Estimating Summer Design Temperatures from Daily Maximum Temperatures in New Mexico

KENNETH E. KUNKEL

New Mexico Department of Agriculture, Las Cruces, NM 88003

(Manuscript received 29 July 1985, in final form 19 October 1985)

ABSTRACT

Many climatological locations report only maximum and minimum temperatures. However, in certain applications, such as estimation of design temperatures, the frequency distribution of hourly temperatures is required. For this reason, a method is developed for estimating the mathematical form of the upper half of the cumulative probability distribution function (CDF) for hourly temperatures from the CDF for daily maximum temperatures for the summer months of June, July and August. In this method, an exponential function is fitted to the daily maximum CDF. A procedure for estimating the hourly CDF from the daily CDF is presented. This method is used to estimate summer design temperatures for a number of stations in New Mexico.

1. Introduction

Hourly temperature information has a variety of applications. One important area is in the construction industry. Design temperatures—temperatures which are exceeded climatically a specific percentage of time (specific number of hours)—are used to assist in estimating the size of heating and cooling systems, in building design, and other related questions. Unfortunately, the availability of hourly temperature data is quite limited. In New Mexico, only a few locations have long-term records of hourly temperatures. However, there are a large number of locations with long term measurements of daily maximum and minimum temperatures. Therefore, methods for estimating hourly temperature distributions from daily maximum and minimum temperature distributions are of widespread potential usefulness.

Doesken and McKee (1983; hereafter referred to as DM) developed a method for estimating winter design temperatures from the cumulative probability distribution function (CDF) of daily minimum temperatures. They found that the lower half of the CDF of both hourly temperatures and daily minimum temperatures could be fit by a power function of the form:

$$P = a(T - T_0)^b \quad (1)$$

where P is the probability of nonexceedance, T is the temperature at that probability level, T_0 is the reference origin temperature, and a and b are coefficients. They also found that a linear relationship existed between the a and b coefficients for the hourly distribution and the a and b coefficients for the daily minimum temperatures. Therefore, for stations without hourly observations, the CDF of daily minimum temperature could be used to estimate the a and b coefficients for hourly temperatures. This synthesized distribution could then be used to estimate winter design temperatures.

In this paper, an approach similar to DM is used to construct a method for estimating summer design temperatures from daily maximum temperatures. In some of the warmer parts of New Mexico, the cooling of buildings may be required for up to six months of the year, contributing a substantial amount to the annual energy costs. Therefore, it is important for buildings to be designed for efficient cooling. This climatic feature provides the impetus for this study.

2. Data

The data used for this analysis were obtained from the National Climatic Data Center in Asheville, NC. Hourly temperatures were extracted from the TD-1440 data set while daily maximum temperatures were obtained from the TD-9727 data set. The primary period of analysis was 1950–64. Six stations were chosen for analysis. These are listed in Table 1 along with climatic temperature data. Their locations are shown in Fig. 1. Note that several of the stations do not have a complete record during this time. In addition, hourly temperatures were also available for Albuquerque for 1973–78. These data were also included in the analysis as if it were a separate station.

3. Analysis

As DM illustrated, the shapes of the CDFs of daily maximum and minimum temperatures and hourly temperatures are qualitatively similar in winter. Figure 2 shows the CDFs for daily maximum temperatures and for hourly temperatures for the summer months

TABLE 1. Station information.

Station	Elevation (m)	Period of record	Temperature (June–Aug) Mean maximum (°C)	Mean minimum (°C)
Albuquerque	1623	1950–64	32.7	16.6
Albuquerque	1623	1973–78	32.7	16.6
Farmington	1675	1953–64	32.7	11.9
Las Vegas	2093	1950–64	27.4	11.0
Truth or Consequences	1469	1951–64	33.1	17.8
Tucumcari	1234	1950–54, 1962–64	33.1	17.6
Zuni	1924	1950–64	29.7	11.1

of June, July and August for Albuquerque. The ordinate is a normal probability scale. The hourly CDF appears to be bimodal with a somewhat different shape than the daily CDF. However, when considering only the upper halves of the CDFs, the shapes are similar, with the hourly CDF exhibiting a slightly larger degree of curvature on this scale. The general similarity of shape suggests that a simple mathematical relationship between the two distributions may exist.

