Identifying Rogue Air Temperature Stations Using Cluster Analysis of Percentile Trends

SCOTT M. ROBESON AND JEFFREY A. DOTY
Department of Geography, Indiana University, Bloomington, Indiana

(Manuscript received 19 June 2003, in final form 14 July 2004)

ABSTRACT

A new and efficient method for identifying “rogue” air temperature stations—locations with unusually large air temperature trends—is presented. Instrumentation problems and spatially unrepresentative local climates are sometimes more apparent in air temperature extremes, yet can have more subtle impacts on variations in mean air temperature. As a result, using data from over 1300 stations in North America, the tails of daily air temperature frequency distributions were examined for unusual trends. In particular, linear trends in the 5th percentile of daily minimum air temperature during the winter months and the 95th percentile of daily maximum air temperature during the summer were analyzed. Cluster analysis then was used to identify stations that were distinct from other locations. Both single- and average linkage clustering were evaluated. By identifying individual stations along the entire periphery of the percentile trend space, single-linkage clustering appears to produce better results than that of average linkage. Average linkage clustering tends to group together several stations with large trends; however, only a handful of these stations appear distinctly different from the large body of trends toward the center of the percentile trend space. Maps of the rogue stations show that most are in close proximity to numerous other stations that were not grouped into the rogue cluster, making it unlikely that the unusually large temperature trends were due to regional climatic variations. As with all approaches for evaluating data quality, time series plots and station history information also must be inspected to more fully understand inhomogeneous variations in historical climatic data.

1. Introduction

Tremendous efforts have been made to create high-quality air temperature archives from instrumental data (e.g., Jones and Moberg 2003; Peterson and Vose 1997). These data are pivotal for estimating regional, continental, and global-scale air temperature variations. When developing these databases, station time series usually are analyzed for homogeneity, that is, temporal variability that is consistent with a number of expectations. These expectations include climatically realistic temporal variability, such as a lack of step changes, and some degree of similarity between proximate stations. Due to variations in elevation and exposure, however, proximate stations frequently have dissimilar mean air temperature. Still, proximate locations should have similar air temperature variability, which is a fundamental reason why air temperature anomalies, and not actual air temperatures, are used in climatic change research. Although they are a primary form of temporal variability, trends are not routinely calculated at the station level. Typically, only after data have been adjusted for inhomogeneities, converted to anomalies, and gridded are trends estimated (DeGaetano and Allen 2002). Trends at the station level, however, contain a wealth of information that can be used to assess the homogeneity of stations for climate change research (Alexandersson and Moberg 1997; DeGaetano and Allen 2002).

Trends in mean air temperature at individual stations can be compared to one another, although progressive changes in the local environment and different sensors can preferentially affect the tails of air temperature distributions. Different sensors and sensor exposures may not represent extremely high or low temperatures as consistently as they represent more central values (Gallo et al. 1992; Kessler et al. 1993; Quayle et al. 1991). Urbanization and modifications to the local environment (such as changes in sky-view factors, soil compaction, and irrigation practices) also can preferentially affect the tails of air temperature distributions, particularly for the daily minimum air temperature (Gallo et
Inhomogeneous changes in the tails of air temperature frequency distributions may be more detectable, yet have subtle effects on trends in mean air temperature that would go unnoticed if only the mean were analyzed (Fig. 1).

As a result, we analyze the spatial homogeneity of trends in time-varying percentiles of daily minimum ($T_{\text{min}}$) and maximum ($T_{\text{max}}$) air temperature over North America (Robeson 2004). Because the expectation is that nearby stations should have similar trends, cluster analysis is used to evaluate the homogeneity of these trends and to aid in the identification of “rogue” stations. In particular, trends in the coldest daily minima during winter and the warmest daily maxima during summer are analyzed to illustrate the benefits of this approach.

