Delineation and Investigation of Temperature-based Climate Regions in the South-Central United States

Gregory S. Bohr
California Polytechnic State University, San Luis Obispo

Homogeneous temperature-based climate regions are delineated for a five-state area of the south-central United States using multivariate cluster analysis. The variables used in the clustering are the monthly means of maximum and minimum temperature as well as the inter-diurnal variability of these temperatures. Both raw data and rotated PCA component scores are used for clustering. Two clustering techniques – Ward's method and average linkage – are compared, and average linkage is found to produce the most meaningful results. Based on multiple clustering solutions, seven regions are defined for the study area. Multivariate ANOVA applied to an alternate temperature data set shows that the defined regions are significantly different from one another. The defined regions are compared to climate division boundaries for the study area. Most divisions are largely contained within a single climate region, but nine divisions across Texas, Oklahoma, and Arkansas are split between multiple zones. Finally, the applicability of these newly defined regions is demonstrated by identifying significant relationships between regional-mean maximum and minimum monthly temperatures and three important hemispheric-scale teleconnection indices. Significant correlations are found between temperature in the study area and both the Pacific/North American pattern and the North Atlantic Oscillation, mainly in winter and spring. The PNA is negatively correlated with maximum and minimum temperatures, while the positive phase of the NAO is associated with warmer conditions. Overall, the sub-regions defined here are believed to represent appropriate spatial units for future studies of temperature change and variability in the southern United States. *Key Words: climate regionalization, cluster analysis, surface temperature, teleconnections, southern United States.*

Introduction

Climate regionalization is the identification of regions – at the global scale or smaller – which can be said to be internally homogeneous; i.e., every place within the region has a similar climate. There are many motivations for climate regionalization, one of which is to facilitate the study of climate change and variability at regional and local scales. As global climate changes, local and regional climates frequently vary in ways that are quite different from the global pattern, as these smaller-scale climates are controlled by a variety of local and global forcings (McGuffie et al. 1999; Easterling et al. 2000; Stott et al. 2000). In order to study regional responses to climate change, it is helpful to have an empirically-defined region; i.e., a region that is internally homogeneous in climate and can be assumed to respond as a unit to variations in climate forcings. For example, regions of this nature are beneficial when downscaling to regional climates from global circulation model scenarios.

Regionalization has a long history within the disciplines of geography and climatology; researchers around the world have defined and classified homogeneous climate regions at all spatial scales using a variety of methods and variables (e.g. Stooksbury and Michaels 1991; Bunkers, Miller, and DeGaetano 1996; Fovell 1997; Feddema 2005; Pineda-Martinez, Carbajal, and Medina-Roldan 2007). For example, the well-known Köppen system defines generalized climate types based on monthly and annual means of temperature and precipitation, while the Thornthwaite system produces climate types based on the regional water balance. These empirical methods classify regions based on (dis)similarities in observed data, but their outcomes are determined to some extent by *a priori* classification rules that are independent of the data being classified. Deterministic methods like cluster analysis can be used to define regions based entirely on observed values, without any pre-defined thresholds to separate classes.

Cluster analysis has been used frequently in the atmospheric sciences, most often for one of two general purposes: synoptic classification (e.g. Kalkstein, Tan, and Skindlov 1987; Davis and Kalkstein 1990; Cheng and Wallace 1993; Santos, Corte-Real, and Leite 2005) and climate regionalization (e.g. Briggs and Lemin 1992; DeGaetano 1996; Coronato and Bisigato 1998; Whitfield, Bodtker, and Cannon 2002; Unal, Kindap, and Karaca 2003). The basic objective of cluster analysis is to identify subsets of a set of objects (in this case weather stations) grouped in such a way as to maximize the similarity within the groups while simultaneously maximizing between-group dissimilarity. The goal is to identify any underlying structure (or spatial pattern, in the case of regionalization) existing in the dataset, which may then provide a framework for further investigation (Gong and Richman 1995; SAS Institute 1999).

The study area addressed here is the south-central United States, comprising the states of Arkansas, Louisiana, Mississippi, Oklahoma, and Texas. While not as topographically complex as the mountain west, for example, this region is climatically variable due to its proximity to the Gulf of Mexico, higher elevations to the west and north, and a strong east/west gradient in temperature and moisture owing to the effects of continentality and the relative locations of general circulation features such as the Atlantic High. While studies of this nature can be carried out anywhere, the region addressed here is chosen because it is an important producer of a variety of agricultural commodities, including several types of livestock, and the implications of changed climate could be severe, particularly in marginal agricultural areas (Parry and Carter 1985). Furthermore, the study area addressed here is of interest due to its unique vulnerability to a wide range of meteorological threats, including extreme heat and cold, as well as more dramatic events such as tropical storms, tornadoes, severe thunderstorms, and winter ice storms. While various climate regionalizations have been performed for the United States as a whole (e.g. DeGaetano 2001; Fovell 1997) and for specific sub-regions (e.g. Bunkers, Miller, and DeGaetano 1996; DeGaetano 1996; Rhee et al. 2008), this project is believed to be the first at this scale specifically for the south-central U.S.

This study uses two different methods of cluster analysis to delineate distinct temperature-based climate regions in the study area, and compares them to the commonly used climate divisions. Once the regions have been delineated, relationships between surface temperatures in the south-central United States and upper-level flow patterns are assessed by relating monthly mean maximum and minimum temperatures for the defined regions to three teleconnection indices that have been shown to be particularly important in the Northern Hemisphere: the El Nino/Southern Oscillation (ENSO, represented by the Southern Oscillation Index (SOI)), the Pacific/North American pattern (PNA), and the North Atlantic Oscillation (NAO). This correlation analysis provides a simple example of how the defined regions may be applied in the investigation of climate variability in the study area.

Data and Methods

Data Sources

To precisely identify the boundaries of climate regions, a dense network of observation sites is beneficial. This study uses the TD3200 dataset, an extremely dense network of weather stations in the U.S., with approximately 8000 currently open sites, the majority of which are volunteer-staffed sites from the National Weather Service (NWS) cooperative program. The TD3200 data have undergone manual and automated quality control to detect spurious values; however, the data have not been adjusted for non-climatological factors that may introduce inhomogeneity (Reek, Doty, and Owen 1992; NCDC 2006).

An source of concern with the TD3200 data is that the time of daily observation (7:00 am or 7:00 pm at most sites) may bias the data by carrying over a particularly low minimum temperature (am observers) or a high maximum temperature (pm observers) from one day to the next. This error would have the effect of erroneously reducing the interdiurnal temperature variabilities used here. Although this potential problem is troubling, the benefit of a dense observation network was deemed to outweigh the risk, and it is hoped that the long averaging period minimized any small biases present in individual months.

To obtain a spatially representative sample of sites in the study region, a stratified-random selection process was applied. The five-state area was divided into $1^{o}x1^{o}$ latitude/longitude grid cells, and two sites were selected at random from each cell. For each site, 50 years worth (1948-1997) of daily maximum and minimum temperatures were obtained. The data from each site were manually screened for obviously erroneous values, and these values were removed and marked as missing. Next, the time series were examined for overall completeness, and any site missing more than 5% of the total number of days was rejected and replaced by another site from the same grid cell. A month was marked as missing when more than 10% of its daily observations were missing, and sites with fewer than 95% of the total months in the study period non-missing were replaced from the same grid cell, if another site was available. The result of this process was a set of 184 sites with 50-year time series of maximum and minimum daily temperatures. These sites are shown in Figure 1, along with the climate division boundaries for the five states.