CDFs for both daily maximum temperatures and hourly temperatures were calculated for each of the seven locations. In these calculations, only data from the summer months of June, July and August were used. Several mathematical forms, including polynomials, power laws, and exponentials were tested to determine an optimum fitting function to the top half of the CDFs. While the power law form used by DM fit the distributions well, a better fit in all cases was obtained using an exponential. Specifically, the following mathematical form was found to be optimum:

$$1 - P = e^{-a(T_0 - T)^b} \tag{2}$$

or

$$\ln[-\ln(1 - P)] = \ln a + b \ln(T_0 - T). \tag{3}$$

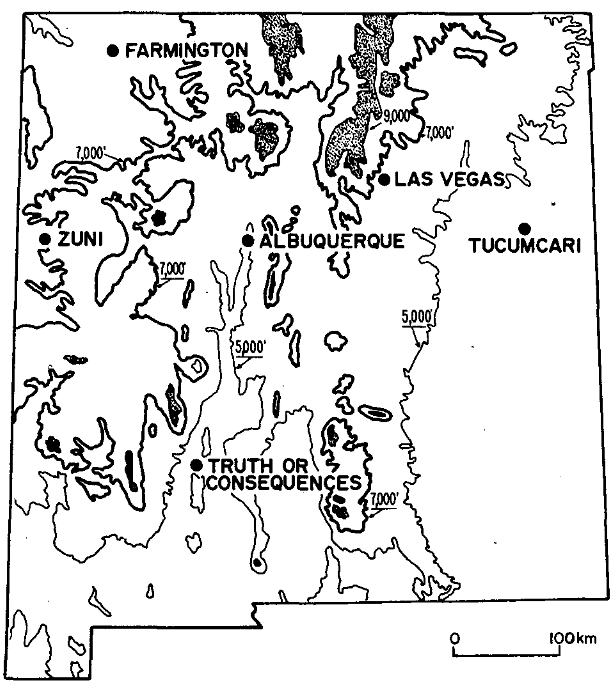

FIG. 1. Locations of the stations used in the analysis. Topographic contours are also shown.

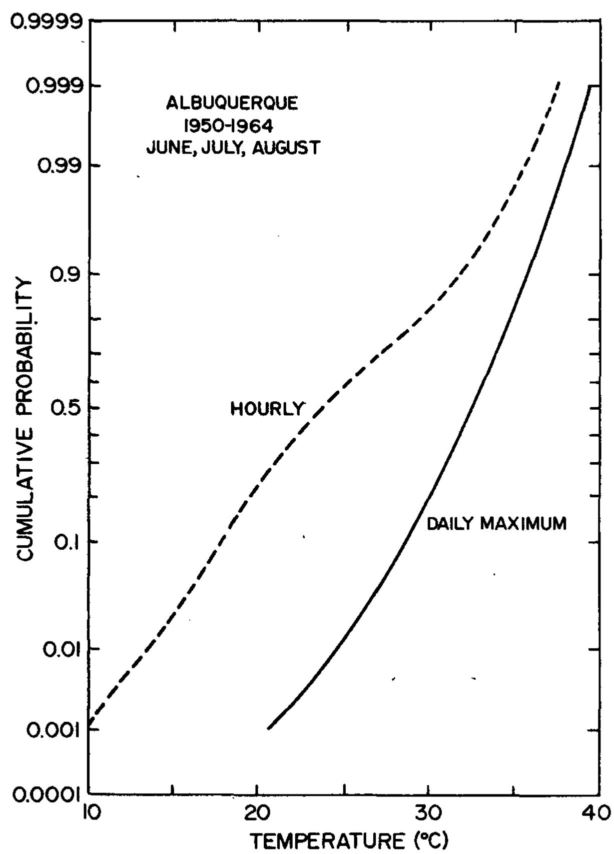

FIG. 2. Summer (June–August) cumulative probability density functions for Albuquerque, NM for daily maximum (solid) and hourly (dashed) temperatures. Ordinate scale is a normal probability scale.

