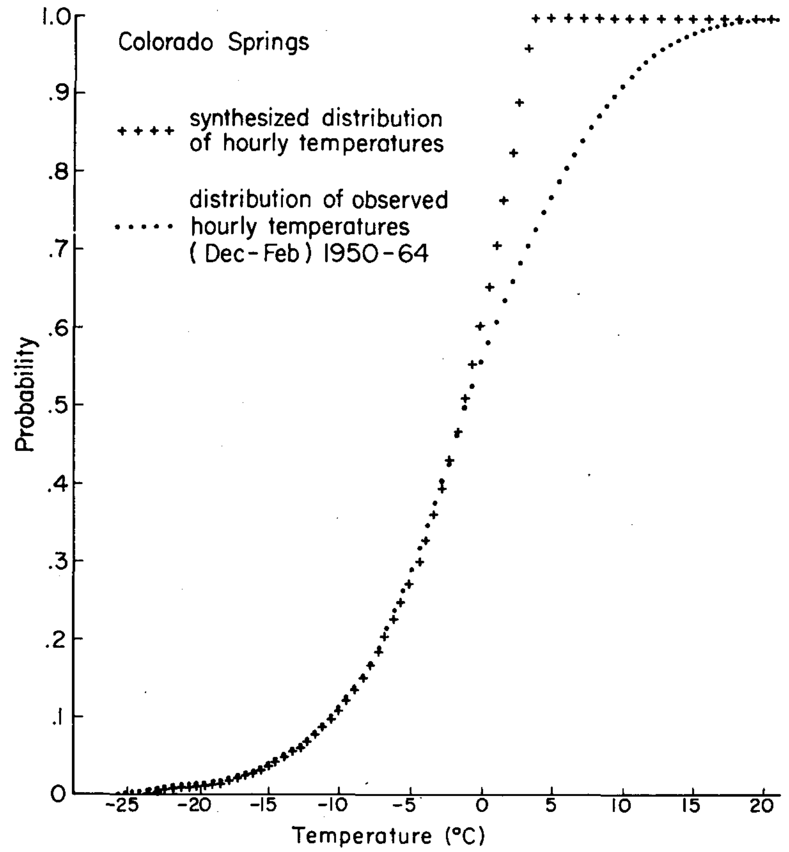

FIG. 5. Comparison of power curve synthetic distribution for bottom half (probability of 0.0–0.5) to observed cumulative distribution function of hourly temperatures for Colorado Springs, December–February, 1950–64.

Equally encouraging results were obtained using a slightly different approach. Regression relationships were developed based on hourly and minimum fitting coefficients for all combinations of six of the seven Colorado stations. These relationships were then used to calculate synthesized hourly CDF's for the seventh station. The synthesized distributions were then compared to the actual observed hourly CDF's. This yielded an independent test of the validity of the method on the seventh Colorado station.

The results of this test are shown in Table 3. Temperatures, synthesized and observed, were compared at several probability levels, including the 0.01 and 0.025 levels (corresponding to the 99 and 97.5% win-

TABLE 3. Comparison of temperature (°C) at specified probability levels on observed and synthesized hourly CDF's.

	Probability							
	0.01		0.025		0.10		0.25	
Station	Obs.	Syn.	Obs.	Syn.	Obs.	Syn.	Obs.	Syn.
Colorado Springs	−20.3	−19.8	−16.8	−16.6	−10.5	−10.4	−5.8	−5.4
Denver	−21.1	−20.4	−17.6	−17.1	−10.7	−10.7	−5.4	−5.4
Eagle	−28.3	−28.2	−25.4	−25.4	−19.2	−19.2	−12.8	−13.5
Grand Junction	−16.8	−16.5	−13.8	−13.7	−9.2	−8.6	−5.3	−4.7
La Junta	−19.9	−20.0	−16.3	−16.6	−10.2	−10.2	−5.3	−4.9
Pueblo	−22.0	−22.6	−18.2	−18.9	−11.4	−11.4	−6.3	−4.8
Trinidad	−17.1	−17.8	−14.6	−14.8	−8.8	−9.4	−4.4	−5.2

Obs. ≡ actual observed distribution.
Syn. ≡ synthesized distribution.