A complete understanding of temperature patterns requires both a measure of the typical value (mean) as well as a measure of how variable temperature is around that mean. In order to capture as much temperature information as possible in the regionalization, four monthly variables were calculated over the study period for each site: monthly mean maximum temperature, monthly mean minimum temperature, mean interdiurnal variability of maximum temperature, and mean interdiurnal variability of minimum temperature. The latter two variables were calculated by averaging the absolute differences between the maximum (minimum) temperatures on each day of the month and the day immediately preceding. The interdiurnal variability is used because it provides insight into the day-to-day (in)consistency of local weather – these short term variations will result from high-frequency synoptic events such as cold front passages and are likely to be highly related to health, energy, and ecological impacts (Driscoll, Rice, and Yee Fong 1994; Williams and Parker 1997). Lastly, long-term means for each of the four variables were calculated by month at each site. This resulted in a total of 48 values (12 months x 4 variables) at each of the 184 sites to be input into the cluster analysis. Each of these long-term means was then standardized to have a mean of zero and a standard deviation of one, to ensure that all variables had the same weight in the cluster analysis.

A significant difficulty in cluster analysis is testing whether the resultant regions are significantly different from one another. To facilitate this testing, a

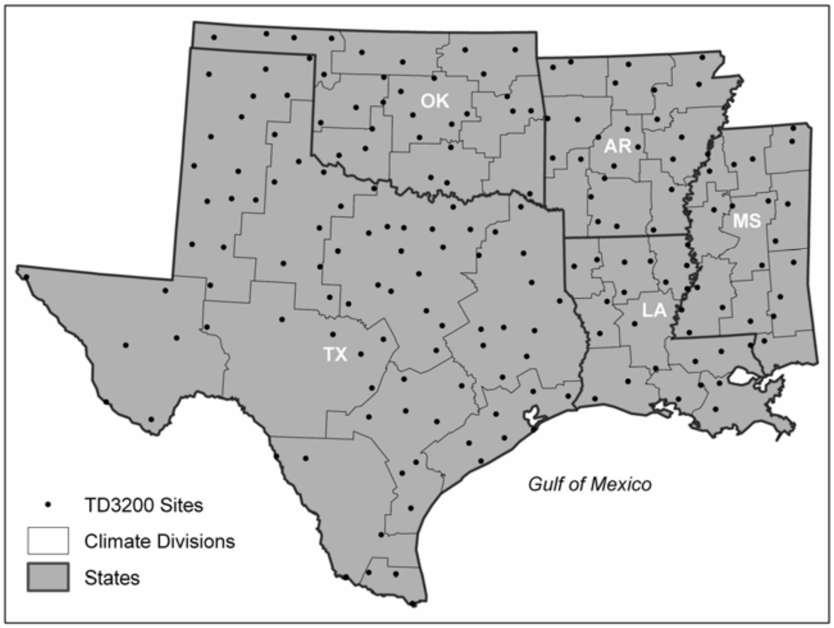

Figure 1. Distribution of 184 TD3200 observation sites used in clustering, along with boundaries of climate divisions in the study region.

second dataset was used for cluster validation: the U.S. Daily Historical Climatology Network (HCND), made up of sites selected for low potential for heat island impacts and a high level of completeness and homogeneity. Sites in the HCND have undergone extensive automated and manual quality control to detect and correct erroneous values (Easterling et al. 1999). Daily data from 1948-2001 for HCND sites in the study region were obtained and subjected to the same completeness tests as the TD3200 data, resulting in 89 sufficiently complete HCND sites. At each of these sites, long-term means of monthly maximum and minimum temperatures as well as the intra-monthly standard deviations of both maximum and minimum temperatures were calculated, producing a set of 48 variables at each site, which is comparable to the TD3200-based data used for the clustering.

Temperature data for the correlation analyses were obtained from the monthly U.S. Historical Climatology Network (HCN), which was developed to aid studies of regional climatic variability and/or climate change. Missing values in the HCN data have been estimated using nearest neighbors, resulting in station time series that are as serially complete as possible. Unlike the TD3200 and HCND, the HCN data are adjusted for a variety of non-climatic inhomogeneities – time of observation bias, station moves, instrument changes, and instrument moves – as well as for the effects of urbanization (Easterling et al. 1996). All available HCN sites in the study area that had complete records for 1950-2000 were used. For each of the sub-regions defined in the cluster analysis, data from all sites with complete and uninterrupted records were averaged to create regional temperature time series of monthly maximum and minimum temperature for the study period.

Each of the three temperature datasets used here has a slightly different period of record: 1948-1997 for the TD3200 data used to create the regions, 1948-2001 for the HCND data used to validate the regions, and 1950-2000 for the monthly HCN data used in the teleconnection analyses. While the temporal inconsistencies may seem odd, they do not negatively affect the results – it is beneficial to validate the regions with a different data set than that used in the clustering (Aldenderfer and Blashfield 1984), and a different time period is immaterial and perhaps even beneficial in this process. Likewise, if the regions are truly coherent climate zones, relationships between regional climate and the circulation indices should be apparent over time periods other than that of the clustering data; thus the inexact overlap with the HCN data (and the teleconnection data) can be ignored.

Clustering Methods

This study uses agglomerative hierarchical clustering methods, in which each observation starts as its own cluster. The clustering algorithm calculates the distance between all pairs of clusters using the selected distance metric, and joins the most similar pair. This process repeats until all sites are grouped into one cluster. From there, the researcher backtracks, breaking apart joined clusters until the appropriate clustering level is reached (Aldenderfer and Blashfield 1984). The measure of distance (similarity) between sites is the squared Euclidean distance, in which each site, described by *n* variables, is treated as a point in an *n*-dimensional space (Gong and Richman 1995; Fovell 1997).

The two clustering algorithms used here are average linkage and Ward's method, which are common in climatological applications. In the average linkage method, the distance between a pair of clusters is the average of the squared Euclidean distances between all possible pairs of points with one in each cluster. This method minimizes within-cluster variance and maximizes between-cluster variance, and has a slight bias toward producing clusters with the same within-cluster variance. In Ward's method, the total within-cluster sum of squares is minimized at each iteration. This method is biased toward producing equal-sized clusters (Aldenderfer and Blashfield 1984; Kalkstein, Tan, and Skindlov 1987; Stooksbury and Michaels 1991).

Typically, all variables should have equal weight in the clustering process, and standardization of the data is often done to ensure this. However, when input variables are interrelated, the redundant information increases the weight of the interrelated variables to an unknown degree (Fovell and Fovell 1993). In order to remove redundant information from the data set, principal components analysis (PCA) is frequently used before the clustering (e.g. Davis and Kalkstein 1990; Briggs and Lemin 1992; DeGaetano 1996). PCA linearly transforms the original data, producing a reduced set of uncorrelated variables (component scores) that explain most of the variability in the original data. If the correct number of components is retained, most of the redundant information will be removed, while the 'good' information in the data is retained (Horel 1981; Dunteman 1989; Jolliffe 1990; White, Richman, and Yarnal 1991).

