2003

Reorganizing National Climate Data Center climate boundaries for the state of Oregon

Stephen Daniel Shelton
San Jose State University

REORGANIZING NATIONAL CLIMATE DATA CENTER CLIMATE BOUNDARIES FOR THE STATE OF OREGON

A Thesis

Presented to

The Faculty of the Department of Geography

San Jose State University

In Partial Fulfillment

Of the Requirements for the Degree

Master of Arts

By

Stephen Daniel Shelton

May 2003

UMI Number: 1415736

UMI

UMI Microform 1415736

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

APPROVED FOR THE DEPARTMENT OF GEOGRAPHY

Dr. Richard Taketa
Associate Professor,
Department of Geography

Dr. David Helgren
Professor,
Department of Geography

Dr. Gary Pereira
Assistant Professor,
Department of Geography

APPROVED FOR THE UNIVERSITY

ABSTRACT

REORGANIZING NATIONAL CLIMATE DATA CENTER
CLIMATE BOUNDARIES FOR THE STATE OF OREGON

by Stephen Daniel Shelton

This thesis creates new climate divisions to replace the current boundaries used by the National Climate Data Center for the state of Oregon. The current boundaries were created over a long time period, before many modern computer advances became available. The thesis contains two parts: First is a qualitative method (visual), and second is a quantitative method (cluster analysis). Although the visual method was generally more homogeneous than the original divisions, the results were not convincing and were possibly inappropriately influenced by using three temperature-related parameters. After removing the superfluous parameters of latitude and longitude, the cluster analysis produced promising results, but the clusters themselves were not geographically organized and were not conducive to the boundary-drawing phase of creating the new divisions.

Acknowledgements

I would like to give thanks to my wonderful and supportive wife for all that she has gone through during the last few years. I love you Taundra.

I would also like to thank my son for giving me the joy in life that has helped me more than anyone will ever know.

Of course, thanks also goes to all those who have taught me throughout my school career. Knowledge is a wonderful thing to share.

Most of all, I give thanks to my Father in Heaven, to whom I owe all.

TABLE OF CONTENTS

List of Figures

CHAPTER 1

Introduction

NCDC Climate Divisions

State climate offices, in conjunction with the National Climate Data Center (NCDC), utilize climate divisions for the purpose of data collection and dissemination. Presently, as one looks at a map displaying the climate divisions of Oregon (Figure 1), one can see that many of the climate division boundaries are based on county boundaries, and some boundaries are unclear as to what they are based on. The opinion of the author is that although climates almost always change gradually and the creation of boundaries require a border line at some point,

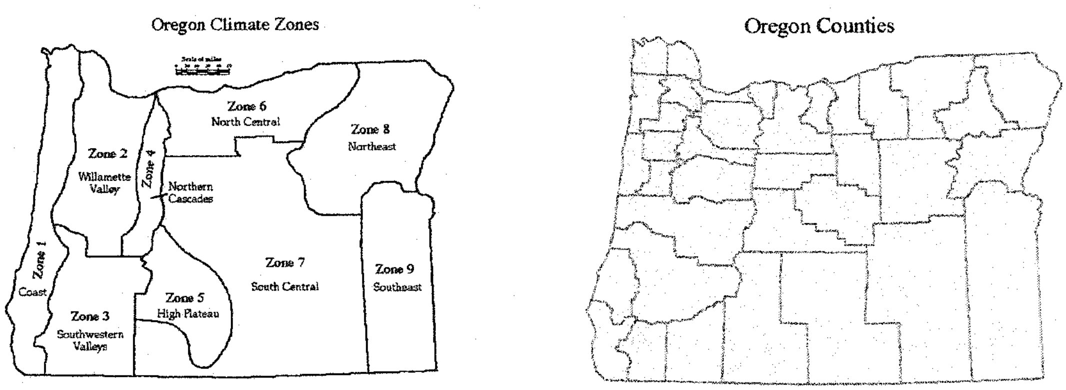

Figure 1: Oregon climate divisions of the National Climate Data Center (left) and counties of Oregon (right).

climate boundaries should not necessarily follow county boundaries, as climate does not change when you cross an arbitrarily placed political line.

This examines and analyzes climate data with the purpose of finding areas of common features. These areas are bounded to create climate divisions. This study also addresses the question: Can new state climate divisions be created that better represent the climatic characteristics of the regions they denote? This question is addressed in two different ways, qualitatively and quantitatively. Specific criteria is developed in order to compare the original divisions to the newly created ones. These criteria are then used to test whether the new divisions are appropriate and more representative of the climate found within each division.

The climate *divisions* used by the National Climate Data Center are different than climate *zones* found within a traditional climate classification system, such as Köppen or Thornthwaite. In this project, the goal is not to determine all regions of a certain type of climate in an area; but rather to construct definitive boundaries (administrative units) that contain areas of similar climate for the purpose of data collection and dissemination. This distinction is not so important in the identification and classification of climate areas, but is important in the creation of the divisions themselves, and the drawing of the boundaries that separate them.

Study Area

The climate divisions of the state of Oregon are the subject for this study. Oregon has areas where the boundaries between regions of different climates are obvious, and other areas where the transition is more subtle. Generally, the western third of the state has a moist climate, where rainfall averages generally increase northward. The Coast and the Cascades (both running north-south) are the two major mountain ranges in Oregon. Between these mountain ranges is the fertile, but rain-shadowed Willamette Valley. The climate on the leeward side of the Cascades is in stark contrast to the wet maritime-like climate of the west. Here, the landscape is high desert (except in the far north east where the landscape is mountainous). The climate here is much dryer and temperatures are more extreme; in other words, winter is colder and summer is hotter (Taylor 1999).

Oregon was not only chosen for its variety of climate types, but also for its compact shape, reducing some of the complexities that a less compact shape might present in a preliminary study such as this. One of the complexities that will be avoided in using a compact shape and reasonable climate pattern is in the process of creating new divisions, or regions.

CHAPTER 2

Concepts & Literature

Regions

The important geographical-organizing concept of regions must be explored before analyzing data, and creating divisions. A region is a concept that is used to identify and organize areas of the Earth's surface for various purposes (Bednarz et. al., 1994). A region may be easy to imagine and describe, but can be difficult to outline on a map. This is because regions can be thought of from different perspectives. Regions are scientific devices that allow one to make spatial generalizations and are based on artificial criteria that the researcher establishes for the purpose of constructing them (De Blij, 1997).

The types of regions utilized in this research are *formal* or *uniform* regions (as opposed to functional or perceptive regions; Bednarz et. al., 1994). Formal regions have certain properties in common. First, they have an area; regions may be intellectual constructs, but they do exist in the real world. Second, formal regions have boundaries. Although on occasion, nature provides sharply defined borders, they usually have to be drawn according to criteria established for a specific purpose. Often, regions transition into each other gradually (this is often the case in working with climates). In reality, this can present more of a transition zone than a *border* where two regions meet. Third, all regions have a location, whether it is an absolute location or relative one. Fourth, and this is an important characteristic for this study, formal regions are marked by a certain homogeneity (De Blij, 1997). Some of these qualities will be discussed later, as criteria for the new climate divisions.