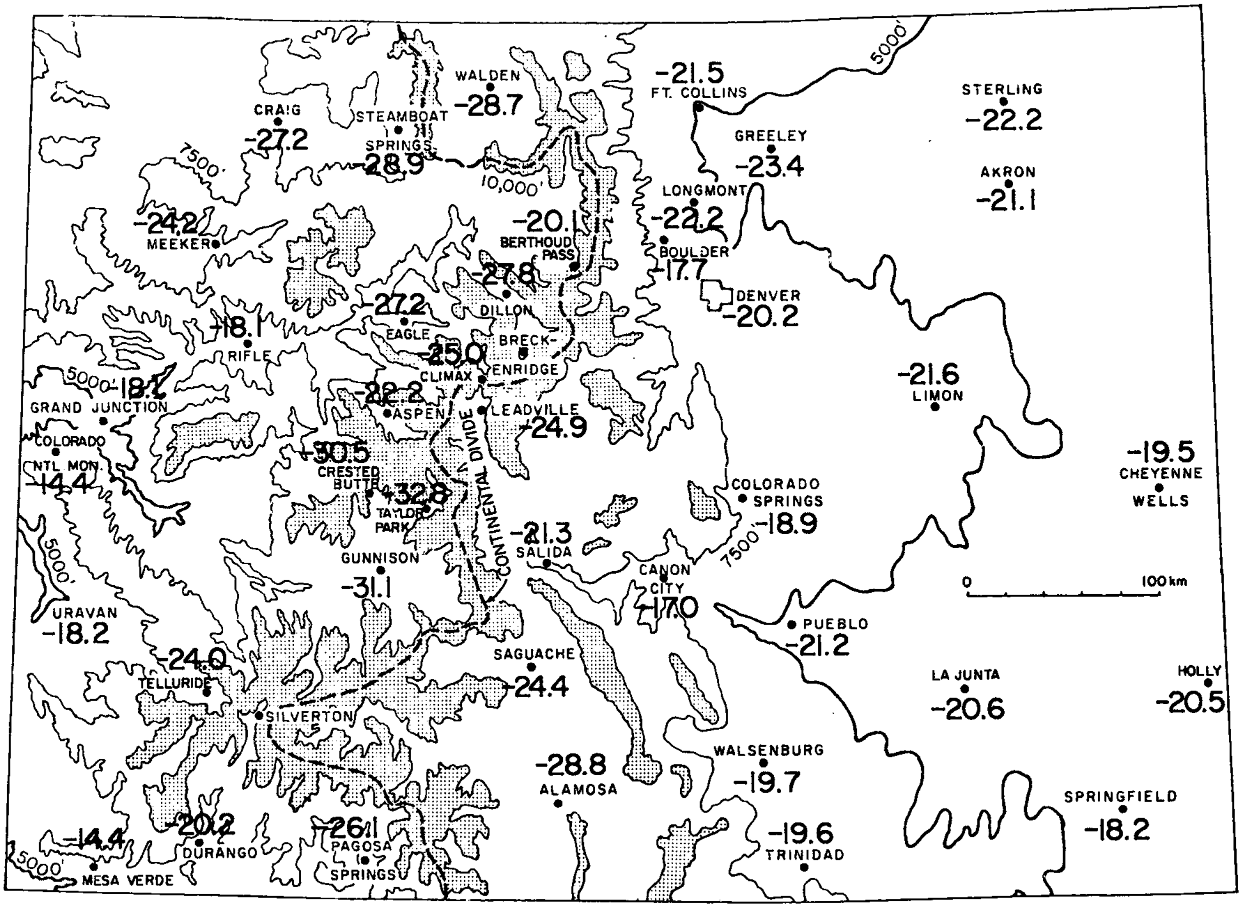

FIG. 6. Winter 99% design temperature (temperature at the 0.01 probability level) for selected Colorado locations based on synthesized hourly temperature distributions. Units: °C.

ter design temperatures used by ASHRAE). The method performed very well, particularly over the probability range 0.01–0.20, but acceptably well over the entire lower half of the CDF for Colorado stations. Synthesized temperatures at a given probability level were usually within 0.7°C of the observed value.

