The Problem of Redistricting: the Use of Centroidal Voronoi Diagrams to Build Unbiased Congressional Districts

Stacy Miller

May 16, 2007

Abstract

This paper is a development of the use of MacQueen's method to draw centroidal Voronoi diagrams as apart of the redistricting process. We will use Washington State as an example of this method. Since centroidal Voronoi diagrams are inherently compact and can be created by an unbiased process, they could create congressional districts that are not only free from political gerrymandering but also appear to the general public as such.

1 Introduction

The purpose of this project is to use centroidal Voronoi diagrams (CVDs) within a mathematical approach to redistricting. One of the benefits of using a mathematical approach instead of wholly relying on the political process and its committees is that the political parties of incumbents do not then have the opportunity to influence the drawing of congressional lines. If they did have the opportunity, they could be motivated to draw the new lines so that they concentrate the voting strength of the other party in as few districts as possible and extend their own party's power as the majority in as many districts as possible. This practice, called gerrymandering, usually results in oddly shaped districts. In response, the general public is very suspicious of non-compactly shaped districts. Since CVDs are inherently compact, they could greatly aid the process of redistricting for any state. As an example, we will draw CVDs over Washington State using a modified MacQueen's method.

The population movement during the Industrial Revolution from rural to urban areas created congressional districts that were very unequal in population to each other. Rural districts would elect a representative for considerably less people so that the concerns of rural residents were given proportionately more attention in the House of Representatives. The Federal Reapportionment Act of 1911 temporarily addressed the problem. It required congressional districts to be equal in population so that they had to be redrawn in response to the changes in population, but this mandate was only to be until 1929 [11]. This problem was not readdressed until 1962. Until then, rural-urban migration continued to cause so much change in the population that by 1946, the largest district in Illinois had more than eight times as many people than the state's smallest district [11].

Supreme Court Mandate This problem of unequal representation eventually came before the Supreme Court in the case of Baker v Carr (1962) [1]. As a part of its ruling, the Supreme Court stated the most important criteria of reapportionment to be "One man, one vote", meaning voting districts had to contain equal portions of the state's population. Districts that did not contain equal populations violated the Fourteenth Amendent: "equal protection of the laws" [11]. The Court however did not give a clear maximum for the allowed and inevitable population deviation. In the Reynolds v Sims case (1964), the Court allowed state legislature districts more of a deviation range than for congressional districts [1]. The size of these deviations are further discussed in cases Wesberry v Sanders (1964), Wells v Rockefeller (1969) and Kirkpatrick v Preisler (1969) [1]. In Wesberry v Sanders, the Court stated that "as nearly as practicable, one man's vote in a Congressional election be worth as much as another's." In the Wells v. Rockefeller the Court further elaborated that "the "nearly as practicable" standard requires that the State make a good faith effort to achieve precise mathematical equity." Without any explicit maximum for the population deviation, the Supreme Court's mandate of equal population comes down to the statements "nearly as practicable" and "a good faith effort." Most importantly, neither of these statements answer the questions of *who* is to carry out the mandate of equal population or *how*.

Enacting the Mandate As state legislators attempted to redraw the voting districts, the problem of gerrymandering quickly became evident. Consequently, some states, including Washington, have turned to bipartisan committees to redraw the districts. This process, though, has proven to be very time-consuming, lasting for months at a time, and is still vulnerable to bias.

An Unbiased Method To achieve unbiased congressional districts of equal population, a mathematical method must be used. The focus of this project is to incorporate CVDs into an unbiased redistricting of Washington state. In Section 2, we will consider criteria used in past redistricing projects. Section 3 will review the past mathematical methods used. Section 4 will describe the available data. Section 5 will introduce CVDs and a modified MacQueen's Method. Section 6 will discuss the results and unresolved problems with the techniques present.

2 The Criteria of Reapportionment

There are many debates surrounding the criteria of redistricting, most of which could be discussed at great length. Below are listed those criteria most mentioned in the literature.

1. **Equal Population:** Congressional districts of equal population are required by the Supreme Court, as discussed in the above section. It should be noted however, that equal population does not mean equal number of registered voters or even of potential voters: all those who are allowed to vote whether or not they are registered (U.S. citizen, no felony convictions, etc.). Equal population refers to equal numbers of the total population.

2. **Compactness:** Districts of compact size are seen by the public as unbiased, while more creatively shaped districts seem suspicious. An unbiased method of redistricting will only be useful if the public believes it to be unbiased. It follows that compactness will be an important criterion in this project.

