2.2 QUALITY ASSURANCE OF MONTHLY TEMPERATURE OBSERVATIONS AT THE NATIONAL CLIMATIC DATA CENTER

Matthew J. Menne*
NOAA/NESDIS/NCDC, Asheville, North Carolina

Claude E. Duchon
School of Meteorology, University of Oklahoma, Norman, Oklahoma

1. INTRODUCTION

Since 1998 a three level, primarily statistical, quality assurance system for temperature observations has been under development at the National Climatic Data Center. Each level deals with a different time perspective and different classes of errors. As described in the 12th Conference on Applied Climatology by Menne and Duchon (2000), the three levels and associated classes are:

Level 1: Short-term perspective (daily values)
- class A: internal consistency errors; detected and corrected by logic tests
- class B: system errors; detected and adjusted for by rules based expert systems
- class C: inhomogeneities; detected by statistical tests

Level 2: Medium-term perspective (monthly values)
- class C: inhomogeneities; detected by applying statistical tests to time series of updated monthly means approximately 10 years in length

Level 3: Long-term perspective (annual values)
- class C: inhomogeneities: detected and adjusted for by applying statistical tests to records longer than two decades

A test involved in Level 1, class C error detection is discussed by Menne and Duchon (2001) and in Level 3 by, for example, Peterson and Easterling (1994) and Easterling and Peterson (1995). The purpose of this paper is to explore application of an existing maximum likelihood technique to the detection of inhomogeneities in monthly mean maximum and minimum temperature observations (Level 2 class C) at a given station called the candidate station. The tests for class C errors at Levels 1 and 2 are part of a strategy aimed at reducing the necessary time between the occurrence and detection of inhomogeneities. The methods developed for use at these levels depart from many existing methods by evaluating daily and monthly time series rather than the more common approach of using annual averages. This increases the number of observations available for analysis in a given time interval, an important issue when early detection is desired, but also raises new concerns, such as the presence of the annual cycle in the time series, that must be addressed.

While there are two general types of inhomogeneities, a step-change or shift and a trend, here only a shift in the average value or level of the candidate series is considered. The term 'monthly temperatures' is used to represent either monthly mean maximum or monthly mean minimum temperatures. The term mean temperature for January, for example, indicates that all available January temperatures (i.e., monthly mean maximums or minimums for 10 or so years) have been averaged.

Corresponding author address: Matthew J. Menne, National Climatic Data Center, Asheville, NC 28801; e-mail: Matthew.Menne@noaa.gov.

2. MAXIMUM LIKELIHOOD RATIO TEST

A practical approach to determining whether a candidate station has an inhomogeneity in temperature is to compare its time series of monthly temperatures to a reference series. A convenient way to form a reference time series is to average monthly temperatures from surrounding stations that have a climatology similar to the candidate station. This is the basis for the inhomogeneity test developed by Alexandersson (1986). A brief mathematical description of the maximum likelihood methodology developed by Alexandersson (1986) and Alexandersson and Moberg (1997) follows.

The comparison between candidate and neighbor time series can be expressed by

$$q_i = \left(y_i - \bar{y}\right) - \frac{\sum_{j=1}^{k}\rho_j^2\left(x_{ji} - \bar{x}_j\right)}{\sum_{j=1}^{k}\rho_j^2}, \quad i = 1,n \qquad (1)$$

where y_i and x_{ji} are monthly temperatures for the candidate and each of the k neighboring stations, respectively, and D_j are correlation coefficients between the candidate and each of the *k* surrounding stations. The quantities with an overbar are mean monthly temperature (see previous section), one for each month of the year, taken over the time series of length *n*. Weighting the candidate's neighbors with the square of the correlation coefficient ensures that those neighbors that have greater variance in common with the candidate carry more weight in creating the reference series than those with lesser variance in common. Thus the q-time series in (1) provides the departures of the candidate series (1st term) from the reference series (2nd term).