The 48 long-term mean temperature values described above are highly intercorrelated, so a rotated PCA using the varimax method was applied to reduce the dataset. The first five components, explaining 96% of the total variation, were retained for analysis. Component scores for each site were calculated, and used as an alternative data set for clustering.

As there is no significance value produced by cluster analysis to determine the optimum solution, a major question is the determination of the 'correct' number of clusters. Fuzzy boundaries and overlap between clusters are to be expected due to both the continuous spatial variation of climate data and the inability of hierarchical methods to reassign poorly-clustered observations (Fovell and Fovell 1993; DeGaetano 1996). Kalkstein (1987) recommends using the squared multiple correlation (R^2), which gives the proportion of total variance explained by the current clustering. It ranges from 1.00 when all observations are their own clusters down to 0.00 when all observations are in a single cluster. A relatively large drop in R^2 from one step in the clustering to the next indicates that two dissimilar clusters have been forced together, meaning that the solution just before the drop may be an appropriate stopping point. Two other statistics useful in selecting the appropriate number of clusters are the pseudo-F and pseudo-T^2. The pseudo-F is the ratio of between-cluster variance vs. within-cluster variance, and the pseudo-T^2 is the ratio of the within-cluster sum of squares for two clusters vs. the within-cluster sum of squares for the one cluster that results from their joining. These statistics are calculated for each step in the hierarchical clustering procedure, and potentially appropriate clustering levels are indicated by local maxima of the pseudo-F, and/or small values of the pseudo-T^2 that are followed by peaks (Stooksbury and Michaels 1991; Fovell and Fovell 1993; SAS Institute 1999). In most cases, it is appropriate to look for agreement between several of these indicators, in conjunction with subjective interpretation of the spatial pattern produced by a particular clustering solution, to determine the 'correct' solution.

Results

Defined Climate Regions

Four separate clustering solutions were produced: average linkage and Ward's method, each applied to both the raw data and the PCA component scores. Potentially appropriate clustering levels were determined using the three measures described above; Table 1 lists the most likely numbers of clusters from each solution, and the cluster analysis statistics that suggest each clustering level. Overall, much stronger agreement exists between the results of the average linkage approach, as only this approach led to clear agreements on appropriate clustering levels between all three indicators. Considering all four solutions together, certain clustering levels are suggested, particularly in the range of 10-12 clusters. Additionally, the two average linkage solutions agree on suggesting a solution in the mid-teens; i.e. 15-17 clusters.

Table 1. Potentially appropriate clustering levels based on the four clustering solutions. The column labeled "Indicators" lists which of the cluster analysis statistics supports the given number of clusters for that solution.

Cluster Solution	Number of clusters	Indicators
Raw data, average linkage	7	all three
Raw data, average linkage	10	all three
Raw data, average linkage	17	all three
PCA scores, average linkage	12	all three
PCA scores, average linkage	15	all three
PCA scores, average linkage	20	all three
Raw data, Ward's Method	10	T^2, R^2
Raw data, Ward's Method	12	T^2, R^2
Raw data, Ward's Method	14	T^2
PCA scores, Ward's Method	10	T^2, R^2, F
PCA scores, Ward's Method	12	T^2, R^2, F
PCA scores, Ward's Method	21	T^2

Distinct patterns can be seen when a clustering level of 15 clusters is selected. Figure 2 shows the 15-cluster average linkage solution using the PCA scores, with coherent clusters are apparent in western Oklahoma (6) and central Texas (3). In the east, the coastal and inland regions (1, 2) are clearly separated, and the panhandles of Texas and Oklahoma are identified as a distinct region (4). The fact that the most immediately coastal locations – Hackberry, LA (10), Galveston, TX (15), Matagorda, TX (8) and Brownsville, TX (8) – are delineated is reassuring, as this level of small-scale differentiation is indica-

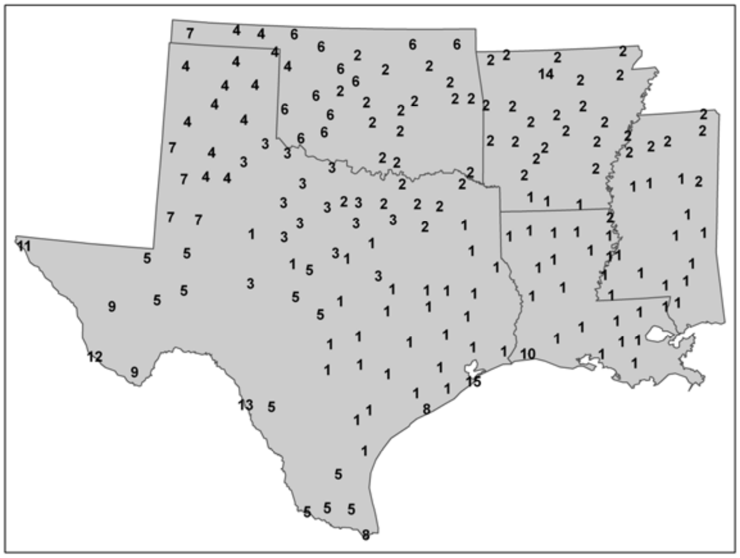

Figure 2. 15-cluster solution obtained from the PCA scores with the average linkage technique. The numbers have no significance other than to identify clusters, and will vary between maps for a given site.

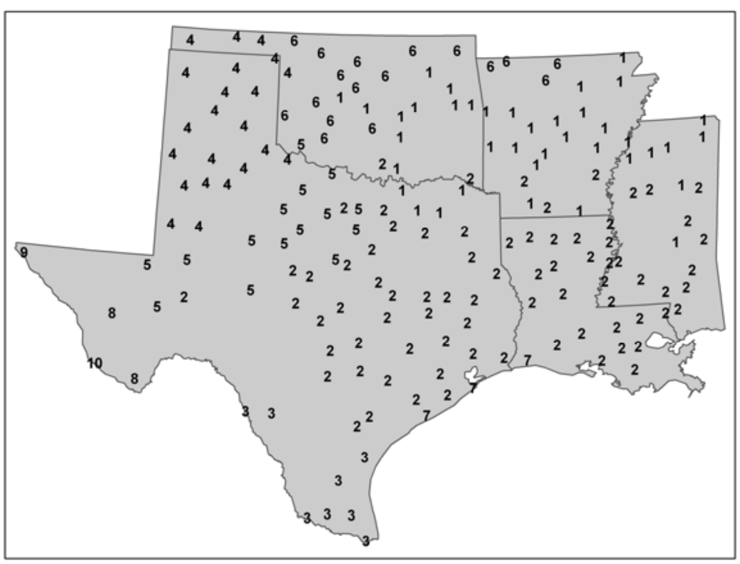

Figure 3. 10-cluster solution from the average linkage technique using the raw data.