3. **Contiguousness:** Keeping districts together as one mass instead of dividing them is widely accepted as practical and essential to proper representation. Also, if the districts were not expected to be contiguous, the possibilities for gerrymandering would be endless.

4. **Respect For City and County Lines:** Some argue that this condition would best represent the interests of individual cities and counties. However, this view is not widely accepted. Many rural communities feel better represented when their district includes a rural community from another county than when it includes a city within the same county. Furthermore, it is difficult to draw compact districts within cities and counties that are not compactly drawn themselves.

5. **Preserving Communities of Interest:** Communities of interest include communities determined geographically (inland vs.coastal), by population (rural vs.urban) and racially (minority communities). This piece of criteria is one of the most hotly debated. While many of the arguments on both sides have merit, drawing district lines that preserve these communities would be an immensely complicated process that not only would include debate over which communities to preserve, but also would conflict with the criteria of compactness.

6. **Creating Competitive Districts:** Creating districts that are competitive (not homogenous in interests) directly conflicts with preserving communities of interest. Competitive districts would include enough of different political opinions and stances so as to be individually representative of the state in general. Again, a lot of debate surrounds this topic and it would likely interfere with the criteria of compactness.

The first three requirements do not conflict with each other and are the most widely accepted. The last three requirements are highly debated. They conflict with each other as well as the important criteria of compactness. The inclusion of the last three requirements will make any redistricting much more complicated and more suspicious to the public. A redistricting that the public will trust to be unbiased must appear as uncomplicated and straightforward: equal in population, contiguous and compact. Hence we will consider past methods of redistricting in regards to this set of criteria.

3 Established Methods of Redistricting

The past methods of redistricting can be divided up into four general methods each designed to redistrict an area made up of population units that can range from city blocks to large rural counties [4]. The population units in a given area to redistrict are not necessarily equal in spatial size or population.

3.1 The Nagel-Kaiser Method

This method was used by Nagel in 1965 and by Kaiser in 1966. What sets this method apart is that it begins with a pre-existing district plan and makes "swaps" of population units from one district to another. These "swaps" minimize an objective function composed of three parameters: the number of over-populated districts, a measure of the average deviation from perfect population equality and the average of the deviation from a population-weighted compactness baseline [11]. In Figure 1 for example, the different shades designate the districts and the lines designate the population units. If the population in unit 1 had declined while the population in unit 2 had increased, then this method would consider switching unit 2 to the dark grey district.

A potential problem of this method is that it does not provide a *global* solution, meaning it does not necessariliy work for each case. However, this method is more flexible in considering other political criteria.

3.2 The Vickery Method

Vickery published this method in 1961. It is an iterative heuristic that was not originally intended for computers. However, Vickery designed this method to be "completely mechanical" with "no room at all for human choice" [11].

1. Begin with an arbitrary population unit A
2. Pick the population unit B_1 farthest from A–the seed base unit: this step forces the unit B_1 to be on the periphery.
3. Add the closest units to B_1 to ensure compactness. To fulfill the criteria of equal population, or that the population of each district is equal to the total population of all units divided by the number of districts, continue

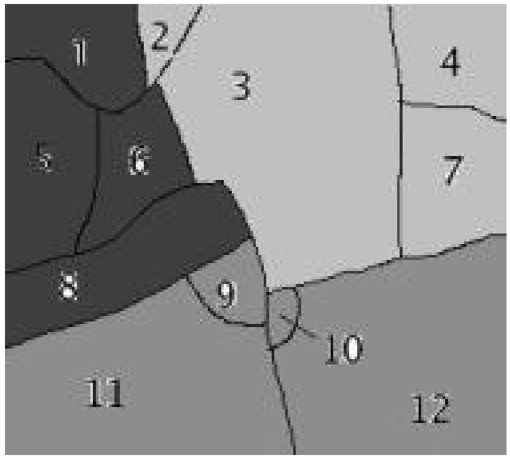

Figure 1: An existing district plan [7]

adding until the condition arises when adding the closest *unassigned* unit would increase the district's population to above the given quota. Figure 2 shows the seed base unit B_1 as the unit farthest from unit A and surrounded by the population units of district 1.

4. Repeat step 3 with the next furthest population unit from A: B_2. Again, this new unit, B_2 will be on the periphery because the furthest units from A are on the periphery.