Climate Classification

Climate classification is a method of dividing climate types into different regions or zones. "Climate classification is an essentially geographic technique. It allows the simplification and generalization of the great weight of statistics built up by the climatologists" (Hare, 1951). While this paper will not focus on the many classification systems used to show the world's climates, one should understand the basics behind the techniques. Lydolph (1985) stated that to fully depict the climate of an area, one must convey the end results of climate (the statistical means and deviations of measured elements such as temperature and precipitation), and the reasons behind those results (day-to-day events, local effects, etc.).

Two systems of classifying climate exist: empirical and genetic. An empirical scheme is based on the statistics from observations of individual elements such as temperature and precipitation. A genetic scheme is based on the causes behind those statistics, or on the observation of weather sequences over time. The latter cannot be precisely measured, and must be handled subjectively. As a result, genetic classifications have not been very successful. Climatologists generally agree to base classifications on observed facts, and then to discuss the causes behind the facts, rather than incorporating them into the classification scheme itself (Lydolph, 1985).

While many climatic features can be handled statistically, most modern classification systems are based only on temperature and precipitation statistics, including the most widely used Köppen and Thornthwaite classifications (Hare, 1951). Other criteria may be used (such snow depth, hours of sunshine, average humidity levels, etc.), but because of complexity, other information (measured or not) is often included as verbal descriptions about each climate type. Additionally, many modern classifications use vegetation and soil as guides for the locations of climatic boundaries. This is partially a historical residual since most of the early scientists in this field were plant-geographers and biologists (Hare, 1951). Lydolph (1985) made the interesting point that because most classifications have utilized vegetation and other natural phenomenon to determine boundaries, they have become systems that more or less classify the state of the surface of the earth rather than the air immediately above it.

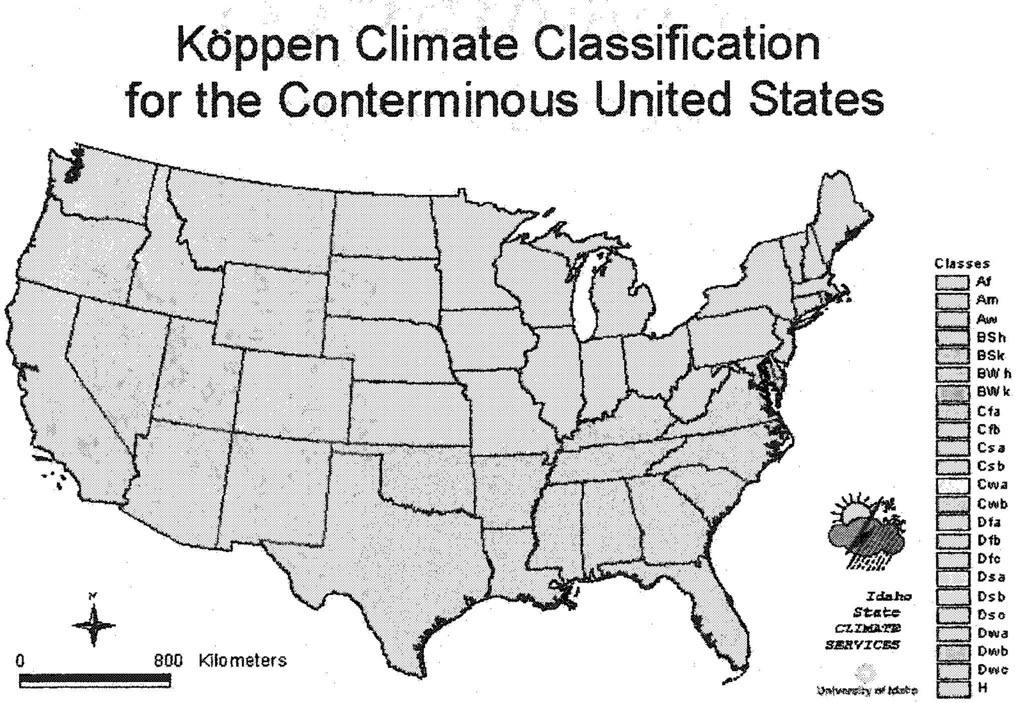

Figure 2: Köppen climate classification of the U.S. (From the Idaho State Climate Service - University of Idaho).

Specific Studies

Guttman and Quayle (1995) compiled a history of the origins and computational methodology of the climate divisions used by the NCDC. The research yields some information on how these divisions came to be. They state, however, that "the history of climatic divisions in the...U.S. has been pieced together from fragmentary documentation." Therefore, this study will not delve into the history of the NCDC climate divisions, though the subject would be an interesting one.

Guttman and Quayle (1995) noted that NCDC's climate divisions are structured such that they often coincide with county boundaries. However, they did not establish a clear explanation for this practice. One gains an impression through reviewing the history of these divisions that many decisions were based on convenience rather than climatological information. (Much of the shifting of the NCDC climate division boundaries over time was made in a period parallel to the creation of well known climate classification systems still in use today.)

Guttman and Quayle (1995) also pointed out weaknesses and strengths of the present divisions. The main weakness they note is the problem of delineating areas of climatological homogeneity using many climate variables. The primary strength that Guttman and Quayle pointed out was the "coherence in space and time" of the data despite the diverse station coverage and varied terrain.

McBoyle (1973) used a method of factor analysis to create a classification which relies solely on climatic variables, and not external factors (such as vegetation). This method was quite flexible in its applications to different scales and showed a moderate correlation with existing classifications such as Köppen's system. McBoyle's study showed the validity of a classification system created using only climate statistics computationally, a method different from classifications of the early 1900s and more similar to the method being employed in this study.

Interpolation

The first part of this study will rely heavily on the process of interpolation (the estimation from a set of observations to a set of unsampled locations for which observations are unavailable). Many different interpolation techniques are available, such as inverse distance weighting, spline, linear, Thiessen, trend, and Kriging (Hartkamp et. al., 1999). Interpolation attempts to create a continuous surface of observational information from a set of recorded observations (points). The rationale behind spatial interpolation is the observation that points closer together in space are more likely to have similar values than points that are further apart (Robeson, 1997). The Kriging method of interpolation will be used in this study. Kriging is based on the assumption that the parameter being interpolated can be treated as a regionalized variable. A regionalized variable is intermediate between a truly random variable and a

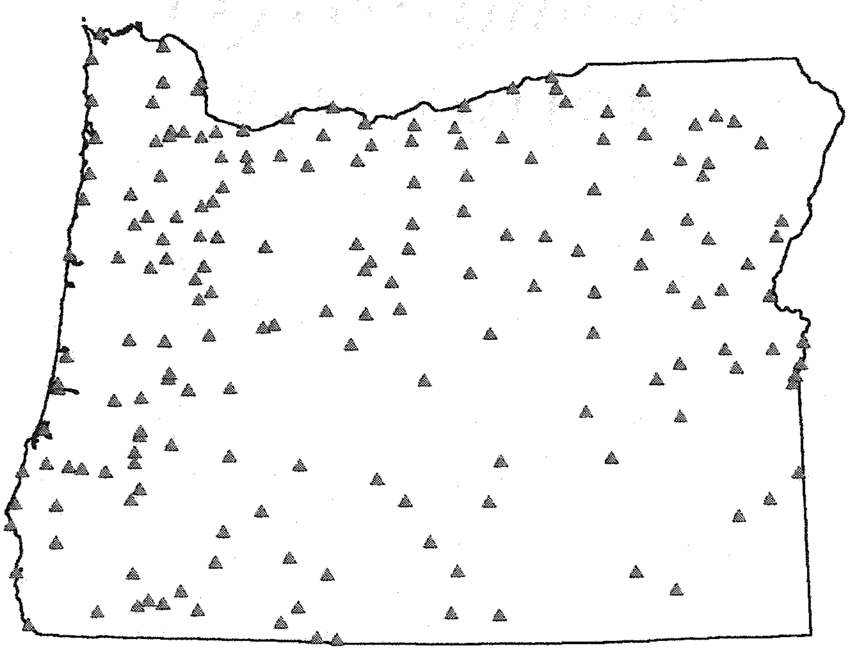

Figure 3: Data points (weather stations) to be interpolated.