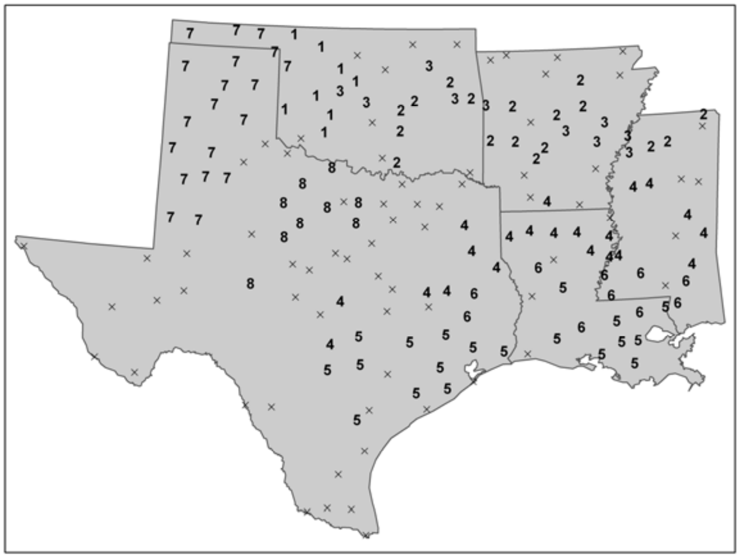

Figure 4. Sites that clustered together in the 10-cluster solutions of all four clustering methods. Only groups of eight of more sites are shown. The X symbol indicates sites that were not members of any consistent groupings.

tive of the power of the technique. Sites in the far west of Texas are climatological outliers by virtue of their isolation from the rest of the region, as well as by their elevational uniqueness. The two sites indicated as cluster 9 are the highest in the sample: Mt. Locke at 2076 meters and Chisos Basin at 1621 meters.

Moving to higher degrees of agglomeration, the 10-cluster average linkage solution with the raw data (Figure 3) provides a coherent and climatologically-rational pattern, with the least overlap of any solution at this level. The Panhandles appear as a solid cluster (4), which is believed to be particularly robust as it appeared earlier in the agglomeration process. The eastern half of the study region is divided into a southern (2) and northern (1) half, while three of the near-coastal sites appear as a distinct micro-cluster (7). The deep-south Texas (3), western Oklahoma (6) and central Texas (5) regions are also apparent. However, even at this level of agglomeration there is overlap between clusters as well as a small number of un-clustered sites, which requires some subjectivity in the final cluster delineation. To guide in the final clustering, sites which clustered together in each of the 10-cluster solutions from all four approaches are shown in Figure 4.

The cores of several distinct clusters are apparent in Figure 4: the panhandles of Texas and Oklahoma (7), western Oklahoma (1), central Texas (8), and the gulf coast of Texas and Louisiana (5). The eastern half of the study region shows a south-north gradient in climate, with overlapping clusters, particularly in the northeast (Arkansas, eastern Oklahoma, and northern Mississippi). Using Figure 4 as a guide, sites were assigned to clusters based on the 10-cluster, raw data, average linkage solution. The eastern half of the study area is divided into two regions, the Gulf Lowland and the Eastern Highland, with the boundary near the Arkansas-Louisiana border. (The region names assigned here are informal, and specific to this project.) The three distinct coastal sites (Hackberry, Galveston, and Matagorda) are subsumed into the larger Gulf Lowland sub-region. A Deep South (Texas) region is delineated, extending from Corpus Christi across to Eagle Pass. The Western Oklahoma region extends eastward across the northern border to encompass the slightly higher-elevation sites in northern Arkansas. The Panhandle and Central Texas clusters complete the main pattern, with the Central Texas cluster extending westward to include four sites in west Texas. Finally, the outliers of far western Texas are merged together in a seventh cluster. During this final clustering, 14 sites (7.6%) were manually reassigned to a neighboring cluster to remove overlap. The entire study area was then divided into these seven regions using Euclidean allocation in ArcGIS, as shown in Figure 5.

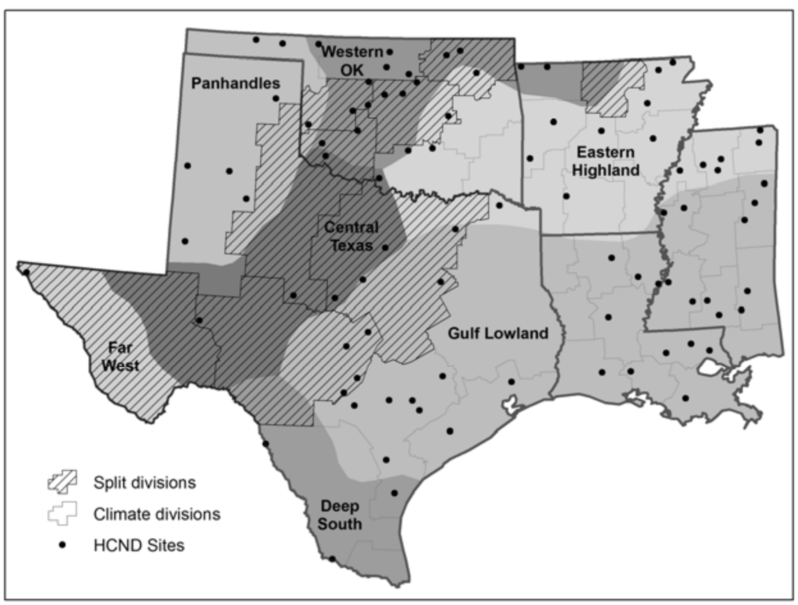

Figure 5. Final regionalization, showing HCND sites used in the cluster validation step as well as the climate divisions in the region. Climate divisions that are less than 80% contained in a single climate region are cross-hatched.

This final pattern is clearly influenced by the combined effects of elevation, latitude, and continentality. The higher elevations in the panhandles correspond well to a clearly-defined cluster, and the intermediate elevations in western Oklahoma and central Texas are also delineated into separate clusters. The eastern half of the study region is uniformly low in elevation, contributing to the difficulty in differentiation of these sites. In the eastern areas, latitude and distance from the Gulf (continentality) are the dominating factors. The western extremity of Texas is cutoff from the Gulf by distance and topography, and extends into the more arid climate zone of the southwestern U.S.

Cluster Validation

Cluster analysis produces no statistical output to assess the significance of the results. Often, the only assessment of whether a clustering is valid is a visual examination of the resulting regions. However, Aldenderfer and Blashfield (1984) suggest that cluster validity be assessed by formal statistical testing on independent data, i.e. data that were not used in the derivation of the clusters. To this end, multivariate analysis of variance (MANOVA) was applied to the HCND data described above to compare the six clusters with a sufficient number of sites (he Far West cluster, with only one acceptable HCND site, was dropped). MANOVA is a multivariate extension of univariate ANOVA, in which group means are compared across a number of variables simultaneously, to detect differences between classes that may not be apparent in individual variables (Johnson and Wichern 1998; SAS Institute 1999). The MANOVA output statistics indicated a statistically significant difference (p-value $<$ 0.0001) between the overall set of clusters, supporting the conclusion that the regions shown in Figure 5 are genuinely distinct climatological sub-regions of the south-central U.S.