The fitting of (2) to the CDFs was done using standard least square regression formulae with b = slope, $\ln a = y -$ intercept, the independent variable $x = \ln(T_0 - T)$, and the dependent variable $y = \ln[-\ln(1 - P)]$. A variety of sensitivity tests were performed to determine the range of CDF values to use in the curve fitting. The most critical area for determination of design temperature is the region from 0.900 to 0.995. Better fits were obtained over this region by restricting the curve fit to a relatively small region of the CDF. However, the fit was relatively insensitive to the exact region chosen. A good fit both to the probability range of 0.90 to 0.995 and to 0.50–0.999 was obtained by using the portion of the curve from 0.70 to 0.993.

In order to keep the method as simple as possible, T_0 was assumed to have the same value for both the hourly and the daily CDF and for all locations. This raises an interesting issue because (2) implies that T_0 is the extreme value of the distribution. Of course, the extreme value of the distribution will vary from location to location. Also, even at the same location, the hourly and the daily CDFs do not necessarily have the same extreme value when finite data sets are used. A related point is that (2) is similar in form to the cumulative Weibull distribution (Hahn and Shapiro, 1967) which has been applied in the analysis of extreme events. However, these issues are not pertinent here because the purpose of this analysis is not the estimation of extremes, and the data have not been processed in a manner appropriate for such an analysis. Therefore, the estimation of temperatures probabilities from (2) should be limited to values less than about 0.999.

The selection of the value of T_0 was accomplished by determining the goodness of fit for a large range of T_0 values. Once again, the fit was quite insensitive to the exact value of T_0. Best overall results were obtained for $T_0 = 60°C$. This value is too large to represent a real extreme of the distributions. This illustrates that the use of (2) for estimating probabilities should not be extended to probabilities very near one. However, use of this value does provide a good fit in the region of most interest to this paper, for probabilities in the range of 0.900–0.995.

Table 2 shows the resulting values of the coefficients for both the hourly CDFs and the daily CDFs. Also shown are the regression coefficients r^2. The r^2 values indicate that the exponential function provides a very good fit to the calculated CDFs.

DM found a good linear relationship between the hourly and daily values of the coefficients $\ln a$ and b. However, in the present study, the correlation between the coefficients in Table 2 was quite poor, too poor, in fact, to be of practical use. Nevertheless, visual examination of the hourly and daily CDF pairs indicated a consistent relationship. Further tests revealed that small changes in the CDF could result in large changes in the coefficients. This is a result of the rather small range of x [$\ln(T_0 - T)$] values that are used to fit the curves. The changes tend to be compensating; e.g., changes in $\ln a$ are compensated by changes in b in such a way that the resulting curve in the relevant range of x-values is changed only slightly, even though the changes in $\ln a$ and b may be substantial. Of course, outside of the actual range of x-values used, the changes would be significant.

A second approach to finding a mathematical relationship between the hourly CDF and the daily CDF was more successful. Figure 3 shows the fitted curves of the hourly and daily CDFs for Albuquerque. On this graph the transformation from the daily to the hourly CDF involves a lateral movement to the right with the lower end of the line moving a somewhat greater lateral distance than the upper end. The degree of rightward movement was quite similar from pair to pair. To quantify this, mathematical relationships were determined between the end points of the CDF segments in the following manner. For probability levels of 0.70 and 0.993, the corresponding x coordinates were calculated for each segment. The coordinate (x_{1h}) for hourly CDF at $P = 0.70$ is

$$x_{1h} = \frac{\ln[-\ln(1 - 0.70)] - \ln a_h}{b_h},$$

$$= \frac{0.1856 - \ln a_h}{b_h}. \quad (4)$$

TABLE 2. Exponential curve fitting coefficients.

Station	Period	Maximum temperature CDF				Hourly temperature CDF			
		a_m	$\ln a_m$	b_m	r^2	a_h	$\ln a_h$	b_h	r^2
Albuquerque	1950–64	6.9×10^{11}	27.26	−8.32	0.996	1.8×10^7	16.73	−4.76	0.999
Albuquerque	1973–78	4.6×10^9	22.25	−6.78	0.996	6.3×10^6	15.66	−4.45	0.999
Farmington	1953–64	8.2×10^{12}	29.74	−9.02	0.997	3.6×10^7	17.41	−4.94	0.998
Las Vegas	1950–64	1.7×10^{12}	28.14	−8.14	0.994	9.9×10^7	18.41	−5.02	0.999
Truth or Consequences	1951–64	8.6×10^9	22.87	−6.98	0.981	2.9×10^7	17.19	−4.94	0.999
Tucumcari	1950–54 1962–64	9.4×10^9	22.96	−7.15	0.995	1.7×10^6	14.36	−4.14	0.997
Zuni	1950–64	6.9×10^{13}	31.86	−9.47	0.999	1.3×10^8	18.68	−5.19	0.996