To form the likelihood ratio test the q-series needs to be standardized according to the usual expression

$$z_i = \frac{(q_i - \bar{q})}{s_q},\ i = 1,n \qquad (2)$$

Assuming that (2) is normally distributed, a single shift in the level of the candidate (y-series) can be determined using the null hypothesis H_o and alternative hypothesis H_1 given by

$$H_o : z_i \rightarrow N(0,1), i = 1,n \qquad H_1 : \left\langle \begin{array}{l} z_i \rightarrow N(\mu_1,1), i = 1,a \\ z_i \rightarrow N(\mu_2,1), i = a+1,n \end{array} \right\rangle$$

where N(g,h) indicates normal distribution with mean g and standard deviation h. If H_o is rejected in favor of H_1, the indication is that there has been a shift in the level of the y-series. The standard form for the likelihood ratio test statistic is (Wilks 1995 p. 135)

$$T = 2\ln\left[\frac{L(H_o)}{L(H_1)}\right] \qquad (3)$$

where the bracketed term is the ratio of likelihood functions and is given by

$$\frac{L(H_o)}{L(H_1)} = \frac{\exp\left[\frac{1}{2}\left(\sum_{i=1}^{a}(z_i - \mu_1)^2 + \sum_{i=a+1}^{n}(z_i - \mu_2)^2\right)\right]}{\exp\left[\frac{1}{2}\sum_{i=1}^{n} z_i^2\right]} \qquad (4)$$

in which the numerator and denominator are directly proportional to normal probability density functions of the z-series. If the ratio in (4) exceeds an appropriate critical value there is a statistical basis for claiming the mean level of the y-series from i = (1,*a*) is different from the mean level from i = (*a*+1,*n*). The criterion for rejection, derived from (3) using (4), in which the maximum likelihood estimators for $\bar{z}_1$ and $\bar{z}_2$ are their sample means, can be shown to be

$$T = a\bar{z}_1^2 + (n-a)\bar{z}_2^2 \ > C \qquad (5)$$

where C is the critical value corresponding to a selected level of significance. We have determined critical values of C for 2.5%, 5%, and 10% levels of significance through extensive simulations of time series of standard normal deviates. The results are, on the whole, similar to Alexandersson and Moberg (1997) but show some systematic differences.

Given a z-series, the value of *a* is systematically changed from some small value > 1 to some large value < *n*. For example, if *a* = 6 and *n* = 120, then $\bar{z}_1$ in (5) is calculated from the first 6 months of the series and $\bar{z}_2$ from the remaining *n*-*a* = 114 months. Next, *a* is increased by 1, $\bar{z}_1$ is recalculated for the first 7 months and $\bar{z}_2$ for the last 113 months. The procedure is repeated until *a* approaches *n*. A new value of T is calculated for each combination (*a*, *n*-*a*). If T > C for one or more combinations, a shift or step-change in the y-series is likely.

3. PROCEDURE

The first step is to select a candidate station and an appropriate number of neighboring stations, say, 5 or 6. The second step is to compute the correlation coefficients D_j in (1) between the candidate and each neighbor. To reduce the impact of a potential step-change or shift on the cross correlation calculation, a first-difference filter is applied to each time series. Then the annual cycle of first-differences is obtained by averaging the differences from all Januaries in the record, all Februaries, all Marches, etc.. The monthly average first difference is then subtracted from the original first-difference data. Removing the annual cycles reduces the otherwise resulting inflation of the magnitude of the correlation coefficient.

The third step is to calculate the q-series in (1) in which $\bar{y}$ and $\bar{x}$ are obtained by averaging the monthly temperatures from all Januaries in the record, all Februaries, Marches, etc. Thus there are 12 values of $\bar{y}$ and 12*k values of $\bar{x}$, i.e., 12 for each neighbor. Again, the purpose is to remove the annual cycle of monthly temperatures. An additional component of the third step is to scale the q-series according to the annual cycle of variance of the monthly temperatures. This is done because, particularly at mid-latitude stations, there is greater interannual variability of January (mean) temperatures than of July (mean) temperatures. The annual cycle of standard deviations that is calculated is representative of all stations in that the sum of the squares of the monthly departures from the mean monthly temperatures involved in computing the standard deviation for each month includes both the candidate and neighbors.