2. Background

Homogeneity research typically focuses on constructing large climatological datasets for comparative analyses and climate change research (e.g., Peterson and Easterling 1994; Vincent et al. 2002). Most methods analyze each individual climate station in comparison to some reference (e.g., a long-term average or some combination of nearby stations). Further, many techniques address the identification and adjustment of temporal discontinuities that are caused by artificial sources, such as instrument or station location changes, but fewer are designed to locate climate stations that, independent of temporal inhomogeneities, may have climatic trends that are different from regional patterns (exceptions include Easterling et al. 1996; Alexanderson and Moberg 1997; Vincent 1998; DeGaetano and Allen 2002). In general, the identification and adjustment of anomalous gradual trends in a station time series, rather than abrupt jumps, is a more difficult task (Easterling and Peterson 1995).

A number of homogeneity methods are dependent on subjective visual comparisons of station data time series. Double-mass plots, for example, plot the cumulative sum of a variable at a candidate station against a reference station, and a sudden change in slope indicates a discontinuity (Peterson et al. 1998). Comparison of only two stations at a time, however, makes it impossible to detect which station represents the problem (Peterson et al. 1998). The inclusion of several reference stations in the analysis increases the likelihood of identifying the station that requires adjustment (Easterling and Peterson 1995). Visual inspection for jumps in side-by-side anomaly time series has also been used to identify discontinuities (Jones et al. 1986).

**Fig. 1.** Time series and ordinary least squares trends lines for the 5th, 50th, and 95th percentiles of Dec $T_{\text{min}}$ at Haliburton, ON (station number: 6163156). The 5th percentile of $T_{\text{min}}$ at Haliburton has a strong trend of 6.7°C (50 yr)$^{-1}$, whereas modest negative trends are evident in the 50th or 95th percentile [−1.3°C and −1.8°C (50 yr)$^{-1}$, respectively]. Unlike the median, the monthly mean is influenced by the fifth percentile values; therefore, the warming in the fifth percentile produces a subtle increase in monthly mean air temperature at Haliburton [trend of 0.33°C (50 yr)$^{-1}$, not shown].
The most common category of inhomogeneity detection and adjustment involves the creation of a reference time series for comparison with candidate stations (e.g., Karl and Williams 1987). This method typically requires the creation of a reference series for each station; however, the selection of appropriate neighboring stations that are to be included in the reference can be problematic (Lanzante 1996). Strong direct relationships (e.g., positive correlation) are often used in the selection of reference stations to ensure that they exhibit a similar climate variability as that of the candidate (Karl and Williams 1987; Peterson and Easterling 1994).

Reference time series studies have differed in the type of series used for comparison, with the most common type being a difference series between a candidate and reference (Alexandersson and Moberg 1997; Vincent 1998) or anomaly series (Jones et al. 1986). Once the series are created for comparison, detection of the inhomogeneity can be accomplished by a variety of statistical methods, such as a two-sample t test (Karl and Williams 1987), linear regression (Vincent 1998), or a standard normal homogeneity test (Alexandersson and Moberg 1997). Another detection method uses multiple-phase regression, which separately analyzes subsets of the stations’ time series before and after a potential discontinuity (Easterling and Peterson 1995; Sow 1987). Adjustments then are made with reference to the differences in the means of the candidate and reference time series (Alexandersson and Moberg 1997). When percentile exceedances are the variables being studied (as opposed to the time-varying percentiles that we utilize), different percentile thresholds may be applied to the inhomogeneous periods (DeGaetano and Allen 2002).

3. Methods

a. Trends in air temperature percentiles

To analyze the spatial homogeneity of air temperature trends, we utilize time-varying percentiles. Analysis of trends in time varying percentiles is a flexible alternative to traditional approaches that focus on variations in the central tendency, which is usually the mean (Robeson 2002b; 2004). In addition, the analysis of changes in low-probability events (such as the 5th or 95th percentile) can provide valuable information for climate impact studies where extreme events are important (Meehl et al. 2000). In this study, however, low-probability events are used because many types of inhomogeneities may differentially affect the tails of air temperature distributions. The methods discussed here are generalizable, though, and could be used on trends in the central tendency as well.