completely deterministic variable. A regionalized variable varies in a continuous manner from one location to the next, and therefore points that are near each other have a certain degree of spatial correlation, but points that are widely separated are statistically independent (Davis, 1986). Some interpolation techniques yield results that stay true to the original data points, but are not very realistic looking. Kriging, however, yields a relatively realistic looking surface in which values fit the known data points relatively well.

Cluster Analysis

The second part of this study will use cluster analysis to help determine new climate divisions. Cluster analysis is the process of classifying objects into subsets that have meaning in the context of a particular problem (Jain, 1988). Its object is to sort cases (people, things, events, etc.) into groups, or clusters, so that the degree of association is strong between members of the same cluster and weak between members of different clusters.

The K-Means cluster analysis will be used in this study. In K-Means clustering, the researcher can specify the number of clusters that will be formed. This is useful because the aim of this study is to stay with the NCDC's nine divisions. The K-means cluster analysis may be thought of as an analysis of variance test in reverse in the sense that the significance test in ANOVA evaluates the between-group variability against the within-group variability when computing the significance test for the [alternative] hypothesis that the means in the groups are different from each other. In K-means clustering, the program starts with *k* random clusters, then tries to move objects (e.g., cases) in and out of groups (clusters) to get the most significant ANOVA results (Jain, 1988).

Analysis of Variance

An analysis of variance (One Way ANOVA) will be performed to determine numerically if the new divisions are better than the original division. Analysis of variance is based on a comparison of two different estimates of the variance common to the different populations.

> "Analysis of variance involves the separation or partitioning of the total variance found in three or more groups or samples into two distinct components: (1) variability ***between*** the group or category means themselves; and (2) variability of the observations ***within*** each group around its group mean (McGrew, 2000)."

The analysis of variance test in this study will not be necessarily testing for whether or not populations are equal, but rather will be contrasting the analyses of variance between the original divisions and the new divisions. This will be done by comparing the F-statistic for each of the parameters used in this study. The F-statistic is *the **between** group variation* *divided by* *the **within** group variation*, ($F = \frac{BetweenGroupVariation}{WithinGroupVariation}$). The more variation between groups, and the less variation within groups (climate divisions in this case), the higher the F-value will be. This will be the more desirable outcome because it shows a measure of homogeneity in the divisions.

CHAPTER 3

Methodology

Climate data have been collected for Oregon cities (weather stations) from the Oregon Climate Service and the Western Regional Climate Center. The climate variables that are being used are: average annual precipitation, average annual temperature, heating degree days, and cooling degree days all averaged over the last 30-years (one heating degree day is given for each degree that the daily mean temperature is below 65 degrees Fahrenheit).

For the first (interpolative/qualitative) part of this study the data was interpolated and displayed in a GIS. Views of different variables were analyzed, and areas with common characteristics were assessed visually. Aligning these common areas with some sort of established, but *natural* borders was the next step before comparing the results with the original climate divisions using the *F*-statistic from the ANOVA test explained earlier.

For the second (cluster analysis/quantitative) part of this study, the data was run through a cluster analysis in SPSS (a statistical software application). The resulting clusters were examined and grouped together. These clusters represented the new climate divisions, and were aligned to established, natural boundaries (as with the first part of the study). The resulting divisions were then numerically compared to the original divisions using the *F*-statistic from the ANOVA test.

The visual/qualitative study is a method more similar to the method used in the original genesis of these climate divisions. However, there are tools available today that were not available then (e.g. displaying climate patterns in a GIS). If this method proves successful, it would be an appropriate and simpler method for determining new climate divisions. The cluster analysis/quantitative method is being used for two reasons. The first reason is of an experimental nature, i.e. to determine if cluster analysis is an appropriate and viable method for forming climate divisions. The second reason for using a statistical method is because it offers a greater likelihood of producing a more accurate result than with a qualitative assessment.

Criteria

In order to assess whether or not the new climate divisions are an improvement over the older divisions, criteria of what constitutes a *good* division must be established.

The first factor taken into consideration was the desired number of climate divisions to be created. If one wanted the study to have an end result of 20 climate divisions within a single state, the criteria for determining those divisions would have to be very specific. On the other hand, if only two divisions were to be created, the requirements would be quite general. The goal for this study is to keep the same number of divisions that the NCDC has determined sufficient. The state of Oregon currently has nine divisions.

The most important criterion that this study deals with is the homogeneity within each climate division (meaning that characteristics found within the climate division are similar). However, the level of homogeneity differs if one was looking at the climate of a city compared to that of a country. This study is examining the climate of the state of Oregon. To define homogenous zones for universal application at any level would be very difficult, if not impossible. This means that the idea of homogeneity depends on many things, such as the climate variables being used, the controlling atmospheric physics, and the use of the data (Guttman, 1995). Despite its difficulties, homogeneity is the most important requirement that the climate divisions possess. As mentioned earlier, this study uses the *F*-statistic from an ANOVA test to compare the homogeneity between the NCDC's original divisions and the newly created ones.

The spatial *sense* of the new divisions must also be taken into account. This means that common sense must also be used in drawing boundaries, even after the statistical analysis has *told* you what to do. For example, a climate division that stretches far across the Cascade Mountains would not make sense.

HUC Boundaries

In the effort to create better climate divisions, a major goal of this study is to use some established, but natural climate boundaries as a guide for drawing the new boundaries that would replace the county and unspecified boundaries of the original divisions. Hydrologic Unit Code (HUC) boundaries are a subdivision of the United States made by the USGS to show major and minor river basins. Each river basin has a numeric code, the major basins have 2 digit HUC boundary code, while the smaller, sub-basins have 4, 6, and 8 digit codes. For example, a large river basin zone may have a HUC2 code of 17. The smaller basins, located within that zone would have HUCs equal to 1701, 1702, etc. The sub-basins can be even further subdivided by using HUC6 and HUC8. The smaller units (HUC8) will be used for guides in drawing the new boundaries.