5. Continue repeating step 3, each time with the next furthest population unit from A until all the population units are assigned, except for perhaps any "residuals." Residuals are units whose adjacent districts cannot add the residual without going over the population quota. Hence, it is impossible to assign them without going over the population quota or violating the criterion of contiguity by assigning it to a non-adjacent district.

6. Repeat steps 1-5 again with a new arbitrary starting population unit in step 1. The two different plans would both be presented to a state legislature who would decide on which is more appropriate for any additional criteria.

As mentioned above, this method can produce residual population units. This hindrance prevents it from guaranteeing a global solution.

3.3 The Garfinkel and Nembauser Method

Garfinkel and Nemhauser presented this method in 1970. Unlike the other methods presented here, this is an exact method, not a heuristic, meaning that the solution *is* optimal for the given objective and constraints. It was specifically

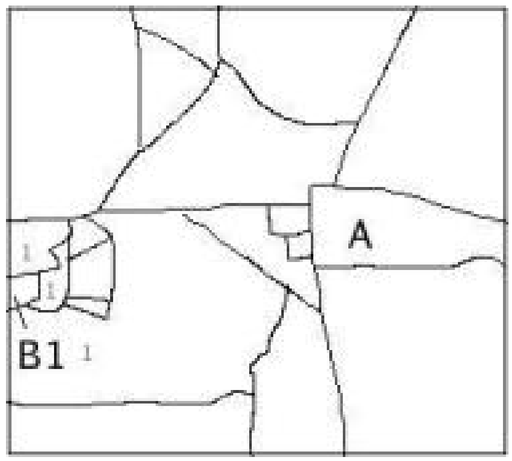

Figure 2: Building the first district by the Vickery Method [7]

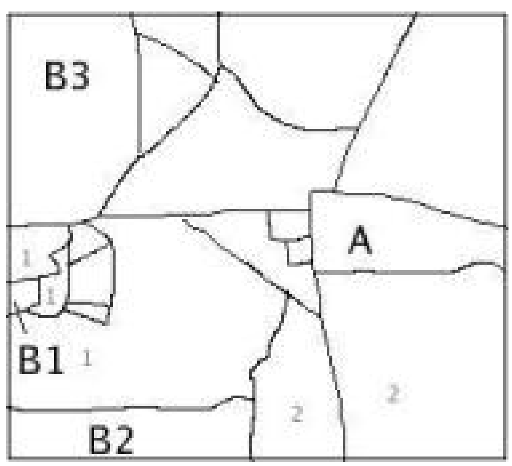

Figure 3: Creating additional districts by the Vickery Method [7]

designed for a computer. Its first step is the tree search method that is somewhat similar to the Vickery method. It generates all the possible districts that have equal populations, are compact enough according to the given constraints and are contiguous. The next step is to build a matrix with the feasible district plans, and to then use the implicit enumeration technique. This technique uses partitioning integer linear programming to remove the less than optimal district plans according to the constraints that correspond to the criteria of contiguity, compactness and equal population [4]. While an exact method like the Garfinkel and Nembauser Method is very appealing, it is computationally very demanding.

3.4 The Weaver and Hess Method

This method, presented in 1963, was the first major computer redistricting method [11]. It is a heuristic optimization model that incorporates the warehouse-location model.

1. Begin with an arbitrary set of district centers
2. Use the Warehouse-Location Model, as described below, to minimize the population weighted distances between each district's center and each of the centers of its population units. This minization can produce "split" population units where part of a unit is assigned to one district and the rest of the unit to another.
3. To correct for any split population units, assign any split units to the districts that already serve the greatest proportion of the split unit's population. These new assignments could cause a significant deviation in the district population.
4. To address the new, possibly larger, population deviations, repeat step 2-3 until the centers converge.
5. Start over with a new set of arbitrary centers
6. Select the best solution according to the criteria of equal population, compactness and contiguity.

The split districts are potentially a large problem as they can interfere with the criteria of equal population and compactness.