Comparison with Climate Divisions

The next step in the project was to assess the degree to which the delineated regions agree with the climate divisions in the region. Climate divisions are frequently used as aggregation units for climate data, but their boundaries are not based on climatological variations, but rather on state and county borders, agricultural divisions, or topographical features such as watersheds. In addition to temperature, divisional-scale datasets include precipitation, heating and cooling degree days, and drought indices such as the PDSI and PHDI (Guttman and Quayle 1996). Figure 5 shows the division boundaries overlaid on the regions delineated in this project. Of the 47 divisions in the study area, 19 are completely contained within one of the regions, 24 are shared between two regions, and 4 are split between three regions. Overall, 38 of the divisions are at least 80% contained within a single region; given the fuzziness of the region boundaries it seems reasonable to assume that that degree of overlap implies a high degree of climatic homogeneity across the divisions. This includes all of the divisions in Louisiana (contained in the Gulf Lowlands region) and Mississippi (seven divisions in the Gulf Lowlands and the three northern divisions in the Eastern Highlands), and all but one division in Arkansas (all but the two northwestern AR divisions are in the Eastern Highlands). In contrast, nearly half of the divisions in Texas and Oklahoma are split between multiple climate regions (Figure 5), suggesting that these divisions are not suitable as climatological aggregation units (at least as far as temperature is concerned). Although this result is not surprising, give that the climate divisions are constructed based on more variables than temperature, the value of this finding is that temperature time series from divisions wholly contained within one of the temperature regions defined here are likely to be well-representative of the division as a whole. In contrast, where divisions are split between regions the implication is that multiple temperature-influencing factors are at work in the division, and thus division-wide temperature data may be suspect.

Correlations with Teleconnection Indices

The final step in this project was an exploration of the relationships between surface temperature in the defined regions and three hemispheric-scale circulation indices. For this analysis, monthly data for the PNA, SOI, and the NAO were obtained online from the Climate Prediction Center (CPC). Although sea surface temperature (SST) based indices such as Nino-3.4 are frequently used to detect the timing and magnitude of El Nino events, Hanley et al. (2003) found that the SOI was approximately equal in sensitivity. Each of the teleconnection index time series was detrended by regressing the monthly values on the time index (year) and retaining the residuals as the detrended series (Raffalovich 1994; Bell et al. 2000). For each teleconnection, the Shapiro-Wilks test indicated no significant departure from normality for the individual detrended monthly time series.

Regional-mean time series of monthly mean-maximum and mean-minimum temperature were constructed by averaging data from all HCN sites in each region with complete and uninterrupted records for the period 1950-2000. These time series were detrended in the same manner as the teleconnection time series. Again, the Shapiro-Wilks test indicated that the monthly series for both maximum and minimum regional-mean temperatures were normally distributed, allowing the use of the parametric Pearson's r correlation coefficient to assess the relationship between the indices and the monthly temperatures. These correlation values were calculated for the period 1950-2000, by month, between each sub-region's mean temperature (maximum and minimum, separately) and the monthly index value for each of the teleconnections.

The results of the correlation analyses are shown in Table 2 (maximum temperature) and Table 3 (minimum temperature). As the tables show, there are distinct seasonal and spatial differences in the degree to which the teleconnection indices explain the regional variations in monthly mean temperatures. Overall, it appears that, for both maximum and minimum temperatures, the large-scale atmospheric circulation as measured by these indices plays a larger role in temperature variation in the south-central U.S. during winter and spring, as opposed to summer. In addition, it appears that the NAO and the PNA indices best explain variations in monthly mean temperatures. Of the three teleconnections, the SOI shows relatively few significant correlations for maximum temperature and none for minimum, and the impacts of the SOI largely appear to be focused in the western and southern parts of the study area.

The Pacific/North America Pattern is a pattern of upper-level circulation characterized by, in its positive phase, ridging over the western portion of North America (approximately centered on the Rocky Mountains) and troughing over eastern North America. The negative phase of the teleconnection pattern is characterized by zonal flow over the continent or, in extreme cases, a reverse-PNA with troughing in the West and ridging over the eastern U.S. For the southern U.S., positive PNA events are strongly correlated with reduced temperatures, owing to the increased advection of cold northerly air into the eastern U.S. The strength of this relation increases (i.e. is more negative) toward the southeast corner of the country (Leathers, Yarnal, and Palecki 1991; Henderson and Robinson 1994). In the South, the PNA pattern is therefore likely to be related to increases in the frequency of extreme cold daily temperatures, as was found by Rogers and Rohli (1991), who showed that positive PNA events are strongly correlated with damaging freeze events in Florida.

In this study, the PNA pattern shows strong negative correlations with both maximum and minimum monthly temperatures throughout winter and early spring, although this relationship is weaker for the Panhandles and Western Oklahoma sub-regions; a pattern similar to that of Leathers et al (1991). The likely reason for this pattern is that the trough over the central and eastern U.S. that characterizes the positive phase of the PNA pattern is responsible for advection of cold northerly air into the region whenever the pattern is in place. Under these circumstances, months dominated by the PNA pattern would experience increased cold air influx. Interestingly, the importance of the PNA declines in February and March before reasserting itself in April, especially over the eastern sub-regions. This intermonthly shift in importance is probably explained by the variations in the mean positions of the ridge and trough from month to month; the precise atmospheric pattern described by the PNA index varies somewhat throughout the winter (Barnston and Livezey 1987). Another interesting pattern detected here is that the PNA is associated with increases in early spring minimum temperatures in the northern sub-regions (Panhandles and Western Oklahoma.)

Table 2. Pearson's r correlations between detrended teleconnection indices and monthly-mean maximum temperatures for 1950-2000. Entries in bold are significant at the 0.05 level. The PNA is not a major mode of circulation in June and July, and is not defined for those months.

	December			January			February		
	NAO	PNA	SOI	NAO	PNA	SOI	NAO	PNA	SOI
Gulf Lowland	**0.33**	**-0.57**	0.19	**0.44**	**-0.54**	0.24	**0.49**	-0.22	0.2
Eastern Highland	**0.35**	**-0.4**	0.1	**0.37**	**-0.31**	0.11	**0.47**	-0.1	0.13
Western OK	**0.32**	-0.21	0.18	**0.32**	-0.12	0.1	**0.5**	-0.04	0.15
Panhandles	0.21	-0.11	**0.33**	**0.31**	-0.18	0.17	**0.55**	-0.03	0.25
Central TX	0.2	**-0.34**	**0.3**	**0.35**	**-0.32**	0.19	**0.51**	-0.08	0.18
Deep South	0.23	**-0.51**	**0.37**	**0.35**	**-0.55**	0.21	**0.38**	-0.09	0.19
Far West	0.02	**-0.31**	**0.38**	**0.38**	**-0.6**	0.27	**0.44**	-0.07	0.26

	March			April			May		
	NAO	PNA	SOI	NAO	PNA	SOI	NAO	PNA	SOI
Gulf Lowland	**0.44**	-0.15	0.16	0.22	**-0.33**	0.15	0.2	-0.25	0.18
Eastern Highland	**0.51**	-0.02	0.06	**0.35**	**-0.29**	0.09	0.21	-0.14	-0.07
Western OK	**0.54**	-0.04	0.1	**0.35**	**-0.3**	0.16	**0.28**	-0.09	0.07
Panhandles	**0.58**	-0.07	0.19	**0.34**	**-0.34**	0.25	**0.29**	-0.13	0.17
Central TX	**0.56**	-0.07	0.2	0.23	-0.24	0.17	0.22	-0.17	**0.33**
Deep South	**0.31**	-0.14	0.23	-0.08	-0.05	0.18	0.02	**-0.35**	**0.32**
Far West	**0.55**	-0.11	**0.4**	0.26	-0.26	**0.33**	0	-0.24	**0.33**

Table 2. Continued.