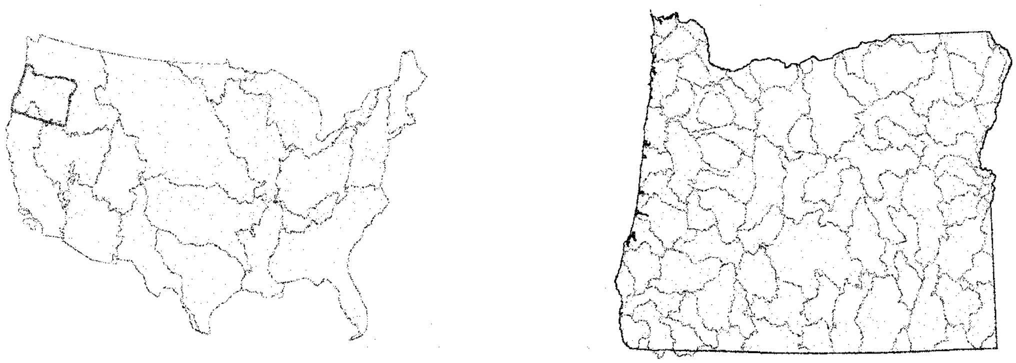

Figure 4: Major Hydrological Units of the U.S. (left), and smallest divisions of Hydrological Units (HUC 8) for Oregon (right).

CHAPTER 4

Results

Interpolative Analysis/Qualitative Study

Figure five on the following page presents the results of the interpolation of the data points. The maps in figure 5 were prepared with some hill-shading effects for visual purposes. Additional maps were also prepared showing some of the interpolations overlaying other parameters by way of a hill-shading technique similar to illuminated contours (Figure 6).

These maps aided in the visual assessment of areas with common climate characteristics as well as verifying the validity of the interpolation used. For instance, it is logical (by definition) that areas with lower temperatures would coincide with areas that experienced more heating degree days.

Another way of checking the interpolations was to compare them with other maps from established agencies. Figure 7 is a map of average annual precipitation created with contours provided by the USGS and a map made with data from the NCDC (the interpolated map is in the center).

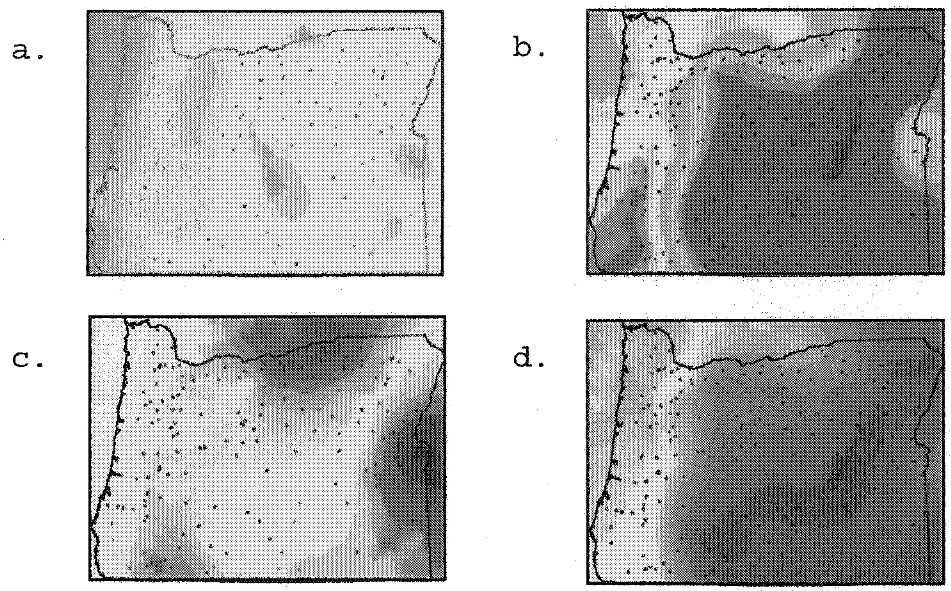

Figure 5: a) average annual precipitation (blues are wetter/reds are dryer, b) average annual temperature (purples are cooler/reds are warmer), c) annual heating degree days (lighter browns are more/darker browns are less, and d) annual cooling degree days (light blues are less/dark blues are more).

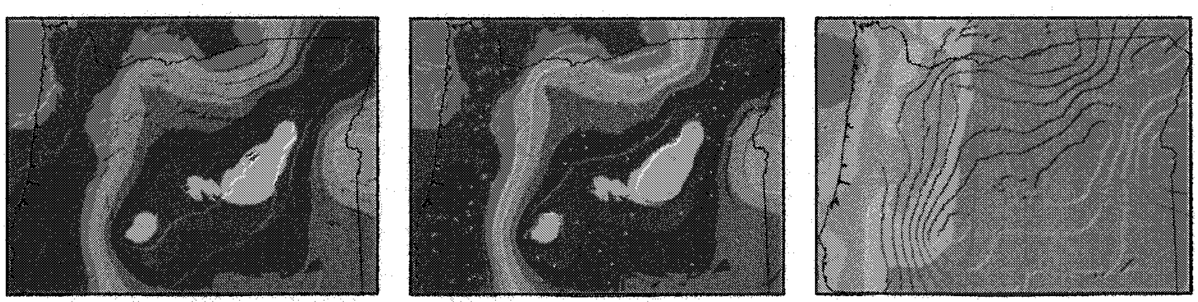

Figure 6: (Left to Right): average annual temperature with cooling degree days contours, average annual temperature with heating degree days contours (showing cities), and average annual precipitation with average annual temperature.

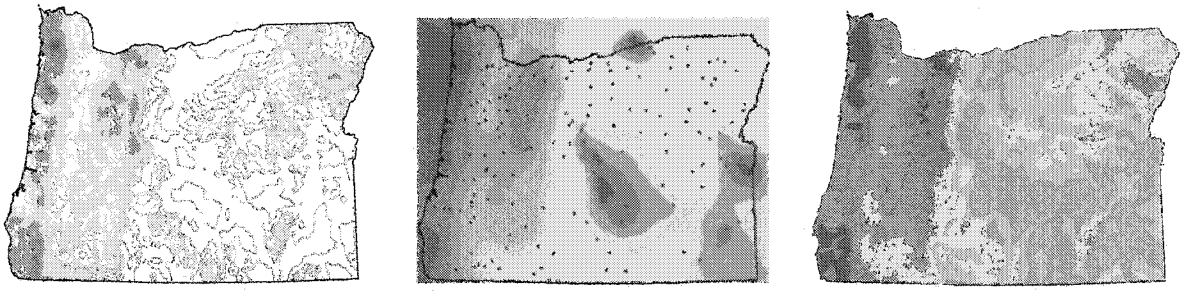

Figure 7: Average annual precipitation from USGS (left), this study (center), and the NCDC.

A visual comparison shows the obvious similarities. The differences are somewhat negligible because of the state-wide scale that this study encompasses.

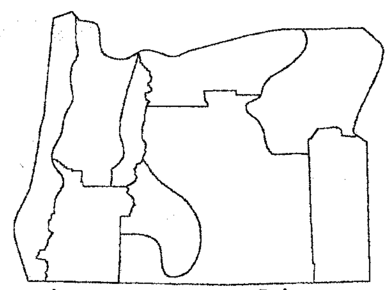

Figure 8: Climate divisions of Oregon, from the NCDC

Figure 8 shows the original climate divisions, created by the NCDC. Figure 9 shows the coincidences between these divisions and the interpolated values for the various climate parameters. Some parameters appear to fit in certain divisions generally well; but as a whole, they do not fit very well.