Warehouse-Location Model

- N: number of population units and consequently the maximum number of districts
- c_{ij} is the Euclidean distance between the center of district i and the center of population unit j, multiplied against the population of population unit j: Q_j

- X_{ij} is the proportion of the population of population unit j that is included in district i ($X_{ij} \leq 1$)
- Constraint: every member of every population unit is assigned to a district
- Objective: minimize $\Sigma_{i=1}^{N} \Sigma_{j=1}^{N} c_{ij} X_{ij}$

Lagrangian Relaxtion Method Hojati [4] uses a Lagrangian Relaxation heuristic to solve this warehouse location problem. Keeping with the above notation, Hojati uses the supply and demand constraints:

- Q: the district quota, or the total population of all the units divided by the number of districts
- $Y_i = 0$: there is not a 'warehouse' or district center in population unit i
- $Y_i = 1$: there is a 'warehouse' or district center in population unit i
- Demand constraint: $\Sigma_{i=1}^{N} X_{ij} = 1$
  This constraint is such that for every population unit j, the sum of the proportion of the unit's population assigned to all districts i, $i = 1, ..., N$, must equal one. More simply, the entire population of each population unit must be assigned to a district.
- Supply constraint: $\Sigma_{j=1}^{N} Q_j X_{ij} = QY_i$
  The inclusion of the supply constraint makes this a capacitated warehouse-location problem. This constraint means that the population of the ith district cannot have more or less than the quota, or that all the proportions of a district j, $j = 1, ..., N$, assigned to all districts i, $i = 1, ..., N$, must equal the given quota.

Hojati then sets up an equation with the objective function including two Langrangian function multipliers multiplied against their respective constraints. With a number of simplifications, he then proceeds to use a sub-gradient optimization to find the best values for the Lagrangian coefficients.

Like the Vickery Method, the Weaver and Hess Method is a heuristic method that includes the risk of split districts, and therefore cannot guarantee a globally optimal solution. The method of redistricting described in this paper is similar to the Weaver and Hess method since they both include converging centers but our method only uses population units to find where the centers converge. Before the discussion of our new method, we will introduce the data that we will be using within that method.

4 Washington State Population Data

The Washington State data is collected from the Census Bureau's 2004 U.S. population estimates since it is the most readily available. The data consists of

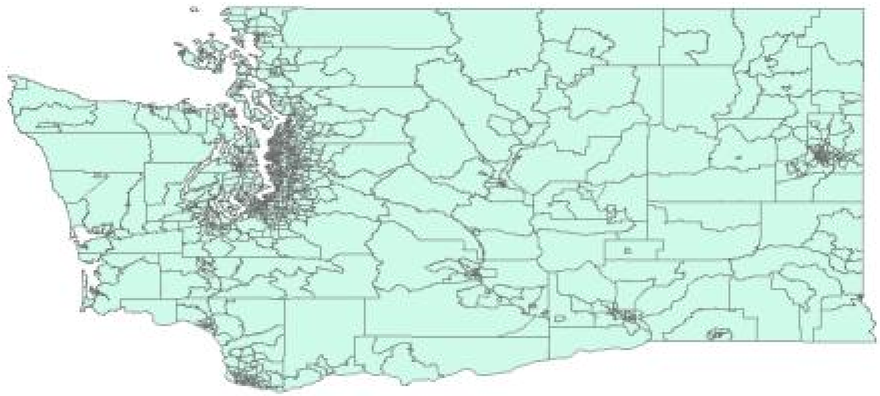

Figure 4: The Census Tracts of Washington State [6]

census tract identification numbers, population, population density (population per squared mile) of each census tract and the latitudes and longitudes of the geographic center of each census tract. There are about 1,318 census tracts in Washington out of 65,443 nationally.

Census tracts are drawn for the U.S. Census Bureau by local committees. They are small sub-divisions of a county and contain between 1,500 and 8,000 people and average about 4,000 people [10]. They are designed to be relatively socio-economically homogeneous such that the habitants have similar characteristics, economic status and living conditions [10]. The spatial size of census tracts vary depending on whether they are in an urban area or a less densely populated rural area as shown by Figure 4. The boundaries usually follow geographical features.

The 2004 population estimates of the census tracts are estimated by the "Distributive Housing Unit Method" that uses a census tract's housing occupency rate, average persons per household and estimated number of housing units (geographically updated in the decennial census) in addition to the county population estimate [10]. The annual county population estimates are based on the decennial census and on the following estimates for the years subsequent to the decennial census:

1. Number of births
2. Number of deaths
3. Net domestic migration

4. Net international migration
5. Net Armed Forces movement to and from overseas
6. Net native emigration from the US
7. Changes in group quarters (correctional institutions, university dormitories, military barracks, nursing homes, etc.)

The population density data will be used to form a population density equation within a modified MacQueen's method that will be described within the next section.