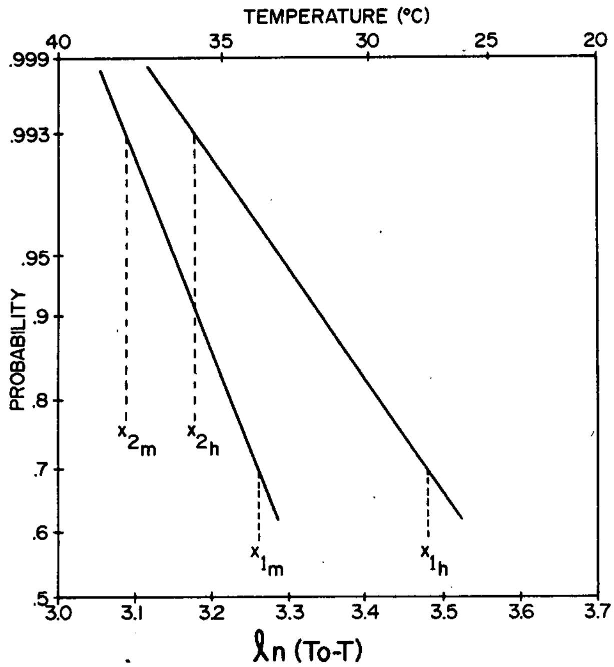

FIG. 3. Fitted functions for CDFs for Albuquerque. Solid line on the left is the fitted function to daily maximum temperatures and line on the right is the fitted function for hourly temperatures. Here x_{1m} and x_{1h} indicate the intercepts of the curves with probability = 0.7 while x_{2m} and x_{2h} indicate the intercepts of the curves with probability = 0.993. The probability scale is linear in $\ln[-\ln(1-P)]$.

For a probability = 0.993 the x coordinate (x_{2h}) is

$$x_{2h} = \frac{1.6018 - \ln a_h}{b_h}. \tag{5}$$

The corresponding x coordinates for the daily CDF are

$$x_{1m} = \frac{0.1856 - \ln a_m}{b_m}, \tag{6}$$

$$x_{2m} = \frac{1.6018 - \ln a_m}{b_m}. \tag{7}$$

Figure 4 shows a plot of x_{1h} vs x_{1m}. The relationship between the two is linear with a very good correlation. A least squares fit yields the following relationship:

$$x_{1h} = 0.99x_{1m} + 0.233. \tag{8}$$

Figure 5 shows the relationship between x_{2h} and x_{2m}. The following linear relationship was obtained:

$$x_{2h} = 0.98x_{2m} + 0.218. \tag{9}$$

These relationships provide the mathematical link between the hourly CDF and the daily CDF. On both Figs. 4 and 5, the elevation of each data point is shown. The data points are ordered according to elevation. This simply reflects the dependence of temperature on elevation since the x-variable is basically a temperature variable.