4. Results for Colorado

Using the methodology described here, hourly probability distributions and design temperatures have been generated for 157 Colorado stations based on 1951–1980 daily minimum temperatures for the winter months (December–February). The 99% winter design temperatures (temperatures which are not exceeded on the average, 1% of the total hours in December, January and February) for selected Colorado locations are shown in Fig. 6.

The results show considerable variation in design temperatures across the state, sometimes over relatively short horizontal distances. The total range of calculated values of the 99% winter design temperature in Colorado went from −14.4°C at Colorado National Monument (near Grand Junction) to −32.8°C at Taylor Park, a high mountain valley near Gunnison.

Design temperatures as a function of elevation for all 157 stations are shown in Fig. 7. Design temperatures decrease with elevation in a general sense at a rate of about 12°C km^{-1} (determined by linear regression with design temperature selected as the independent variable). However, the regression line explains only 34% of the variation. Other topographic factors have considerable significance. The lower left-hand edge of the distribution represents basin and valley bottom locations which are consistent cold air trapping sites. The upper right-hand edge includes well-mixed, thoroughly drained sites such as mountain passes, ridge tops and windswept valley locations which are not susceptible to the formation of temperature inversions. All other points lie somewhere between and indicate the varying degree of cold air trapping and draining at each site.

Finally, Table 4 presents a comparison between ASHRAE-generated design temperatures and current results generated by the Colorado Climate Center

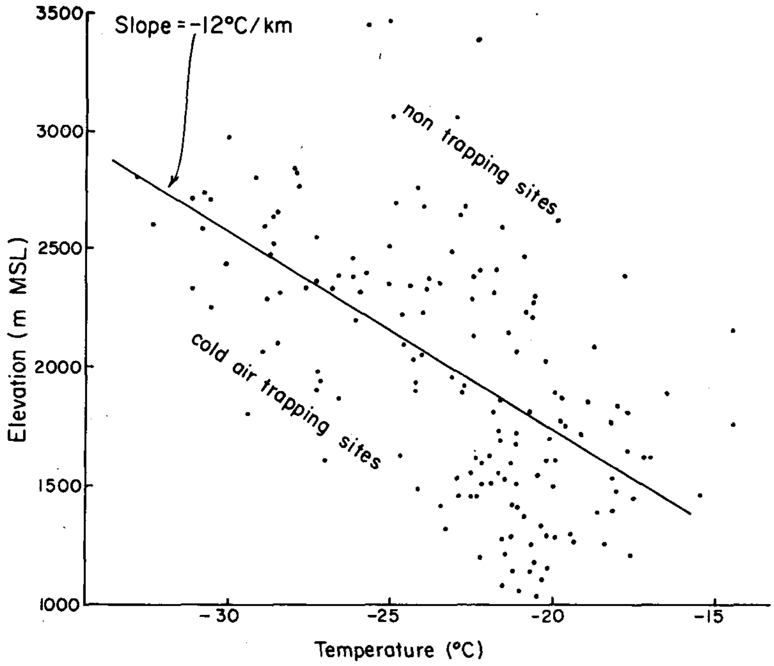

FIG. 7. Winter 99% design temperature (temperature at the 0.01 probability level) as a function of elevation based on synthesized hourly temperature distributions at 157 locations in Colorado. Units: °C.

(CCC). Differences of 1°C or more are frequently noted. Differences between successive editions of ASHRAE were significantly larger than differences between CCC and the individual ASHRAE editions. The most similar sets of results were those of the 1976 ASHRAE Rocky Mountain Chapter and the current CCC calculations. For these data sets, the average difference in calculated design temperatures was 1.2°C. While it is impossible to determine the "correct" results without actual hourly data, the results on the CCC method appear consistent and realistic. We place a high degree of confidence in these results.