5 The Centroidal Voronoi Diagram

The Voronoi diagram is essentially a partition of an area, Ω, in reference to a set of points called cluster centers, defined below, such that each point of the area is grouped with the cluster center to which it is closest. The resulting boundary line between any two cluster centers is a bisector of the line segment connecting the two points. An example is shown in Figure 5 with a bisector illustrated in the lower right corner. Another characteristic of Voronoi diagrams is that they are inherently compact because they assign points of the area to the closest cluster point.

5.1 Voronoi Polygons

We denote the ordered set of cluster centers as $C = \{\mathbf{z}_k \in \Omega \mid k = 1, 2, ..., K\}$, with K as the number of cluster points, and the ordered pair $\mathbf{z}_k$ containing the coordinates of the cluster point k. We designate the Voronoi polygons (or Voronoi clusters) as

$$P(\mathbf{z}_k, C, \Omega) = \{\mathbf{x} \in \Omega \mid d(\mathbf{x}, \mathbf{z}_k) \leq d(\mathbf{x}, \mathbf{z}_j)\},\ j = 1, 2, ..., K$$

$$\text{If } d(\mathbf{x}, \mathbf{z}_k) = d(\mathbf{x}, \mathbf{z}_k) \text{ then } k \leq j,\ k = 1, 2, ..., K$$

The point $\mathbf{x}$ contains the coordinates of a point in the set Ω. The quantity $d(\mathbf{x}, \mathbf{z}_k)$ represents the distance between the point $\mathbf{x}$ and the cluster center point $\mathbf{z}_k$. Our representation of the Voronoi polygon of $\mathbf{z}_k$ means that it contains all the points $\mathbf{x}$ in Ω that are closer to $\mathbf{z}_k$ than any other $\mathbf{z}_j$, $j \neq k$. However, if $\mathbf{x}$ is equidistant from $\mathbf{z}_k$ and $\mathbf{z}_j$, it is assigned, arbitrarily, to the cluster center with the smaller index. For basic properties of Voronoi diagrams see Atsuyuki Okabe's "Spatial Tessellations: Concepts and Applications of Voronoi Diagrams" [8].

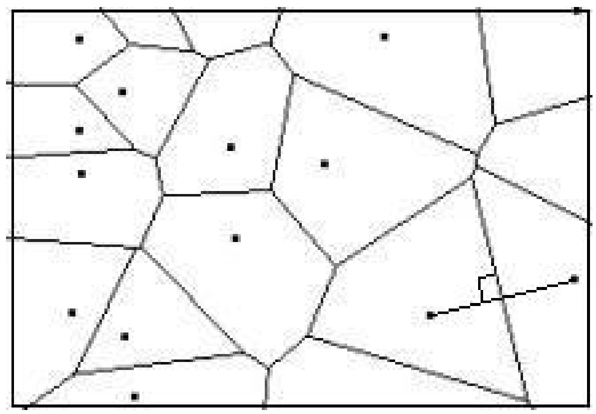

Figure 5: An example of a Voronoi diagram [7]

5.2 Centroidal Voronoi Diagrams

Given a density function $\rho(\mathbf{y}) \geq 0$ for $\mathbf{y} \in \Omega$ and a region $V \in \Omega$, let

$$\mathbf{z} = \frac{\int_V \mathbf{y}\rho(\mathbf{y})d\mathbf{y}}{\int_V \rho(\mathbf{y})d\mathbf{y}}$$

then $\mathbf{z}$ is the mass centroid of the region V [3]. When both the following are true:

1. The polygons $\{V_i\}_{i=1}^k$ are Voronoi polygons for the set of points $\{\mathbf{z}_i\}_{i=1}^k$
2. The points $\{\mathbf{z}_i\}_{i=1}^k$ are the mass centroids of the polygons $\{V_i\}_{i=1}^k$

then the Voronoi diagram may be classified as a CVD of the area $\bigcup_{i=1}^k V_i$ [3]. While CVDs contain compact polygons, they also contain equal energy meaning the total distance between each point in a polygon and a polygon's centroid is equal to that of the other polygons in the diagram [3]. Another characteristic of CVDs are that they are more aesthetically pleasing than general Voronoi diagrams. Figure 6 compares a general Voronoi diagram and a CVD over the same size area with constant density. The general Voronoi diagram has more variation in the area of its polygons and in general the polygons are less compact.