The fourth step is to standardize the q-series leading to the z-series in (2), which, therefore, has unit variance and zero mean. The fifth and final step is to compute the test statistic T for each value of *a* that was chosen and examine the test statistic time series for values of T > C for the chosen significance level. If the critical value has been exceeded, there is reason to believe an inhomogeneity of the form of a step-change or shift in level of the candidate station has occurred at or near the largest value of T.

4. APPLICATION

The maximum likelihood ratio (MLR) test was applied to monthly maximum and minimum time series from approximately 240 First-Order stations over the period 1991-2000 in order to assess the homogeneity of each series. This sample of station records was selected since the potential for an inhomogeneity exists in nearly each of the time series associated with the commissioning of the Automated Surface Observing System (ASOS). ASOS station commissioning was carried out systematically at these locations beginning in the early 1990s, with the majority completed during the mid- to late 1990s, as part of the National Weather Service's modernization program. Commissioning at a handful of First-Order stations occurred after the year 2000. The changeover to ASOS brought a change in temperature measurement equipment as well as possible changes in instrument siting characteristics. Each factor may contribute to discontinuities in temperature observations from First-Order stations.

Excluding those occasions where the maximum likelihood ratio test statistic, T, critical value is reached within the first or last few observations only (see section 5), our analysis shows that the number of stations where the MLR test (5% significance level) suggests a change in mean level somewhere in the 10-year period is large: 165 for monthly maximum series and 148 for monthly minimum temperature. An example where the likelihood ratio test suggests one change in mean level, coincident with ASOS commissioning, is shown in Fig. 1. Fig. 1(a) shows mean monthly maximum temperature departures at Phoenix, AZ, as well as departures from the reference series calculated using (1). The standardized differences between the candidate and reference series are shown in Fig. 1(b), and the time series of the test statistic, T, is shown in Fig. 1(c). ASOS commissioning at Phoenix occurred in March 1994 (month 40 of the 120-month time series), which is the month where the magnitude of the test statistic reaches a maximum. The maximum value of the test statistic is the most likely location in the time series where a change in mean level occurs. The dashed line in Fig. 1(c) shows the magnitude of the T-critical value for a time series of this length.

It should be noted that the critical value of the MLR test statistic is commonly exceeded more than once during the 1991-2000 period for a number of First-Order monthly temperature series. The presence of multiple "peaks" in the test statistic suggests at least the potential for multiple changes in the mean level of monthly temperature series within the 10-year period. Fig. 2 shows an example with multiple peaks in the test statistic and provides the same time series for Charlotte, NC as shown in Fig. 1. Note that one of the peaks in T (Fig. 2.c.) is coincident with the ASOS commission date of July 1998 (month 90 of the 120-month series).

Since the dates of ASOS station commissioning are known, the presence of a statistically significant peak in the likelihood ratio test statistic close to the time of ASOS commission date is of interest. Fig. 3 shows histograms of the difference, in months, between the date of ASOS commissioning and the date of the nearest significant peak (if more than one) in the T test statistic for both monthly maximum and monthly minimum temperature series. Stations commissioned after 2000 were excluded. It is clear from Fig. 3 that the MLR test suggests that a change in the mean level of monthly maximum and minimum temperatures series likely occurred at a number of First Order stations coincident with ASOS station commissioning. The results summarized in Fig. 3 suggest also that the MLR test is amenable to the evaluation of serial monthly temperature observations, subject to the modifications described in section 3.