	June			July			August		
	NAO	PNA	SOI	NAO	PNA	SOI	NAO	PNA	SOI
Gulf Lowland	**-0.31**	.	-0.04	**-0.39**	.	-0.03	-0.26	-0.21	-0.09
Eastern Highland	-0.15	.	-0.05	-0.26	.	0.03	-0.14	-0.27	0.01
Western OK	-0.03	.	0.03	-0.14	.	0.04	0	-0.22	0.04
Panhandles	-0.05	.	0.16	-0.1	.	-0.08	-0.01	-0.22	0.04
Central TX	-0.09	.	0.04	-0.18	.	0	-0.22	**-0.28**	-0.09
Deep South	-0.08	.	0.09	-0.21	.	0	**-0.4**	-0.15	-0.27
Far West	-0.16	.	0.06	-0.02	.	**-0.31**	-0.14	-0.24	-0.23
	September			October			November		
	NAO	PNA	SOI	NAO	PNA	SOI	NAO	PNA	SOI
Gulf Lowland	-0.02	**-0.34**	-0.23	-0.01	-0.21	0.03	-0.09	**-0.67**	0.13
Eastern Highland	-0.11	**-0.32**	-0.08	0.11	-0.24	0.08	-0.06	**-0.49**	0.1
Western OK	-0.15	-0.23	-0.09	-0.13	-0.15	0.01	-0.09	**-0.34**	0.12
Panhandles	-0.1	-0.16	-0.14	-0.23	-0.14	0.09	-0.17	**-0.32**	0.16
Central TX	-0.05	**-0.34**	-0.17	-0.19	-0.15	-0.01	-0.19	**-0.49**	0.12
Deep South	0.01	**-0.36**	**-0.32**	-0.2	-0.07	0	-0.18	**-0.54**	0.19
Far West	0.01	-0.18	-0.03	-0.21	0.04	-0.05	-0.15	**-0.56**	0.2

Table 3. Pearson's r correlations between detrended teleconnection indices and monthly-mean minimum temperatures for 1950-2000. Entries in bold are significant at the 0.05 level.

	December			January			February		
	NAO	PNA	SOI	NAO	PNA	SOI	NAO	PNA	SOI
Gulf Lowland	**0.36**	**-0.56**	-0.05	**0.42**	**-0.6**	0.07	**0.37**	**-0.31**	0.04
Eastern Highland	**0.43**	**-0.44**	-0.03	**0.41**	**-0.45**	-0.05	**0.37**	-0.11	-0.03
Western OK	**0.42**	-0.24	-0.02	**0.29**	-0.22	-0.22	**0.36**	0.15	-0.11
Panhandles	**0.39**	-0.05	0.06	**0.29**	-0.14	-0.17	**0.33**	0.26	-0.12
Central TX	**0.36**	**-0.28**	0	**0.32**	**-0.34**	-0.18	**0.36**	0.1	-0.08
Deep South	0.22	**-0.35**	0.03	**0.35**	**-0.52**	-0.07	**0.42**	-0.19	0.05
Far West	0.19	**-0.46**	0.03	**0.34**	**-0.37**	-0.1	**0.36**	0.11	-0.09
	March			April			May		
	NAO	PNA	SOI	NAO	PNA	SOI	NAO	PNA	SOI
Gulf Lowland	**0.37**	-0.13	0.07	0.16	**-0.51**	0.12	0.15	-0.19	-0.08
Eastern Highland	**0.44**	0.04	-0.02	**0.28**	**-0.45**	0.08	0.18	-0.23	-0.06
Western OK	**0.48**	**0.28**	-0.07	**0.34**	**-0.38**	0.17	0.25	-0.19	0.03
Panhandles	**0.48**	**0.32**	-0.07	**0.44**	**-0.33**	0.21	**0.29**	-0.14	0
Central TX	**0.48**	0.13	0.02	0.26	**-0.52**	0.24	0.25	-0.2	0.12
Deep South	**0.37**	-0.08	0.16	0.07	**-0.55**	0.26	0.07	-0.26	0.19
Far West	**0.45**	0.11	0.1	**0.28**	**-0.44**	0.24	0.08	**-0.29**	0.17

Table 3. Continued.

	June			July			August		
	NAO	PNA	SOI	NAO	PNA	SOI	NAO	PNA	SOI
Gulf Lowland	**-0.33**	.	0.01	-0.2	.	0.06	**-0.32**	-0.26	0.02
Eastern Highland	-0.21	.	0.08	-0.18	.	0.08	-0.25	**-0.39**	0.07
Western OK	-0.11	.	0.03	-0.17	.	0	-0.13	**-0.39**	-0.03
Panhandles	-0.06	.	0.03	-0.13	.	-0.04	-0.05	-0.25	-0.13
Central TX	-0.16	.	0.02	-0.23	.	0.04	**-0.3**	**-0.29**	-0.09
Deep South	**-0.28**	.	0.03	-0.09	.	0.04	**-0.4**	-0.06	0
Far West	-0.11	.	0.05	-0.1	.	-0.21	**-0.38**	-0.11	-0.23

	September			October			November		
	NAO	PNA	SOI	NAO	PNA	SOI	NAO	PNA	SOI
Gulf Lowland	-0.06	**-0.49**	0.06	0.14	-0.17	-0.02	-0.1	**-0.63**	-0.11
Eastern Highland	-0.05	**-0.45**	0.03	0.18	-0.27	0.12	-0.1	**-0.57**	-0.1
Western OK	-0.09	**-0.36**	-0.06	0.11	**-0.34**	0.11	-0.04	**-0.41**	-0.03
Panhandles	-0.06	-0.12	-0.16	0.13	**-0.32**	0.05	-0.08	**-0.34**	-0.02
Central TX	-0.09	**-0.34**	-0.08	0.15	**-0.35**	0.05	-0.14	**-0.5**	-0.07
Deep South	-0.09	**-0.33**	-0.05	0.08	-0.12	-0.05	-0.25	**-0.62**	-0.04
Far West	-0.02	-0.27	-0.12	0.21	**-0.29**	-0.04	-0.15	**-0.52**	-0.12

The NAO has previously been shown to have significant impacts on surface temperature conditions over large parts of North America, particularly the East. During the positive phase of the oscillation, anomalous southerly flow brings warmer than normal conditions to the eastern and southern U.S, under the influence of the enhanced Atlantic High. The negative phase is associated with lower temperatures in the Eastern U.S. (Hurrell 2000; Sheridan 2003)

As Tables 2 and 3 indicate, the NAO index shows strong positive correlations with regional monthly-mean maximum and minimum temperatures in the study area during winter. This winter warming during the positive phase of the NAO may relate to increased southerly advection into the study area due to the above normal SLP across the central North Atlantic, although processes related to the larger-scale Arctic Oscillation (with which the NAO is intertwined) may also play a role. During summer, the relationship, although weaker, appears to be reversed, with negative correlations particularly with minimum temperature. Spatially, the NAO first becomes significant for maximum temperature in the eastern sub-regions (in December), before reaching significance in all of the study area. Conversely, the significant positive relationship with maximum temperature lingers later into the year in the northernmost sub-regions. Conceivably, the NAO impact fades along the coast during summer because these areas would be subject to moist maritime air with or without the enhanced southerly flow; however, during positive NAO seasons, this flow extends further inland and is directed further west by the expanded sub-tropical high. The impact of the NAO on the immediate Gulf Coast is not surprising, but the degree to which this teleconnection extends influence westward through the spring is of interest.