Similar to the way that a monthly mean is calculated, monthly percentiles of air temperature can be estimated at each station. These percentiles can then be analyzed in a variety of ways (Robeson 2004). In this research, trends in percentiles are estimated using linear least squares regression, although methods that are less sensitive to influence points (Huber 1981) or autocorrelation (Zhang et al. 2000) also could be used. Linear trends in December $T_{\text{min}}$ percentiles at Haliburton, Ontario, Canada, for instance, show that the 5th percentile has a strong upward trend over the last 50 yr, while the 50th and 95th percentiles show virtually no change (Fig. 1). If stations close to Haliburton do not have similar trends in the 5th percentile during December, this suggests that the lower tail of this station’s frequency distribution is unrepresentative of large-scale spatial patterns. Its station record could then be examined more closely, or it could be removed from the network. It is important to note that the methods developed here are aimed at identifying stations with unusual trends; therefore, stations with other forms of inhomogeneities (such as multiple, offsetting changes) should be identified using other methods.

b. Clustering alternatives

Once percentile trends are calculated, cluster analysis is used to analyze the spatial homogeneity of trends and to help identify “rogue” stations. Numerous clustering methods are available for use in identifying homogeneous spatial patterns. The choice of clustering methods depends primarily on whether the goal of cluster analysis is to identify cohesive regional trends in the data or to identify locations that are most unusual and, therefore, may represent an inhomogeneity or other problems in the data. Commonly used methods of cluster analysis include single-linkage, average linkage, complete linkage, Ward’s, centroid, and $k$ means (Kalkstein et al. 1987; Fovell and Fovell 1993).

The nature of the methods, as well as previous work with climate data (Kalkstein et al. 1987; Fovell and Fovell 1993; Gong and Richman 1995; Jackson and Wein and 1995), suggests that average linkage or $k$ means would be useful for identifying cohesive regional percentile trends. However, where the goal is to identify rogue stations, a method such as single linkage, which tends to produce small-element clusters that represent the most unusual locations, would be desirable. Average linkage or $k$ means may be useful if they can identify a cluster of rogue stations with similar characteristics (e.g., increasing urbanization). In this study, both single- and average linkage clustering methods are used and compared. As expected, the average linkage method produces multiple-station clusters, whereas the single-linkage method typically produces single- or two-element clusters. The rogue stations from the single-linkage clusters were compared to those in the average linkage cluster(s) to see if both methods were identifying the same stations.

c. Determining the number of rogue clusters

Most algorithms for hierarchical clustering methods, such as single or average linkage, agglomerate clusters.
Solutions, as a result, range from \( n \) clusters, with each observation in its own cluster, to a single cluster that contains every observation (nonhierarchical methods such as \( k \) means require that the number of clusters be specified). Identifying the “correct” number of clusters, therefore, is a compromise between the idiosyncrasies of individual stations and the generality of merging large numbers of stations into groups that are deemed similar. A number of stopping rules are available for determining an appropriate clustering solution, including both graphical and statistical approaches. Using randomly generated data, Milligan and Cooper (1985) compared 30 different computational methods. Although some methods clearly perform better than others, Milligan and Cooper (1985) warn users that their results may be data dependent.

Statistical stopping rules, such as the pseudo \( F \) statistic (Calinski and Harabasz 1974) recommended by Milligan and Cooper (1985), can be used in the current context. We choose, however, to evaluate the linkage distance between successive clustering stages graphically—the approach that is used by Kalkstein et al. (1987) and discussed by Wilks (1995). By evaluating when dissimilar clusters are being merged, which is the same information that is used by statistical procedures, and ensuring that the linkage distance is sufficiently large [e.g., > 1°–2°C (50 yr)\(^{-1}\)], this approach provides a scientifically meaningful termination of the clustering procedure. Otherwise, statistical stopping rules may find a break in the clustering procedure that does not identify clusters with unusually large trends.

d. A demonstrative example

To demonstrate the effectiveness and efficiency of applying cluster analysis to air temperature trends, we created a number of random and perturbed time series with various characteristics and then subjected them to the methods outlined above. In total, synthetic time series for 3 months at 10 different “stations” were created using a uniform random number generator. Two of these stations had single months that were perturbed: one station had a modest step change during 1 month (Fig. 2a), and another station had a weak trend added to its time series (Fig. 2b). Single-linkage cluster analysis was then performed on the 10 \( \times \) 3 matrix of trends derived for each of the 3 months at the 10 locations.