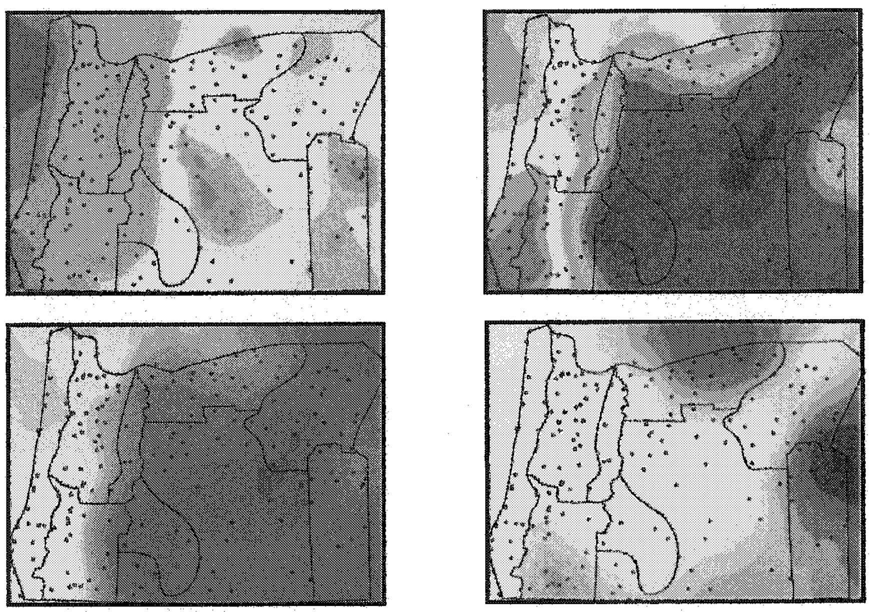

Figure 9: (Clockwise from upper-left): Interpolations of average annual precipitation, average annual temperature, cooling degree days, and heating degree days, all shown with original divisions.

After examining the interpolated climate maps, a visual interpretation had to be made of where new climate division boundaries should be placed. These visually "eyeballed" boundaries were fit to the nearest HUC boundary, resulting in the new climate divisions (Figure 10). Some irregularities in the new divisions' boundaries were due to the sometimes odd shape of the hydrological units.

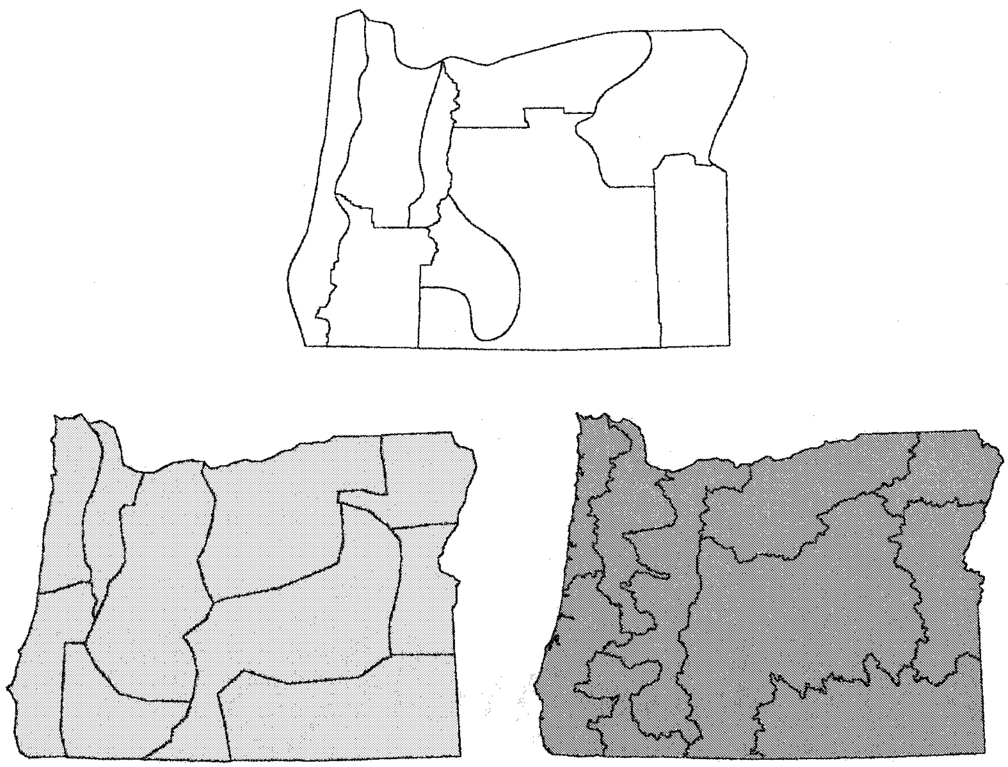

Figure 10: Original NCDC divisions (top), visually "eyeballed" boundaries (bottom-left), and new climate divisions for Oregon (bottom-right).

Most of the new divisions correspond generally to one of the original ones. The exception is the division in the southeast, and the coast, which is now separated into two distinct divisions. Although the goal in this study was to stay with nine divisions, as with the original configuration, the differences between the southern and northern coasts could not be ignored. Figure 11 shows the new climate divisions with the four climate parameters used to help create them. Although it cannot be quantified, certain variables were visually *weighted* more heavily than others in certain parts of the state. For example, precipitation was considered more important in defining divisions in the western third of Oregon (west of the Cascades). The Eastern two-thirds did not have as much variability in precipitation. Although precipitation *was* taken into consideration in creating new divisions,

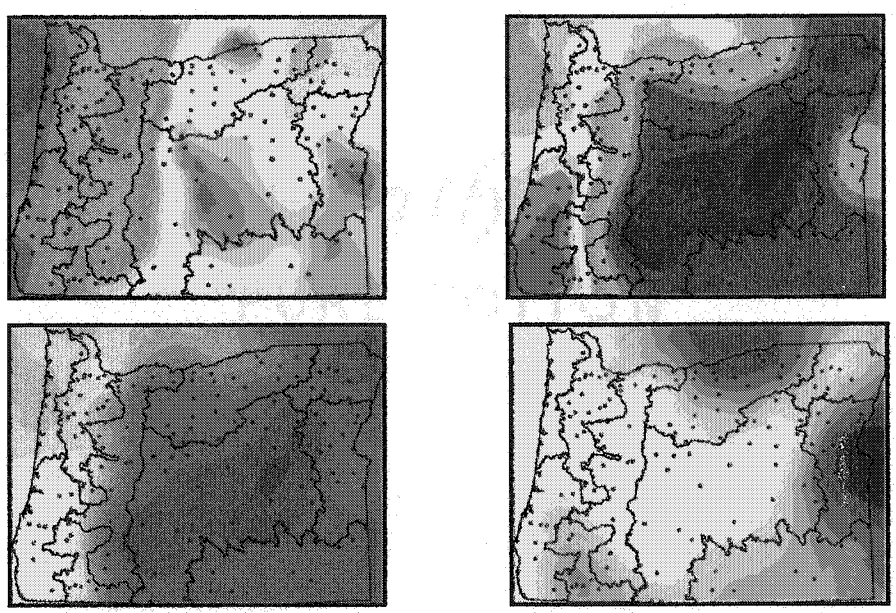

Figure 11: (Clockwise from upper-left): Interpolations of average annual precipitation, average annual temperature, cooling degree days, and heating degree days, all shown with new divisions.

temperature and the heating and cooling degree days parameters were more telling in this part of the state.