As stated in the introduction, the motivation for this work was to provide a contribution to the design of a QA system that includes tests capable of reducing the time between the occurrence of a discontinuity its detection. Consequently, we also ran the MLR test on the same monthly temperature series, but this time truncated 10 months after the ASOS commission date. This was done to shed some light on the capability of the MLR test to identify inhomogeneities occurring near the end of a station's period of record since the ultimate use of the MLR test will be in a monitoring capacity, evaluating the homogeneity of temperature time series updated each new data month. Fig. 4 shows histograms identical to those in Fig. 3, except in this case the differences between the peak in the test statistic T and the commission date are for the truncated time series. The histograms of differences indicate a slight increase in the number of time series having an apparent change in mean level that coincides with the beginning of ASOS period. A possible contributing factor behind this slight increase in MLR test rejections is discussed briefly in section 5. An example of an ASOS-coincident MLR test rejection occurring using the truncated time series, but not using the original series is shown in Fig. 5. The figure shows the same suite of time series as in Figs. 1 and 2, in this case for mean monthly maximum temperatures at Des Moines, IA. The difference between Fig. 5 and the earlier figures is that two time series for the T test statistic are given in Fig. 5(c). The solid line represents the time series of T for the full 120 months while the dashed line represents the magnitude of T for the 69-month temperature series ending in October 1996, 10 months following the December 1995 ASOS commissioning.

5. CAVEATS

Evaluation of a large number of simulated homogeneous time series has shown that some preference exists for the maximum value in the T test statistic to occur near the ends of a period of record. This is a consequence of the relatively few observations available to provide a stable estimate of the mean level for the shortest segments that occur at the beginning or end of a time series. In the application of t e MLR test on the First-Order temperature time series for the decade of the 1990s, however, the maximum value of T occurred near the ends of the time series relatively infrequently. In fact, when a change in mean level is present somewhere else in the time series, the value of the test statistic near the ends of the time series rarely approaches the 5% significance level. Nevertheless, the slight increase in the number of rejections near the end of the truncated time series as indicated through a comparison of Figs. 3 and 4, even though coincident with ASOS commissioning, may be evidence of the need to modify the magnitude of the critical value of the test statistic as a function of the position of the observation in the time series.

In addition, experience has shown that when more than one change in mean level is present in a series, however brief, the time series of the test statistic can sometimes be difficult to interpret since the presence of one change in mean level can affect the magnitude of the test statistic in the vicinity of a second change. The solution to this problem could take the form of analyzing successive segments of a time series separately, as suggested by Alexandersson and Moberg (1997), and/or through the successive adjustment for earlier discontinuities before any evaluation of more recent segments of a temperature series are carried out.

6. CONCLUSION

The modification of the MLR test described in this paper appears to be quite sensitive to changes in mean monthly temperature associated with the commissioning of ASOS at First-Order stations. In fact, the test statistic frequently points to the most likely month for a discontinuity that is exactly coincident with the ASOS commission date. In addition, a simple test, conducted by truncating each of the approximately 240 time series 10 months after ASOS commissioning suggests, at least for this sample of stations, that there little reduction in the capability of MLR test to identify the same changes when they occur near the end of the period of record. The MLR test also may be used as a means to test for the presence of trends in a time series. Future reports will address the characteristics of the MRL test when used for the identification of non-climatic trends in monthly temperature series.

References

Alexandersson, H., 1986: A homogeneity test applied to precipitation data. *J. Climatol*., **6**, 661-675.

Alexandersson, H., and A. Moberg, 1997: Homegenization of Swedish temperature data. Part I: Homogeneity test for linear trends. *Int. J. Climatol*., **17**, 25-34.

Easterling, D. R., and T. C. Peterson, 1995: A new method for detecting and adjusting for undocumented discontinuities in climotological time series. *Int. J. Climatol*., **15**, 369-377.

Menne, M. J., and C. E. Duchon, 2000: Quality assurance of daily temperature observations at the National Climatic Data Center. Preprints, Twelfth Conference on Applied Climatology, Asheville, NC, Amer. Meteor. Soc., J48-J50.