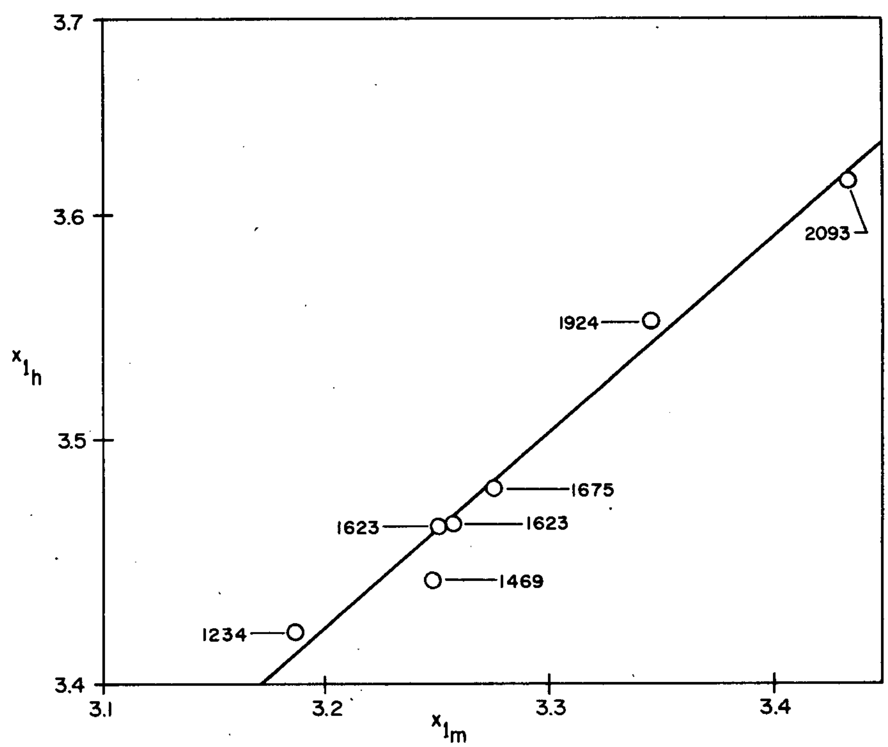

FIG. 4. The x-intercept of the fitted function of hourly temperature (x_{1h}) vs the x-intercept of the fitted function of daily maximum temperature (x_{1m}) with probability = 0.7. Elevation (m) is given next to each data point.

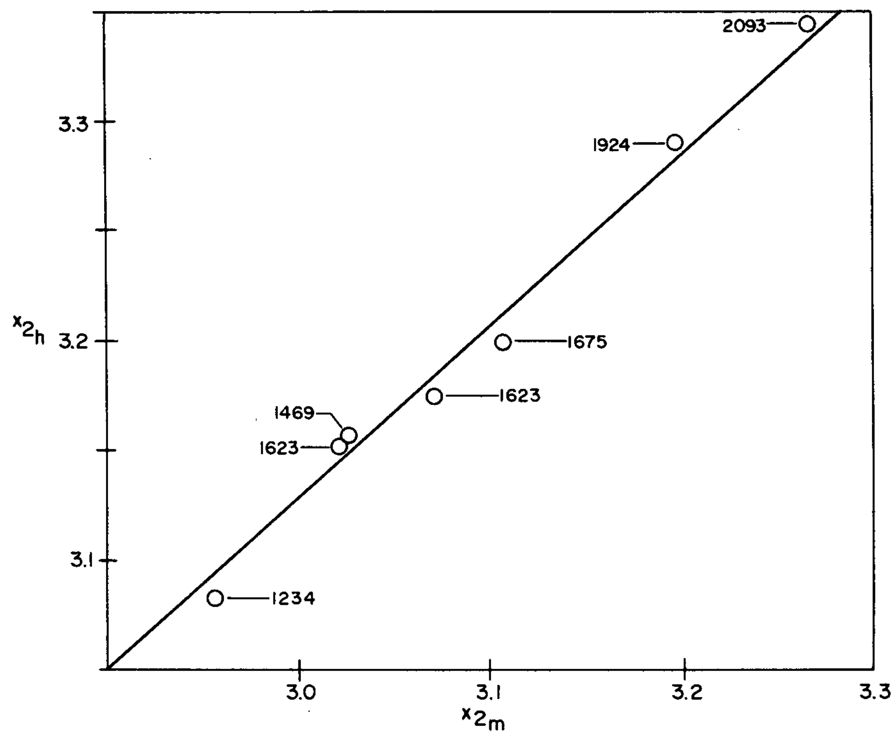

FIG. 5. The x-intercept of the fitted function of hourly temperature (x_{2h}) vs the x-intercept of the fitted function of daily maximum temperature (x_{2m}) with probability = 0.993. Elevation (m) is given next to each data point.

The procedure, then, for determining summer design temperatures from daily maximum temperatures is as follows:

1) Calculate the CDF for daily maximum temperatures using data from June, July and August.

2) Use linear regression formulae to determine the coefficients $\ln a_m$ and b_m for the probability range 0.70–0.993.

3) Calculate x_{1m} and x_{2m} from (6) and (7).

4) Calculate x_{1h} and x_{2h} from (8) and (9).