Visually, these new divisions appear to coincide with the interpolated maps of the various climate parameters better than the original divisions in most cases. In addition, Figure 12 shows the new divisions with vegetation patterns, agricultural patterns, and aquatic ecoregions (all very much influenced by climate). Again, visually, there is a general coincidence between the new climate divisions and these patterns.

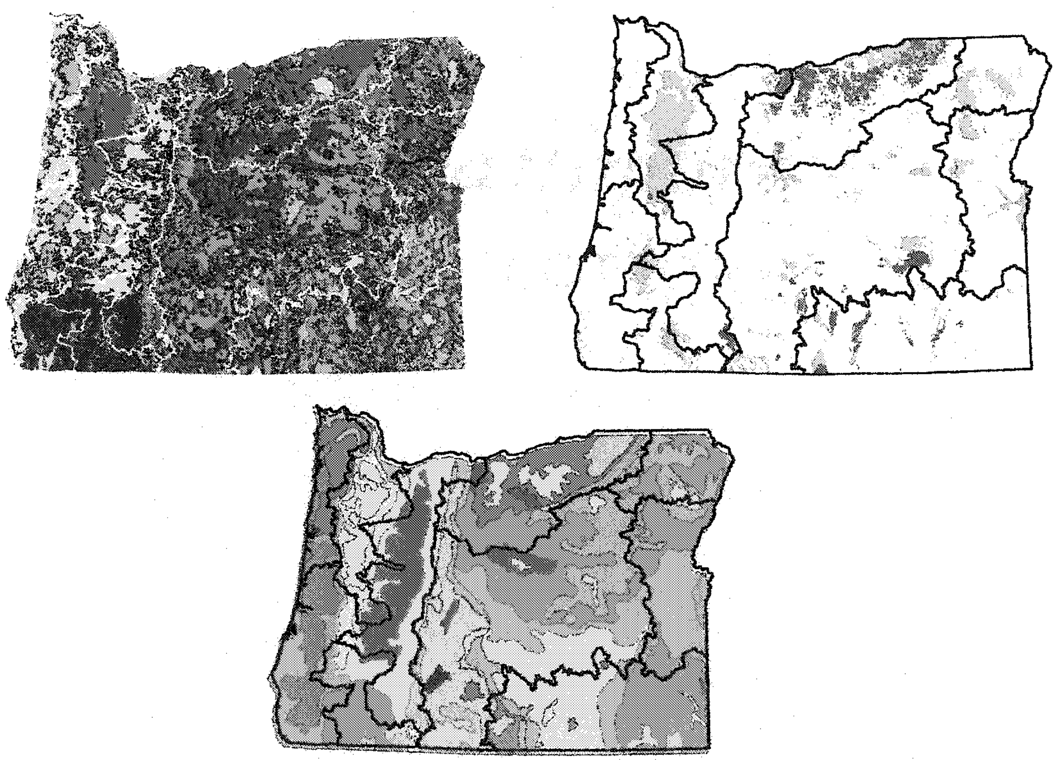

Figure 12: Vegetation (upper-left), agriculture (upper-right), and aquatic ecoregions (bottom).

Below is a table with the resulting *F*-Values from the ANOVA test. The higher, more desirable values have been colored red, the lower values are blue.

F-Value Results from Interpolative/Qualitative Study

	Orig. Divisions	New Divisions
	F-Value	F-Value
Average Annual Precipitation	91.089	71.059
Average Annual Temperature	12.000	13.048
Heating Degree Days	17.356	17.565
Cooling Degree Days	9.263	12.366

Conclusions for Interpolative Analysis/Qualitative Study

In all cases but one, the *F*-statistic was higher in the new divisions than in the original ones (meaning the within-division variation was lower and the between-division variation was higher). This leads to the conclusion that the newly created divisions are more homogeneous than the original divisions. However, the differences between *F*-values were not extraordinary, and the *F*-value for precipitation was *higher* in the original divisions. Although the author believes that these results are a step in the right direction, more care needs to be taken in the assessment of the interpolated climate data. Perhaps a more quantitative method will yield more convincing results.

Cluster Analysis/Quantitative Study

Below are maps showing the resulting clusters from the cluster analysis. The first (Figure 13) shows the

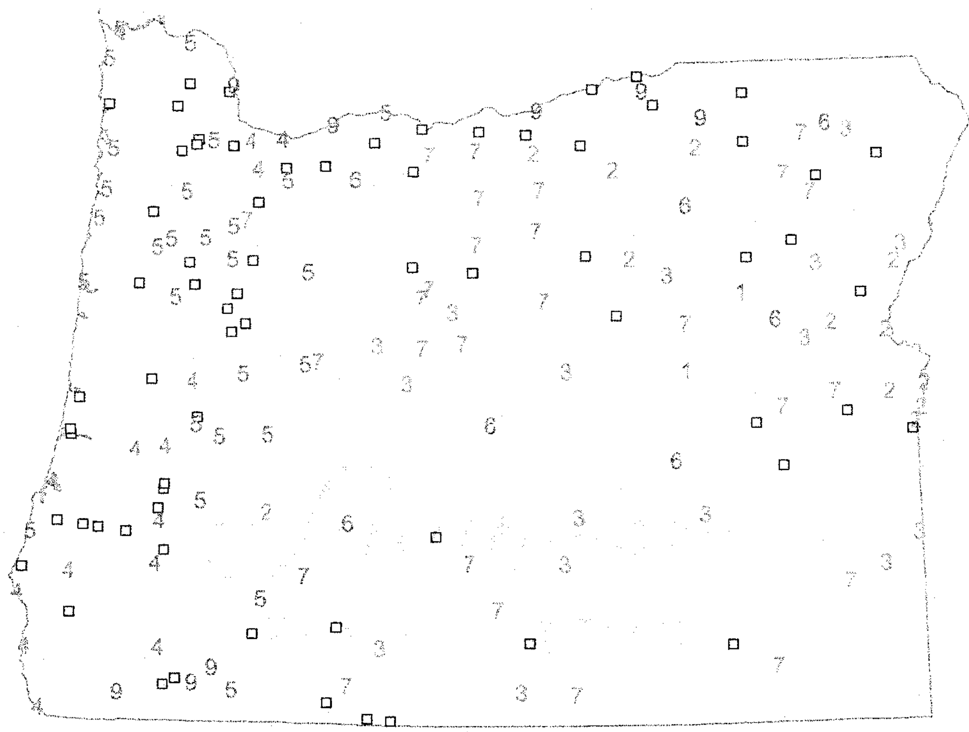

Figure 13: Cluster results before including latitude and longitude (squares represent weather stations not included in analysis).

results of the cluster analysis before including locational information (latitude and longitude). The lower map (Figure 14) is after including locational information. (The different colors and numbers are only to help distinguish between the different clusters and do not represent specific climate variables.) The separations between divisions are much clearer when the locational information is used; these are the results that were used in drawing the new division boundaries.