Both the dendrogram (Fig. 2c) and the linkage-distance diagram (Fig. 2d) identify a two-cluster solution, with the two perturbed stations (stations 1 and 2, as indicated in Fig. 2c) being grouped together. Further sensitivity analysis was performed by adding additional stations with progressively smaller step changes and

![Fig. 2. Overview of a demonstrative example of cluster analysis applied to air temperature trends at 10 “stations”: (a) time series for 3 months at station 1, which includes a step change during 1 month (bold line); (b) time series for 3 months at station 2, which includes a modest trend during 1 month (bold line); (c) dendrogram; and (d) linkage distance diagram of the single-linkage cluster results, showing that a two-cluster solution is appropriate, with stations 1 and 2 forming a distinct (rogue) cluster that should not be joined with the other cluster that includes the eight stations with random time series.](image-url)
trends. Essentially, as the trends or step changes become smaller and smaller, one has to be increasingly cautious with the number of clusters (i.e., allowing a conservatively large number of clusters to be identified as rogue) in order to identify all of the perturbed stations. Occasionally, a few of the random time series can be included in this “super set” of rogue stations, making it clear that subsequent analysis of the rogue station clusters is needed. The key point here, however, is that very little information is needed to identify the rogue stations—only the trends at the 10 stations. While results from the synthetic data can be limiting, they do demonstrate both how the method works and that it works in an efficient manner. In particular, if thousands of stations were subjected to this type of analysis, a super set of rogue stations could be identified, without the need to develop a reference series for each station or to visually inspect the time series for every station. Only those in the rogue clusters would need further analysis.

FIG. 3. Linkage distances for 5th percentile trends of winter $T_{\text{min}}$ and 95th percentile trends of summer $T_{\text{max}}$ for (a) single and (b) average linkage as a function of clustering level. Moving from large numbers of clusters to smaller numbers of clusters, a large jump in linkage distance indicates that two dissimilar clusters are being merged.
Fig. 4. Results of single-linkage cluster analysis of trends in the fifth percentile of $T_{\text{min}_5}$ ($T_{\text{min}_5}$) during winter months from 1948 to 2000 across North America: (a) six single-station clusters were identified as being anomalous (solid circles, other station locations used are indicated by the plus), (b) three-dimensional stem plot of $T_{\text{min}_5}$ trends ($^\circ$C (50 yr)$^{-1}$) during Dec, Jan, and Feb at rogue stations (solid circle) and other locations in North America.
4. Data

To examine trends in time-varying percentiles, high temporal resolution data are needed over relatively long periods of time. Relatively high-quality daily $T_{\text{min}}$ and $T_{\text{max}}$ data have recently become available for much of the United States and Canada. The data used here are derived from (i) the daily United States Historical Climatology Network (HCN) archive (Easterling et al. 1999), which contains 1062 stations over the 48 contiguous states, (ii) the daily Historical Adjusted Climate Database for Canada (HACDC; Vincent et al. 2002), and (iii) a subset of the Alaska stations from the Global Daily Climatology Network (GDCN; National Climatic Data Center 2002), selected for long-term, nonurban locations. The combination of the daily HCN, HACDC, and Alaska subset of the GDCN produces a network of 1324 stations that have records spanning from 1948 to 2000 (see Fig. 4a for spatial distribution of stations; all stations had data for at least 80% of the months used). The daily HCN and Alaska stations have been selected for their long-term quality, based on criteria such as consistent observation times, a low potential for urban bias, and other quality assessments that were developed for the monthly HCN. The Canadian station data already have been homogenized (Vincent 1998) and, therefore, provide a useful test of whether stations with spatially inhomogeneous trends in the tails of frequency distributions still exist within an adjusted dataset.