5) Calculate estimates of $\ln a_h$ and b_h from the following formulae:

$$b_h = \frac{\ln[-\ln(1-0.993)] - \ln[-\ln(1-0.7)]}{x_{2h} - x_{1h}},$$

$$= \frac{1.4162}{x_{2h} - x_{1h}}, \tag{10}$$

$$\ln a_h = \ln[-\ln(1-0.7)] - b_h x_{1h},$$

$$= 0.1856 - b_h x_{1h}. \tag{11}$$

6) Calculate design temperatures for any desired probability level by solving for T in (2).

An independent test of the accuracy of this method was done in a manner similar to DM. Six of the seven stations were used to derive regression relationships for x_{1h} and x_{2h} similar to (8) and (9). The resulting relationships were used to calculate design temperatures for the seventh station using the method above. These design temperatures were then compared to the actual temperatures from the hourly CDF for that station. This process was repeated seven times with each station being used as the control.

Table 3 shows the results of that test for each station when that station was used as the control. Temperatures for probability levels of 75%, 90%, 97.5% and 99% are shown. The synthesized temperatures compare very favorably. The average difference between the observed and synthesized temperatures is 0.2°C. The largest difference is only 0.9°C. These favorable results imply that this method may be applied to stations without hourly data with a high level of confidence.

4. Application to New Mexico

This method was also applied to a number of stations in New Mexico with only daily observations of maximum and minimum temperatures. Data from the period 1948–84 were used for this separate analysis. The resulting design temperatures were strongly dependent on elevation, as might be expected. Figure 6 shows the 99% summer design temperatures plotted as a function of elevation. Also plotted is a best fit straight line. The deviations from the best fit line are relatively small, although there is some indication that the slope may

TABLE 3. Comparison of temperatures (°C) at specified probability levels on observed (obs) and synthesized (syn) CDFs.

		Probability							
		0.75		0.90		0.975		0.99	
Station	Year	obs	syn	obs	syn	obs	syn	obs	syn
Albuquerque	1950–64	28.7	28.7	31.9	31.8	34.5	34.3	35.6	35.5
Albuquerque	1973–78	28.7	28.9	32.2	32.3	34.9	35.2	36.1	36.5
Farmington	1953–64	28.1	28.1	31.4	31.1	33.9	33.6	34.9	34.7
Las Vegas	1950–64	23.3	23.3	27.0	26.9	29.9	29.9	31.1	31.2
Truth or Consequences	1951–64	29.7	28.8	32.8	32.3	35.3	35.1	36.4	36.4
Tucumcari	1950–54 1962–64	30.2	30.9	33.9	33.9	36.5	36.5	37.6	37.6
Zuni	1950–64	25.5	26.1	29.0	29.2	31.5	31.8	32.5	32.9

change slightly at higher elevations. The 99% design temperature decreases by about 7.3°C per km.

Table 4 shows a comparison of the 99% summer design temperature between these results and those published by ASHRAE (1972, 1981). In general, the difference between these results and the ASHRAE results are less than 1°C, although some larger differences do occur. The 1972 ASHRAE results are most similar to these results. Differences between the two ASHRAE editions are not consistent in sign and are as large as the differences between these results and the ASHRAE results.

5. Conclusion

The method for estimation of summer design temperatures developed here is relatively straightforward