ENSO tends to have more pronounced impacts on the south-central United States during winter and spring. During warm phase (El Nino) events, the dominant circulation change is a strengthening of the southern branch of the Pacific jet stream across the southern tier of the U.S., along with a reduction in the amount of cold air advecting southward from Canada because of the zonal flow of the northern branch of the polar front jet. The net result is wetter and milder spring and winter seasons in the southern U.S. during warm events (Vega, Rohli, and Henderson 1998; Wolter, Dole, and Smith 1999; Higgins, Leetmaa, and Kousky 2002). Cold phase events are characterized by increased meridionality in the flow across North America with frequent ridging over the eastern Pacific. As a result of the northward displacement of the jet stream, the southern U.S. has a tendency to experience drier and warmer winter and spring conditions (Vega, Rohli, and Henderson 1998; Wolter, Dole, and Smith 1999; Higgins, Leetmaa, and Kousky 2002).

The correlations between SOI and regional maximum temperatures shown in Table 3 are consistent with these synoptic patterns (note that because a negative SOI indicates an El Nino event, positive correlations in Table 3 mean cooler temperatures for this phase), with a stronger relationship. The Far West region in particular is influenced by the SOI pattern throughout the spring, with cooler temperatures when the SOI is negative. A likely explanation for this response is increased moisture and cloud cover in this typically dry part of the study area.

Summary

Daily maximum and minimum temperatures from 184 weather stations in the south-central United States have been used to derive climate regions that are internally homogeneous in terms of temperature. Although such classifications previously have been carried out in other parts of the U.S., this project represents the first such classification at this scale for the south-central part of the country. Multivariate tests show that these regions are significantly different from one another. These regions are believed to represent appropriate aggregation units for regional-scale studies of temperature variability in this part of the continent. In addition, because their definition is based on long-term means of temperature variables, they provide insight into where various climate-producing processes dominate, such as how far inland the maritime influence of the Gulf of Mexico extends.

The derived regions were compared to the widely-used climate divisions, and significant overlap was found. The majority of the divisions in the study area appear to be reasonably homogeneous in terms of temperature. However, nine divisions, extending in a swath from the southwest to the northern part of the region across Texas and Oklahoma, are found to be split between temperature regions and should be viewed with caution as climatological enumeration units. Temperature time series from divisions that are split between regions are likely to be combinations of multiple climate-producing forces, and thus may not be ideal for analyses of temperature trends or teleconnection impacts on regional temperature. In contrast, the cluster-based regions presented here are likely to be internally homogenous in terms of the temperature-controlling processes affecting them. Therefore, regional temperature time series produced from high-quality observing sites within each region should be reliable for use in trend analysis or downscaling studies for the study area.

Monthly maximum and minimum temperatures in the defined regions show significant correlations with large-scale atmospheric circulation, particularly in winter and spring, which demonstrates the utility of these defined areas as enumeration units for analyses of climate patterns. Both the PNA and the NAO have been found to significantly influence maximum and minimum temperatures in the study area, with varying intensities both seasonally and spatially.

In conclusion, this project has identified seven regions across the south-central United States that are climatologically homogeneous in terms of temperature. As global and regional climate variability continues to be an important topic of study, coherent climate regions such as these serve as useful enumeration units for identifying the factors responsible for producing spatial variations in climate, as well as in quantifying relationships between regional climate and hemispheric scale circulation, and downscaling from global or regional climate models.

References

Aldenderfer, M. S., and R. K. Blashfield. 1984. *Cluster Analysis*. Edited by M. S. Lewis-Beck. Vol. 44, *Sage University Paper Series on Quantitative Applications in the Social Sciences*. Newbury Park, CA: Sage Publications.

Barnston, A. G., and R. E. Livezey. 1987. Classification, Seasonality, and Persistence of Low-Frequency Atmospheric Circulation Patterns. *Monthly Weather Review* 115:1083-1126.

Bell, J., P. Duffy, C. Covey, L. Sloan, and the CMIP investigators. 2000. Comparison of Temperature Variability in Observations and Sixteen Climate Model Simulations. *Geophysical Research Letters* 27 (2):261-264.

Briggs, R. D., and R. C. Lemin, Jr. 1992. Delineation of Climatic Regions in Maine. *Canadian Journal of Forest Research* 22:801-811.

Bunkers, M. J., J. R. Miller, Jr., and A. T. DeGaetano. 1996. Definition of Climate Regions in the Northern Plains Using an Objective Cluster Modification Technique. *Journal of Climate* 9:130-146.

Cheng, X., and J. M. Wallace. 1993. Cluster Analysis of Northern Hemisphere Wintertime 500-hPa Height Field: Spatial Patterns. *Journal of the Atmospheric Sciences* 50 (16):2674-2696.

Coronato, F., and A. Bisigato. 1998. A Temperature Pattern Classification in Patagonia. *International Journal of Climatology* 18:765-773.

Davis, R. E., and L. S. Kalkstein. 1990. Development of an Automated Spatial Synoptic Climatological Classification. *International Journal of Climatology* 10:769-794.

DeGaetano, A. T. 1996. Delineation of Mesoscale Climate Zones in the Northeastern United States Using a Novel Approach to Cluster Analysis. *Journal of Climate* 9:1765-1782.

———. 2001. Spatial Grouping of United States Climate Stations Using a Hybrid Clustering Approach. *International Journal of Climatology* 21:791-807.

Driscoll, D. M., P. B. Rice, and J. M. Yee Fong. 1994. Spatial Variation of Climatic Aspects of Temperature: Interdiurnal Variability and Lag. *International Journal of Climatology* 14:1001-1008.

Dunteman, G. H. 1989. *Principal Components Analysis*. Edited by M. S. Lewis -Beck. Vol. 69, *Sage University Paper Series on Quantitative Applications in the Social Sciences*. Newbury Park, CA: Sage Publications.

Easterling, D. R., T. R. Karl, K. P. Gallo, D. A. Robinson, K. E. Trenberth, and A. Dai. 2000. Observed Climate Variability and Change of Relevance to the Biosphere. *Journal of Geophysical Research* 105 (D15):20,101-20,114.