5. Results and discussion

a. Clustered percentile trends

To demonstrate the methods described here, single-linkage clustering was applied to trends in the 5th per-

![Graphs showing time series of daily minimum air temperature for December, January, and February at Stambaugh, MI, and its three nearest neighbors.](image)

Fig. 5. Time series of $T_{\text{min},5}$ during Dec, Jan, and Feb at Stambaugh, MI, and at its three nearest neighbors.
centile of $T_{\text{min}}$ ($T_{\text{min,5}}$) during winter (December–January–February) and the 95th percentile of $T_{\text{max}}$ ($T_{\text{max,95}}$) during summer (June–July–August). As a result, each clustering analysis utilizes three variables (trends for each of the 3 months) at 1324 locations. These variables were chosen for the following two reasons: (i) the response of instrumentation is such that the lowest and highest temperatures are often not represented as well as more typical temperatures (particularly for electronic instruments; DeFelice 1998), and (ii) local scale biases are often more apparent in the lowest values of winter $T_{\text{min}}$ (e.g., less radiational cooling at sites in built environments; Oke 1981) and the highest values of summer $T_{\text{max}}$ (e.g., increased sensible heat and decreased evapotranspiration in many built environments; Oke 1982). Therefore, trends in the lower tail of $T_{\text{min}}$ during winter and the upper tail of $T_{\text{max}}$ during summer should be useful for identifying potential rogue stations. It should be noted, however, that trends for additional percentiles and other times of the year are easily integrated into this procedure. One of the primary advantages of using cluster analysis in homogeneity research is that it can classify stations within a large, multidimensional space.

For single-linkage analysis of $T_{\text{min,5}}$ trends, graphical evaluation of the linkage distance between successive clustering stages (Fig. 3a) suggests a seven-cluster solution. At that clustering level, it is clear that continuing to merge clusters would combine two $T_{\text{min,5}}$ clusters that (i) are distinct and (ii) have large linkage distances from one another. Of the seven clusters, six clusters (each having one station) are identified as rogue clusters (Fig. 4). The six rogue clusters had $T_{\text{min,5}}$ trends during the winter months that were distinct from all of the other stations in North America and lie along the periphery of the percentile trend space for the three