Menne, M. J., and C. E. Duchon, 2001: A method for monthly detection of inhomogeneities and errors in daily maximum and minimum temperatures. *J. Atmos. Oceanic Technol*., **18**, 1136-1149.

Peterson, T. C., and D. R. Easterling, 1994: Creation of homogeneous composite climatological reference series. *Int. J. Climatol*., **14**, 671-697.

Wilks, D. S., 1995: Statistical Methods in the Atmospheric Sciences. International Geophysics Series, Vol. 59, Academic Press, 464 pp.

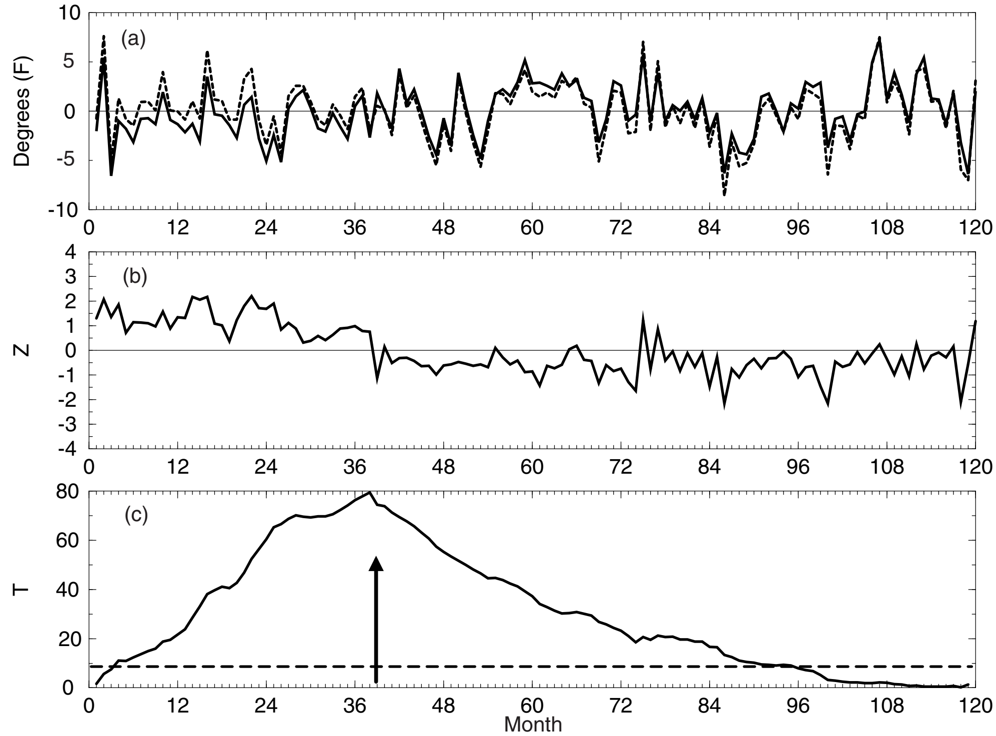

Fig. 1. (a) Time series of monthly maximum temperature departures January 1991 through December 2000 at Phoenix, Arizona (solid line) and for the reference series comprised of nearby stations (dashed line), calculated using Eq. (1); (b) standardized difference (z) between the candidate and reference series; and, (c) time series of the maximum likelihood ratio test statistic, T, for a change in mean level. The dashed line is the magnitude of the 95% critical value for a time series with 120 observations. The date of ASOS commissioning is noted by the arrow at month 40 (March 1994).