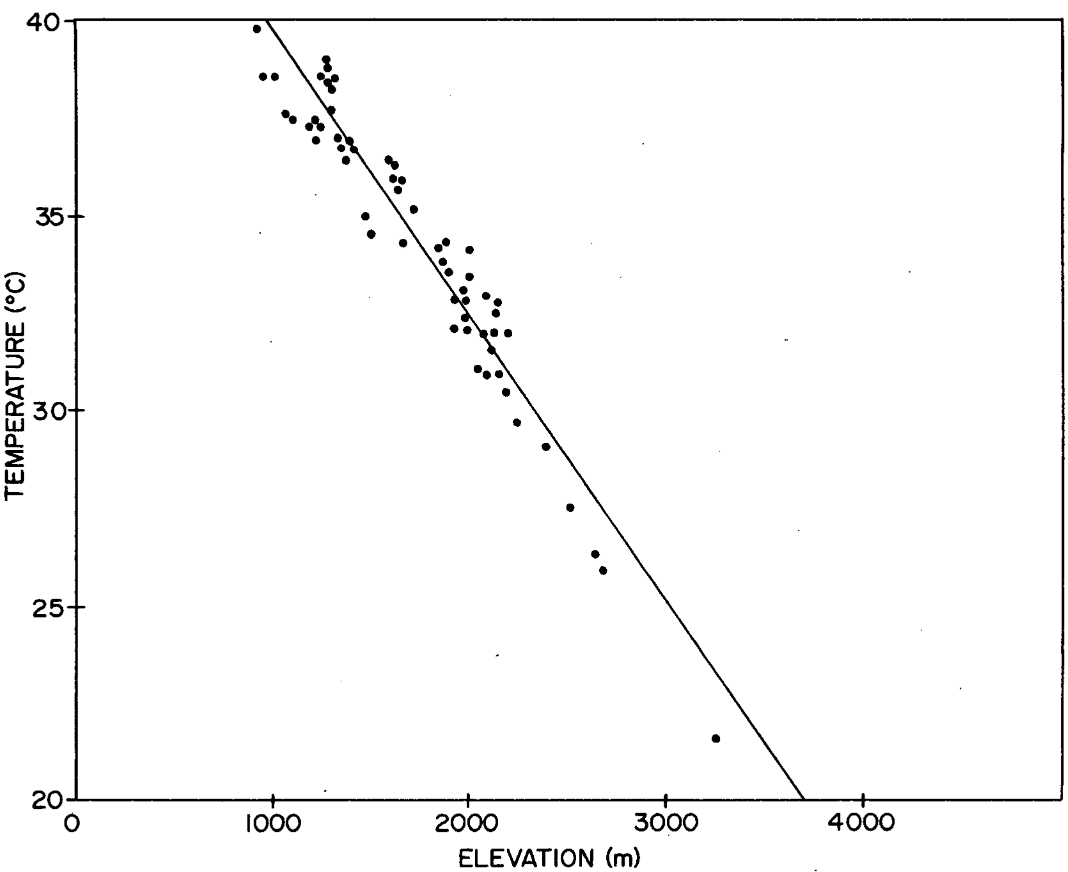

FIG. 6. 99% summer design temperature vs elevation for numerous New Mexico locations. Solid line shows a least squares fit to the data.

TABLE 4. Comparison with ASHRAE 99% summer design temperatures (°C).

Station	ASHRAE (1972)	ASHRAE (1981)	Current results
Alamogordo	37.8	36.7	38.6
Albuquerque	35.6	35.6	35.9
Artesia	38.3	39.4	38.6
Carlsbad	38.3	39.4	38.6
Clovis	37.2	35.0	37.7
Farmington	35.0	35.0	35.8
Gallup	33.3	32.2	33.4
Grants	32.7	31.7	33.1
Hobbs	38.3	38.3	37.4
Las Cruces	38.9	37.2	37.3
Los Alamos	31.1	31.7	29.7
Raton	33.3	32.8	30.5
Roswell	38.3	37.8	37.5
Santa Fe	32.2	32.2	31.9
Socorro	37.2	36.1	36.9
Tucumcari	37.2	37.2	37.4

and easy to apply. It also appears to be quite accurate and reliable. All of the stations used to develop it have relatively high average diurnal ranges of temperature, ranging from 15° to 20°C. This method is expected to be applicable to other areas with high diurnal ranges of temperature such as most of the western half of the United States. However, it is not clear whether the method would be applicable to areas with smaller diurnal ranges of temperature. It is recommended that the method be tested with hourly data before application to such areas.

In addition, the estimation of the hourly temperature CDF has been developed using only summer data. More work is required to determine whether such a method may be successful in other seasons such as spring and autumn.

Acknowledgments. I am grateful to Nolan Doesken and Thomas McKee for their review of this manuscript.

REFERENCES

American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc. (ASHRAE), 1972: Weather data and design conditions. *ASHRAE Handbook of Fundamentals*, ASHRAE, 667–668.

——, 1981: Weather data and design conditions. *ASHRAE Handbook 1981 Fundamentals*, ASHRAE, 24.1–24.15.

Doesken, N. J., and T. B. McKee, 1983: Estimating winter design temperatures from daily minimum temperatures. *J. Climate Appl. Meteor.*, **22**, 1685–1693.

Hahn, G. J., and S. S. Shapiro, 1967: *Statistical Models in Engineering*. Wiley and Sons, 355 pp.