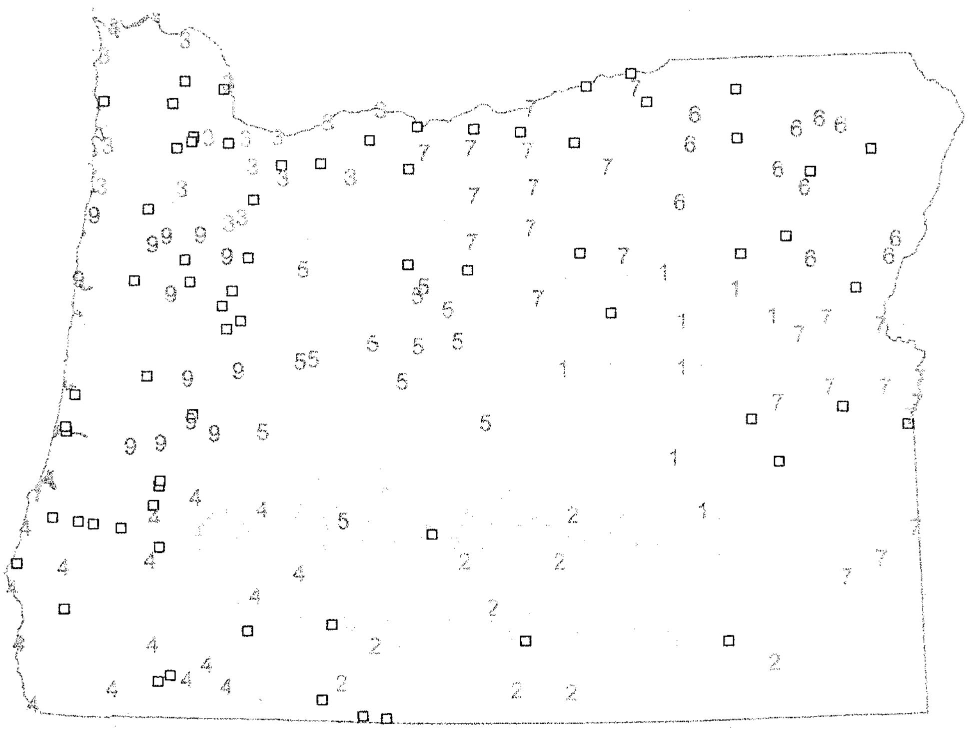

Figure 14: Cluster results after including latitude and longitude (squares represent weather stations not included in analysis).

With each station assigned to a specific cluster, the HUC boundaries served as guides in creating the new divisions. Here the criteria of making spatial sense came into play. In particular, a basic knowledge of the state of Oregon played a part in keeping the newly-drawn divisions consistent with the topography of the land. Not every member of every cluster could always be included together in the same division due to the fact that there might be members of different clusters in the same HUC (as illustrated in Figure 15). But this happened rarely. Below are maps of the newly created divisions and the original NCDC divisions (Figure 16) with which the new divisions will be compared.

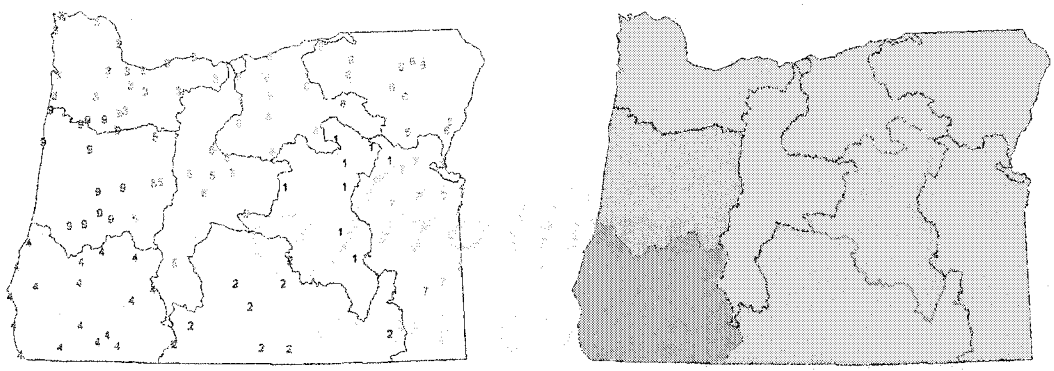

Figure 15: New divisions with clustered results (left; showing that some cluster members were split up because of shared location in HUCs with other cluster members), and without clustered results (right).

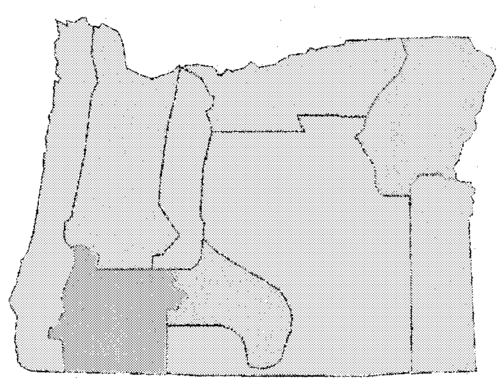

Figure 16: Original Divisions (from NCDC).

The new divisions are similar, in some respects, to the original divisions. Obvious similarities are seen in the Cascades and along the eastern and eastern-central borders of Oregon. The biggest difference seems to be along the Pacific coast and the western third of the state. Instead of a long climate division along the coast, and divisions for the Willamette Valley and southern mountains, the coast and interior area were divided into three divisions, stacked from north to south.

Below are two tables. The first, labeled *1*, is a comparison of the original divisions and the new divisions. The second table, labeled *2* (on page 33), is a comparison of the original divisions and the raw clusters (before assigning new divisions). As with the table from the preceding chapter, these are the resulting *F*-Values from the ANOVA test. The higher, more desirable values have been colored red, the lower values are blue.

1) *F-Value Results from Cluster Analysis/Quantitative Study*

	Original Divisions	New Divisions
	F-Value	F-Value
Average Annual Precipitation	91.089	62.185
Average Annual Temperature	12.000	7.351
Heating Degree Days	17.356	11.997
Cooling Degree Days	9.263	4.637

Conclusions for Cluster Analysis/Quantitative Study

The *F*-values for the original and the newly created climate divisions show that the original divisions are more homogeneous. *F*-values are higher for the original divisions for every climate variable tested. In fact, these values for the new divisions in table *1* are lower than the values for the visually determined divisions in the previous section. How can this be if the cluster analysis is theoretically supposed to create the most homogeneous solution? As stated earlier, in the previous section, some members from different clusters fell within the same HUC boundary. Therefore, some cluster members had to be grouped with other cluster members in drawing the boundaries. Although this did not occur often (see Figure 15), perhaps including members of different clusters in the same climate divisions was enough to throw the homogeneity off for the new divisions.

The second table (table *2*), however, does not necessarily support this conclusion. This table compares the ANOVA results from the original divisions to the raw cluster results (the cluster results before being assigned a division). Overall, the *F*-values are much lower for the raw clustered numbers. Although these are theoretically supposed to have the highest *F*- values, the results from the raw clusters most certainly are not the very high at all.

2) *F-Value Results from Cluster Analysis/Quantitative Study*

	Original Divisions	Raw Clusters
	F-Value	F-Value
Average Annual Precipitation	91.089	31.155
Average Annual Temperature	12.000	10.233
Heating Degree Days	17.356	15.448
Cooling Degree Days	9.263	6.749

This study used the cluster results (from Figure 14) that included latitude and longitude. This was done because the results appeared more geographically conducive for drawing boundaries around them (they cluster were tightly grouped and contiguous. However, the original clustering, without latitude and longitude (Figure 13) delivers higher *F*-values than the cluster scenario just tested (see table 3), which indicates that using the latitude and longitude in the cluster analysis overwhelmed

3) *F-Value Results from Cluster Analysis/Quantitative Study*

	Original Divisions	Raw Clusters(No lat/long)
	F-Value	F-Value
Average Annual Precipitation	91.089	18.918
Average Annual Temperature	12.000	360.487
Heating Degree Days	17.356	468.110
Cooling Degree Days	9.263	18.074

the climate parameters. In fact, a cluster analysis on *only* latitude and longitude yields identical results to the analysis that includes latitude and longitude with the four climate parameters.