<table>
<thead>
<tr>
<th>Station name and coop identification</th>
<th>Record date</th>
<th>Observer change</th>
<th>Station move</th>
<th>Obs time (LST)</th>
<th>Elev (m)</th>
<th>Instrument type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stambaugh MI 207812</td>
<td>2 Dec 1949</td>
<td>0.48 km southeast</td>
<td>Unknown</td>
<td>492</td>
<td>38 m south-southwest</td>
<td>CRS</td>
</tr>
<tr>
<td></td>
<td>26 Aug 1954</td>
<td>N</td>
<td>1700</td>
<td>492</td>
<td>CRS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15 Aug 1962</td>
<td>Y</td>
<td>1600</td>
<td>543</td>
<td>CRS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 Dec 1975</td>
<td>Y</td>
<td>1600</td>
<td>543</td>
<td>CRS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 Sep 1982</td>
<td>N</td>
<td>0700</td>
<td>452</td>
<td>CRS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 Dec 1986</td>
<td>Y</td>
<td>1600</td>
<td>475</td>
<td>CRS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>29 Oct 1987</td>
<td>N</td>
<td>0700</td>
<td>475</td>
<td>CRS</td>
<td></td>
</tr>
<tr>
<td>Kaslof, AK 504425</td>
<td>19 Mar 1952</td>
<td>N</td>
<td>1800</td>
<td>24</td>
<td>CRS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5 Nov 1954</td>
<td>Y</td>
<td>1800</td>
<td>24</td>
<td>CRS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 Jun 1966</td>
<td>Y</td>
<td>1900</td>
<td>23</td>
<td>CRS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>21 Jan 1969</td>
<td>N</td>
<td>1930</td>
<td>23</td>
<td>CRS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 Oct 1972</td>
<td>N</td>
<td>1930</td>
<td>23</td>
<td>CRS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 May 1973</td>
<td>Y</td>
<td>1930</td>
<td>23</td>
<td>CRS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>29 Apr 1977</td>
<td>Y</td>
<td>0630</td>
<td>21</td>
<td>CRS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>26 May 1983</td>
<td>Y</td>
<td>0630</td>
<td>21</td>
<td>CRS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7 Aug 1984</td>
<td>N</td>
<td>0700</td>
<td>21</td>
<td>CRS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5 Jun 1986</td>
<td>N</td>
<td>0700</td>
<td>21</td>
<td>CRS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>28 Oct 1986</td>
<td>Y</td>
<td>0700</td>
<td>21</td>
<td>CRS</td>
<td></td>
</tr>
</tbody>
</table>
Fig. 6. Results of single-linkage cluster analysis of trends in the 95th percentile of $T_{\text{max}}$ ($T_{\text{max},95}$) during summer months from 1948 to 2000 across North America: (a) Kasilof, AK (station number: 504425), and Baker Lake, Nunavut (station number: 2300500), were the two locations identified as being anomalous (solid circle, other station locations used are indicated by the plus); and (b) three-dimensional stem plot of $T_{\text{max},95}$ trends [°C (50 yr)$^{-1}$] during Jun, Jul, and Aug at rogue stations (solid circle) and other locations in North America (open circles).
the rogue nature of this station is due to genuinely unusual trends in this region. The station in Alaska (Kasilof), conversely, had trends in $T_{\text{max},95}$ that were consistently large and negative during June, July, and August. Although a few other stations had large negative trends during all 3 months, none of these stations were in Alaska (all of other Alaska stations had trends that were weak or positive during these months). The July and August trends in $T_{\text{max},95}$ at Kasilof are particularly distinct from those at other stations in Alaska.

Inspection of the $T_{\text{max},95}$ time series at Kasilof shows that it was consistently among the warmest of its three nearest neighbors from 1950 through the mid-1970s, when it became consistently colder (Figs. 7a–c). Examination of the station record at Kasilof shows that the station moved 4.8 km northwest in 1977 (Table 1) to a location that was within 100 m of Cook Inlet. This change likely exposed the Kasilof station to more frequent and stronger cold-air advection from Cook Inlet during the warm months, resulting in lower daytime high temperatures during the summer at this station. Specifically, there was a decrease in the frequency of very high $T_{\text{max}}$ values at the Kasilof station. The change in observation time from evening to morning at Kasilof in 1977 likely also contributed to this change. Evening observation times typically produce higher monthly mean temperatures, although observation time differences are less pronounced during summer than in other seasons (Karl et al. 1986).

b. Sensitivity to clustering method

The results of the single-linkage clustering for winter month $T_{\text{min},5}$ and summer month $T_{\text{max},95}$ trends confirm the expected characteristics of that method. Single-linkage clustering tends to produce one large cluster, with secondary clusters containing small numbers of

![Fig. 7. Time series of $T_{\text{max},5}$ during Jun, Jul, and Aug at Kasilof, AK, and at its three nearest neighbors.](image)
Fig. 8. Results of average-linkage cluster analysis of trends in the fifth percentile of $T_{\text{min}}$ ($T_{\text{min5}}$) during winter months from 1948 to 2000 across North America: (a) 20 stations in three clusters were identified as being anomalous (solid circles, other station locations used are indicated by the plus), (b) three-dimensional stem plot of $T_{\text{min5}}$ trends [$^\circ$C (50 yr)$^{-1}$] during Dec, Jan, and Feb at rogue stations (solid circle) and other locations in North America.
elements. For both $T_{\text{min},5}$ and $T_{\text{max},95}$ trends, single-linkage clusters distinguished a handful of rogue stations. To evaluate the sensitivity of the results to the clustering method, we also applied average linkage clustering to the same winter month $T_{\text{min},5}$ and summer month $T_{\text{max},95}$ trends that were used in the single-linkage analysis.