Easterling, D. R., T. R. Karl, J. H. Lawrimore, and S. A. Del Greco. 1999. United States Historical Climatology Network Daily Temperature, Precipitation, and Snow Data for 1871-1997, ORNL/CDIAC-118, NDP-070, 23. Oak Ridge, TN: Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory.

Easterling, D. R., T. R. Karl, E. H. Mason, P. Y. Hughes, and D. P. Bowman. 1996. United States Historical Climatology Network (U. S. HCN) Monthly Temperature and Precipitation Data, ORNL-CDIAC-87, NDP-019/R3, 27. Oak Ridge, TN: Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory.

Feddema, J. J. 2005. A Revised Thornthwaite-Type Global Climate Classification. *Physical Geography* 26 (6):442-466.

Fovell, R. G. 1997. Consensus Clustering of U.S. Temperature and Precipitation Data. *Journal of Climate* 10:1405-1427.

Fovell, R. G., and M.-Y. C. Fovell. 1993. Climate Zones of the Conterminous United States Defined Using Cluster Analysis. *Journal of Climate* 6 (11):2103-2135.

Gong, X., and M. B. Richman. 1995. On the Application of Cluster Analysis to Growing Season Precipitation Data in North America East of the Rockies. *Journal of Climate* 8:897-931.

Guttman, N. B., and R. G. Quayle. 1996. A Historical Perspective of U.S. Climate Divisions. *Bulletin of the American Meteorological Society* 77 (2):293-303.

Hanley, D. E., M. A. Bourassa, J. J. O'Brien, S. R. Smith, and E. R. Spade. 2003. A Quantitative Evaluation of ENSO Indices. *Journal of Climate* 16 (8):1249-1258.

Henderson, K. G., and P. Robinson, J. 1994. Relationships between the Pacific/North American Teleconnection Patterns and Precipitation Events in the South-Eastern USA. *International Journal of Climatology* 14:307-323.

Higgins, R. W., A. Leetmaa, and V. E. Kousky. 2002. Relationships between Climate Variability and Winter Temperature Extremes in the United States. *Journal of Climate* 15:1555-1572.

Horel, J. D. 1981. A Rotated Principal Component Analysis of the Interannual Variability of the Northern Hemisphere 500 mb Height Field. *Monthly Weather Review* 109:2080-2092.

Hurrell, J. W. 2000. Climate Variability: North Atlantic and Arctic Oscillation (NAO/AO). In *Encyclopedia of Atmospheric Sciences*, eds. J. Holton, J. Pyle and J. Curry, 439-445. Burlington, MA: Academic Press.

Johnson, R. A., and D. W. Wichern. 1998. *Applied Multivariate Statistical Analysis*. 4th ed. Upper Saddle River, NJ: Prentice Hall.

Jolliffe, I. T. 1990. Principal Component Analysis: A Beginner's Guide - I. Introduction and Application. *Weather* 45:375-382.

Kalkstein, L. S., G. Tan, and J. A. Skindlov. 1987. An Evaluation of Three Clustering Procedures for Use in Synoptic Climate Classification. *Journal of Climate and Applied Meteorology* 26:717-730.

Leathers, D. J., B. Yarnal, and M. A. Palecki. 1991. The Pacific/North American Teleconnection Pattern and United States Climate. Part I: Regional Temperature and Precipitation Associations. *Journal of Climate* 4:517-528.

McGuffie, K., A. Henderson-Sellers, N. Holbrook, Z. Kothavala, O. Balachova, and J. Hoekstra. 1999. Assessing Simulations of Daily Temperature and Precipitation Variability with Global Climate Models for Present and Enhanced Greenhouse Climates. *International Journal of Climatology* 19:1-26.

NCDC. 2006. Data Documentation for Data Set 3200 (DSI-3200): Surface Land Daily Cooperative Summary of the Day, 19 pp. Asheville, NC: National Climatic Data Center.

Parry, M. L., and T. R. Carter. 1985. The Effect of Climatic Variations on Agricultural Risk. *Climatic Change* 7:95-110.

Pineda-Martinez, L. F., N. Carbajal, and E. Medina-Roldan. 2007. Regionalization and Classification of Hioclimatic Zones in the Central-Northeastern Region of Mexico Using Principal Component Analysis (PCA). *Atmosfera* 20 (2):133-145.

Raffalovich, L. E. 1994. Detrending Time Series: A Cautionary Note. *Sociological Methods & Research* 22 (4).

Reek, T., S. R. Doty, and T. W. Owen. 1992. A Deterministic Approach to the Validation of Historical Daily Temperature and Precipitation Data from the Cooperative Network. *Bulletin of the American Meteorological Society* 73 (6):753-762.

Rhee, J., J. Im, G. J. Carbone, and J. R. Jensen. 2008. Delineation of Climate Regions Using in-situ and Remotely-Sensed Data for the Carolinas. *Remote Sensing of Environment* 112:3099–3111.

Rogers, J. C., and R. V. Rohli. 1991. Florida Citrus Freezes and Polar Anticyclones in the Great Plains. *Journal of Climate* 4:1103-1113.

Santos, J. A., J. Corte-Real, and S. M. Leite. 2005. Weather Regimes and Their Connection to the Winter Rainfall in Portugal. *International Journal of Climatology* 25:33-50.

SAS Institute. 1999. *SAS Online Documentation*. 8 ed. Cary, N.C.: SAS Institute, Inc.

Sheridan, S. C. 2003. North American Weather-Type Frequency and Teleconnection Indices. *International Journal of Climatology* 23 (1):27-45.

Stooksbury, D. E., and P. J. Michaels. 1991. Cluster Analysis of Southeastern U.S. Climate Stations. *Theoretical and Applied Climatology* 44:143-150.

Stott, P. A., S. F. B. Tett, G. S. Jones, M. R. Allen, J. F. B. Mitchell, and G. J. Jenkins. 2000. External Control of 20th Century Temperature by Natural and Anthropogenic Forcings. *Science* 290:2133-2137.

Unal, Y., T. Kindap, and M. Karaca. 2003. Redefining the Climate Zones of Turkey Using Cluster Analysis. *International Journal of Climatology* 23:1045-1055.

Vega, A. J., R. V. Rohli, and K. G. Henderson. 1998. The Gulf of Mexico Mid -Tropospheric Response to El Nino and La Nina Forcing. *Climate Research* 10:115-125.

White, D., M. Richman, and B. Yarnal. 1991. Climate Regionalization and Rotation of Principal Components. *International Journal of Climatology* 11 (1):1-25.

Whitfield, P. H., K. Bodtker, and A. J. Cannon. 2002. Recent Variations in Seasonality of Temperature and Precipitation in Canada, 1976-1995. *International Journal of Climatology* 22:1617-1644.

Williams, K. R. S., and K. C. Parker. 1997. Trends in Interdiurnal Temperature Variation for the Central United States, 1945-1985. *Professional Geographer* 49 (3):342-355.

Wolter, K., R. M. Dole, and C. A. Smith. 1999. Short-Term Climate Extremes over the Continental United States and ENSO. Part I: Seasonal Temperatures. *Journal of Climate* 12:3255-3272.