TABLE 4. Comparison of previously published 99% winter design temperatures (°C) with results generated by the Colorado Climate Center.

Station	ASHRAE (1972)	ASHRAE (1976)	ASHRAE (1977)	CCC (1983)
Alamosa	−27.2	−29.4	−23.9	−28.8
Boulder	−15.6	−16.1	−21.1	−17.7
Colorado Springs	−18.3	−18.3	−19.4	−18.9
Denver airport	−18.9	−19.4	−20.6	−20.2
Durango	−17.8	−18.3	−21.1	−20.2
Fort Collins	−22.8	−22.8	−20.6	−21.5
Grand Junction	−13.3	−13.9	−16.7	−18.1
Greeley	−22.8	−23.3	−18.9	−23.4
La Junta	−21.1	−20.0	−19.4	−20.6
Leadville	−22.8	−23.3	−27.8	−24.9
Pueblo airport	−20.6	−21.1	−21.7	−21.2
Sterling	−21.1	−21.1	−21.7	−22.2
Trinidad airport	−17.2	−21.1	−18.9	−19.6

5. Further applications

There is growing interest among engineers, architects and builders, to have design temperature and hourly temperature distribution information available monthly as well as seasonally. This method could quickly be used to generate that information for all stations having a sufficient record length of daily temperatures.

The upper half of the empirical CDF of hourly temperatures is related in a similar way to the upper half of the cumulative distribution of daily maximum temperatures (Fig. 2). Therefore, this method should also be tested as a means of generating probabilities of warm temperature extremes. Combining results for each half separately, we find it reasonable to approximate the entire distribution of hourly temperatures based on minimum and maximum temperatures.

The nationwide applicability of this method has also been examined. Daily data from Seattle, Minneapolis, Boston, Houston and St. Louis were used to generate hourly distributions (using the relationship derived from Colorado). The results were compared to actual hourly distributions.

Results were not as consistent as in Colorado. Differences between synthesized and actual CDF's at specific probability levels were as much as 3°C. However, considering that a constant $T_0 = -37.2$°C was used for the entire country (actual coldest minimum temperatures for the stations used ranged from −8.3°C at Houston, Texas to −47.8°C at Eagle, Colorado) and considering the great differences in temperature characteristics nationwide, these results are very good. Minor adjustments based on regional climatic differences appear to be all that will be required to make this technique usable for the entire country.

Acknowledgments. This research was funded by the National Climate Program office, NOAA, under Grant NA80AA-D-00118. The idea was initially developed with the support of the Colorado State University Experiment Station.

REFERENCES

American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc. (ASHRAE), 1972: Weather data and design fundamentals. *ASHRAE Handbook of Fundamentals*, Chapt. 33, 667–688.

——, 1976: *Climate Data for Air Conditioning Design, Rocky Mountain Chapter Region, Colorado, Wyoming, Montana and Environs*. Rocky Mountain Chapter ASHRAE, 1st ed. Centennial.

——, 1977: Weather data and design conditions. *1977 Fundamentals Handbook*, Chapt. 23, 23.1–23.22.

Crow, L. W., 1963: *Study of Weather Design Conditions for American Society of Heating, Refrigerating, and Air Conditioning Engineers, Inc.* Res. Proj. No. 23, LWC Rep. No. 40, 52 pp. [Loren Crow Consultants, 2422 S. Downing St., Denver, CO 80210].

Nicodemus, M. L., and N. B. Guttman, 1980: *Probability Estimates of Temperature Extremes for the Contiguous United States*. NUREG/CR-1390. Prepared by the National Climatic Center for the U.S. Nuclear Regulatory Commission.

Tattelman, P. I., and A. J. Kantor, 1977: A method for determining probabilities of surface temperature extremes. *J. Appl. Meteor.*, **16**, 1175–1181.