Although the above table (table 3) reveals much higher *F*-values for three of the four parameters, precipitation still has a low *F*-value (just as in the visual analysis). Table 4 shows the results of performing a cluster analysis using only temperature and precipitation (leaving heating and cooling degree days out because of their high correlation with temperature). This produced the best ANOVA results so far (i.e., each parameter used has a higher *F*-value than the original zones). Just as latitude and longitude overwhelmed the cluster analysis, in this

4) *F-Value Results from Cluster Analysis/Quantitative Study*

	Original Divisions	Raw Clusters Using Only Precip. & Temperature (No lat/long)
	F-Value	F-Value
Average Annual Precipitation	65.414	1045.525
Average Annual Temperature	12.000	16.004
Heating Degree Days	x	x
Cooling Degree Days	x	x

example, using three temperature-related parameters overwhelmed the cluster analysis. Although it cannot be quantified, this is perhaps what occurred in the visual analysis as well.

CHAPTER 5

Final Conclusions

Although the visual analysis produced results that were generally more acceptable than the original divisions, the results were not fully convincing, and seemed to be affected by the use of three temperature-related climate parameters. The cluster analysis that incorporated locational information was affected by two problems, the use of latitude & longitude in the analysis and the use of three temperature-related climate parameters.

While the results from the cluster analysis based only on temperature and precipitation did not necessarily yield *F*-values higher than the *F*-values of the other ANOVA tests, the resulting values *were* all higher than the *F*-values from the original divisions (i.e., these clusters were more homogeneous than the original NCDC climate divisions). However, these clusters (as was pointed out in Chapter four, Figure 13) were not contiguous. While this cluster analysis produced the homogeneous numbers desired, the appropriate new divisions cannot be drawn so as to have one division for each cluster. Either compromises must be made in drawing the boundaries (including members of different clusters in the same division, as was demonstrated earlier) or a different method of grouping the climate stations will be needed.

Another possible weakness in this study may be the use of the hydrological units as guides for building the new boundaries. Although the main goal of this study was to use something other than arbitrary or political boundaries, the hydrological unit boundaries may have deemed themselves inappropriate for this task. Perhaps there are other sets of known, natural boundaries that could be used that would be more flexible. It appears that smaller units should be sought out in order to eliminate the possibility of members of different clusters falling within the same guide boundary (the HUC boundaries in this case). Using smaller boundaries to guide the boundary-drawing phase would enable all the members of a particular cluster to remain together.

Obviously, further research needs to be conducted to take the results found here to the next level. The author is confident that a better set of boundaries than that of the National Climate Data Center can be created. Whether it is through a more visual approach (which worked relatively successfully in this study) or a more statistical approach remains to be seen. Although one must also give some validity to the observation that sometimes systems that have been created over a long period of time, and adjusted over the years, can be a powerful tool, and a challenge to replace.

BIBLIOGRAPHY

Bednarz, S.W., Boehm, R.G., Downs, R.M. (1994). Geography for Life: National Geography Standards. Washington, D.C.: National Geographic Society.

Block, J. (2000). The Use of Historical, Current, and Forecast Weather and Climate Data in a GIS: 2000 ESRI International User Conference. San Diego, CA.

Davis, J.C. (1986). Statistics and Data Analysis in Geology. New York: John Wiley & Sons.

De Blij, H.J., Muller, P.O. (1997). Geography: Realms, Regions, and Concepts. New York, NY: John Wiley & Sons.

Everitt, B. (1974). Cluster Analysis. London: Heinemann Educational Books, Ltd.

Guttman, N.B., Quayle, R.G. (1995). A Historical Perspective of U.S. Climate Divisions. National Climatic Data Center. Ashville, NC.

Hare, F.K. (1951). *Climate Classification.* In McBoyle, G (Ed). Climate in Review. (pp. 96-109). Boston, MA: Houghton Mifflin.

Hartkamp, A.D., De Baeurs, K., Stein, S., White, J.W. (1999). Interpolation Techniques for Climate Variables. NRG_GIS Series 99-01. Mexico, D.F.: CIMMYT.

Jackson, B.B. (1983). Multivariate Data Analysis: An Introduction. Homewood, Ill: Richard D. Irwin, Inc.

Jain, A.K., Dubes, R.C. (1988). Algorithims for Clustering Data. Englewood Cliffs, NJ: Prentice Hall.

Lancaster, P., Salkauskas, K. (1986). Curve and Surface Fitting: An Introduction. London: Academic Press.

Linacre, E. (1992). Climate Data and Resources: A Reference and Guide. London: Routledge.

Lynch, S.D., Schulze, R.E. (1995). Techniques for Estimating Areal Daily Rainfall. 1995 ESRI International User Conference. Palm Springs, CA.

Lydolph, P.E. (1985). The Climate of the Earth. Totowa, NJ: Rowman & Allanheld.

McBoyle, G.R. (1971). *Climate Classification of Australia by Computer*. In McBoyle, G (Ed). Climate in Review. (pp.110-118). Boston, MA: Houghton Mifflin.

McGrew, J.C., Monroe, C.B. (2000). An Introduction to Statistical Problem Solving in Geography. Boston, MA: McGraw-Hill.

Mather, J.R. (1974). Climatology: Fundamentals and Applications. New York: McGraw-Hill.

Plantico, M.S., Goss, L.A., Daly, C., Taylor, G.H. (2000). Development of a new U.S. Climate Atlas: 11th Symposium on Global Change Studies, Paper P1.12. Long Beach, CA.

Polyak, I. (1996). Computational Statistics in Climatology. New York: Oxford University Press.

Robeson, S.M. (1997). *Spherical methods for spatial interpolation: Review and Evaluation*. Cartography and Geographic Information Systems, 24 (1). 3-20.

Sanderson, M. (1999). *The Classification of Climates from Pythagoras to Koeppen*. Bulletin of the American Meteorological Society, 80 (4). 669-773.

Shipley, S.T., Graffman, I.A., Ingram, J.K. (2000). GIS Applications in Climate and Meteorology. 2000 ESRI International User Conference. San Diego, CA.

StatSoft, Inc. (2001). Electronic Statistics Textbook. Tulsa, OK: StatSoft. WEB: http://www.statsoft.com/textbook/stathome.html.

Taylor, G.H., Hannan, C. (1999). The Climate of Oregon: From Rain Forest to Desert. Corvallis, OR: Oregon State University Press.

Triola, M.F. (1998). Elementary Statistics. Reading, MA: Addison-Wesley.

Data Sources

Oregon Climate Service (OCS) - Climate data
http://www.ocs.orst.edu/

Oregon Geospatial Data Clearinghouse (OGDC)
http://www.sscgis.state.or.us/

United States Environmental Protection Agency (EPA)
http://www.epa.gov/airmarkets/cmap/data/

Western Regional Climate Center (WRCC) - Climate data
http://www.wrcc.dri.edu/