For average linkage analysis of $T_{\text{min},5}$ trends, graphical evaluation of the linkage distance between successive clustering stages (Fig. 3b) suggests a four-cluster solution. The $T_{\text{min},5}$ average linkage rogue clusters contain all of the stations that are identified by single-linkage clustering, as well as 14 other stations. Unlike the single-linkage analysis, which produced a number of single-station rogue clusters, average linkage analysis produced larger rogue clusters that were not spatially isolated (Fig. 8). Because the rogue stations that are identified by average linkage are proximate to other rogue stations, it is likely that these trends are due to regionally coherent, yet distinct, climatic variations. The five-cluster solution for average linkage analysis of the $T_{\text{max},95}$ trends (not shown) produced a similar result, with 40 stations that are identified as rogue and many of these stations being spatially proximate.

By forcing all of stations in the rogue clusters to have somewhat similar trends, average linkage clustering likely includes stations in the rogue cluster that have large, but not unusually large, trends. Average linkage clustering produces larger clusters that are more spherical because the mean distance between clusters is the distinguishing criteria. Single-linkage clustering, on the other hand, has a chaining tendency—most stations tend to be associated with a large existing cluster rather than form new clusters. Single-linkage rogue clusters, therefore, can identify multiple stations that have unusually large trends, but that are not necessarily similar to one another. All of the stations identified by single-linkage clustering were along the periphery of the percentile trend space, but were not required to be similar to one another.

6. Summary and conclusions

The goal of this research was to develop a method for identifying “rogue” air temperature stations—locations with unusually large trends. Trends in the tails of air temperature frequency distributions were examined, because instrumentrom problems and spatially unrepresentative local climates are more apparent in air temperature extremes. Specifically, trends in the 5th percentile of $T_{\text{min}}$ and the 95th percentile of $T_{\text{max}}$ were analyzed. Cluster analysis was used to identify stations that were distinct from other locations. Both single- and average linkage clustering were evaluated.

Single-linkage cluster analysis appears to produce better results. Average linkage clustering tends to identify numbers of stations with similar trends, while single-linkage clustering identifies unusual stations from the entire periphery of the percentile trend space. Spatial isolation of a rogue station, however, may simply be indicative of (genuine) regional climatic variability that is not detectable in more distant stations. Most of the stations that are identified by single-linkage clustering, however, are not isolated spatially; they are close to numerous other stations that were not grouped into the rogue clusters, making it unlikely that the unusually large temperature trends were due to regional climatic variations. As with all approaches to evaluating data homogeneity, inspection of time series and station history records also is necessary to fully evaluate a station’s record. While the approach used here evaluated and compared trends across all of North America, it is likely that a more regional application also would be useful. Trends in air temperature percentiles at a regional scale should be relatively consistent, although boundary biases may become more important at the regional scale (Fovell and Fovell 1993). Nonetheless, if trends are not consistent across the region, the most unusual stations can be further analyzed for inhomogeneities.

An advantage of the method described here is the use of low-probability percentiles of daily data to detect subtle influences on historical air temperature records. Particularly as new daily data are ingested by the global climate community [e.g., the pre-1948 cooperative data in the United States, as well as the National Climatic Data Center’s (NCDC’s) Global Daily Climatology Network], the methods presented here should be a useful addition to existing homogeneity methods, which were primarily developed for monthly and annual data. Daily air temperature data increasingly are important for climate change detection (Robeson 2002a, 2004); therefore, improved methods for evaluating the quality of daily climatic data are essential.

Acknowledgments. This research is based upon work supported by the National Science Foundation (NSF) under grant 0136161. Useful comments on this paper by Justin Schoof, Francis Zwiers, and two anonymous reviewers are appreciated.

REFERENCES


Lanzante, J. R., 1996: Resistant, robust and non-parametric techniques for the analysis of climate data: Theory and examples, including applications to historical radiosonde station data. *Int. J. Climatol.,* 16, 1197–1226.