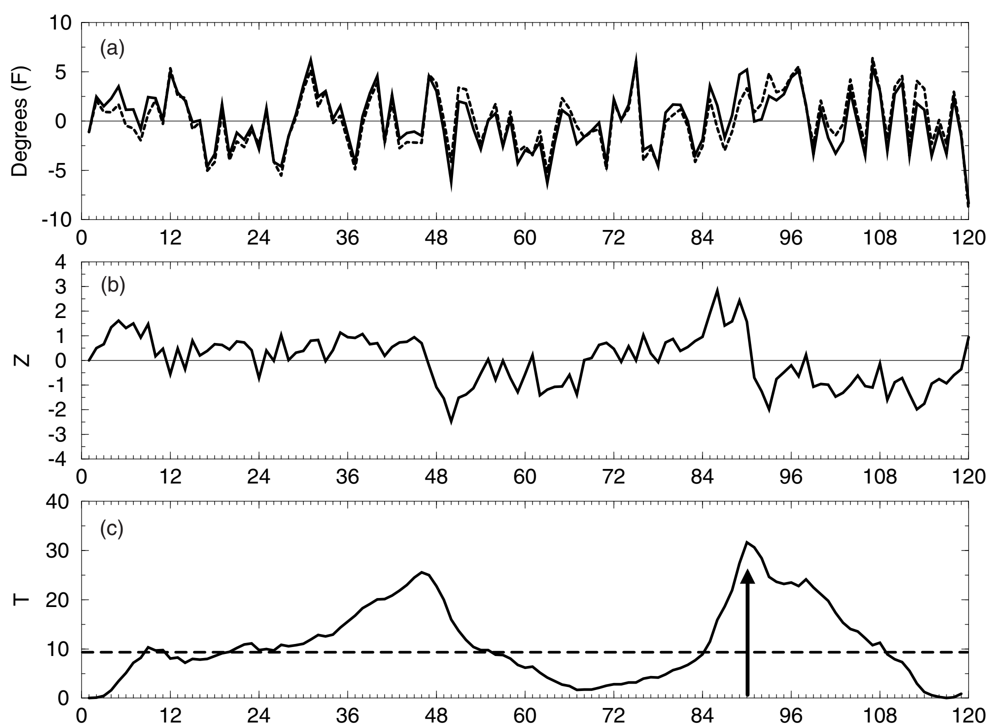

Fig. 2. As in Fig. 1, except for Charlotte, North Carolina.

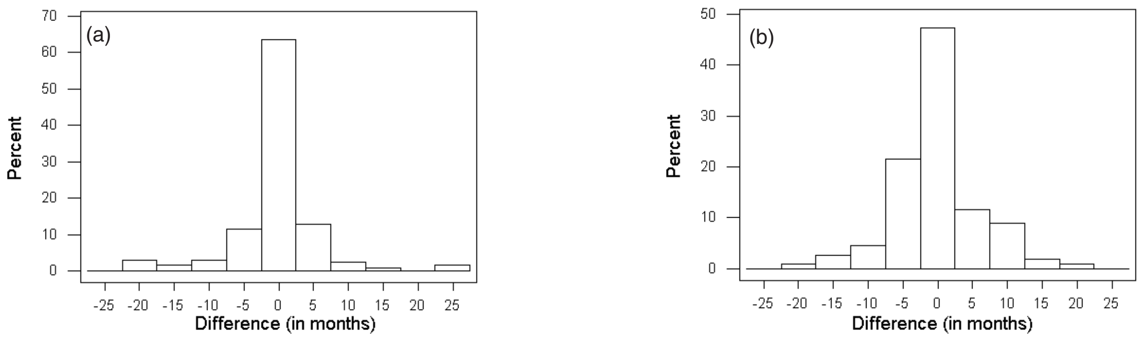

Fig. 3. Histogram of the differences between the date of nearest statistically significant peak in the maximum likelihood ratio test statistic, T, and the date of ASOS station commissioning for at First-Order stations. (a) Monthly maximum temperatures; (b) Monthly minimum temperatures.

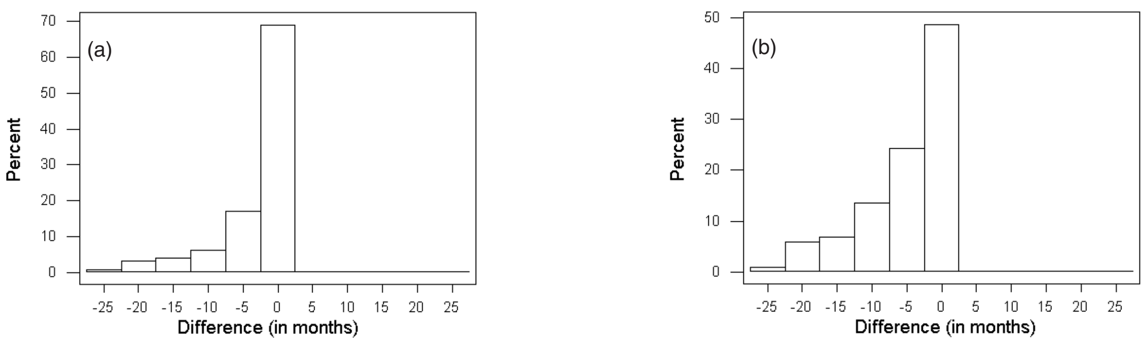

Fig. 4. As in Fig. 3, except each First-Order temperature series was truncated 10 months following the ASOS commission date before application of the MLR test.

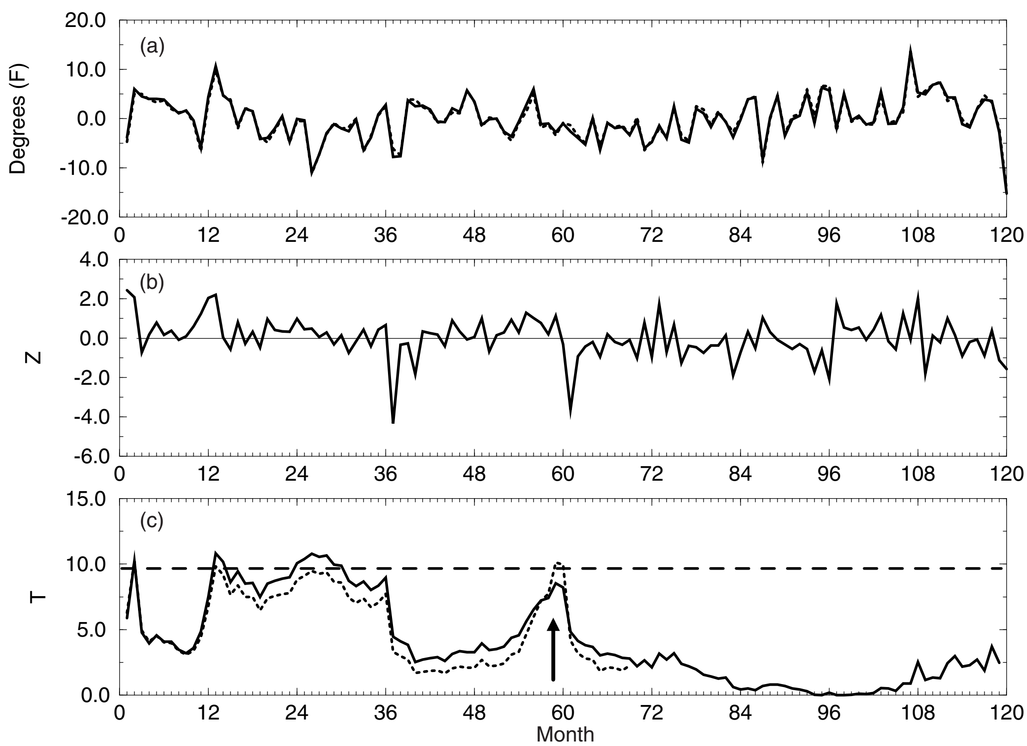

Fig. 5. As in Figs. 1 and 2, except for Des Moines, Iowa, and in (c), the time series of the test statistic, T, is given for an evaluation of the full 10-year period (solid line) and for a period of record ending 10 months after the December 1995 ASOS commission date (dashed line).