5.3 A Modified MacQueen's Method

MacQueen's method is designed to generate CVDs as described by Du, Gunzburger and Ju in their article "Probabilistic Methods for Centroidal Voronoi Tessellations and their Parallel Implementations" [5]. We will apply a modification of MacQueen's method that is more suited to the available census tract data.

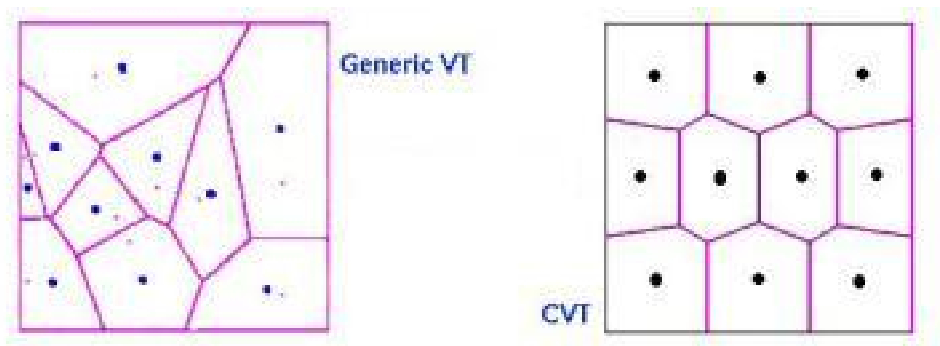

Figure 6: Voronoi Tessellation and Centroidal Voronoi Tessellation over an Area with Constant Density [2]

1. **Begin with an inital set of k points, $\{z_{i=1}^{k}\}$ in the region Ω found by a Monte Carlo method. Let $j_i = 1$ for all $i = 1, ..., k$.**
   A Monte Carlo method is any method that generates random numbers but only uses those that satisfy previously determined properties [5]. In this project however, we will not use a Monte Carlo method. Instead, we will chose points from the top nine population centers in Washington State. In our case of redistricting, we will let the area Ω be Washington State and we will let $k = 9$ because Washington has 9 congressional districts.

2. **Select a census tract and its geographic center, x, in Ω at random using a probability density function based on population: $\rho(y)$.**
   The point, **x**, will be the geographic center of a randomly selected census tract. The probability that the census tract will be chosen is the population density of the census tract divided by the sum of the population densities of all the census tracts in Washington state.

3. **Find the closest initial point, z_i, from step 1, to the randomly selected point x from step 2.**
   To put it in terms of redistricting, we will find the closest center point to the randomly selected census tract using the Haversine function that is the most appropriate function for geographic distances since it considers the curve of the earth's surface within its calculation. The longitude and latitude components, both in radians, of the first point are entered in as (x_1, y_1), and the second point is similarly entered in as (x_2, y_2). The distance, D, is given in kilometers:

$$\Delta x = x_2 - x_1$$

$$\Delta y = y_2 - y_1$$

$$a = (\sin\left(\frac{\Delta y}{2}\right))^2 + \cos(y_1)\cos(y_2)(\sin\left(\frac{\Delta x}{2}\right))^2$$

$$c = 2\arctan\left(\frac{\sqrt{a}}{\sqrt{(\sqrt{a})^2 + (\sqrt{1-a})^2} + \sqrt{1-a}}\right)$$

$$D = (6371)(0.62)(c)$$

4. **The coordinates of z_i will be reset such that**

$$\mathbf{z}_i \leftarrow \frac{j_i\mathbf{z}_i + \mathbf{x}}{j_i + 1}$$

$$j_i \leftarrow j_i + 1$$

   What these expressions mean is that the closest center point, $\mathbf{z}_i$, is reset in the first iteration of this method ($j_i = 1$) by an average of its coordinates and the coordinates of the tract center, $\mathbf{x}$. The parameter j_i is reset to two such that for the next iteration, the next randomly selected tract center that is closest to this center point, $\mathbf{z}_i$, will have less influence on the coordinates of the new $\mathbf{z}_i$. For each time a center point is found closest to a randomly selected census tract, the tract center will have less and less influence over the new coordinates of $\mathbf{z}_i$.

5. **When the $\mathbf{z}_i$'s satisfy a certain convergence criteria, end the iterations**.
   We will end the iterations after a set number have been performed. Other possible convergence criteria includes stopping once a certain number of consecutive movements of $\mathbf{z}_i$, for $i = 1, 2, ..., k$, are each less than a specified distance.

With this modified MacQueen's method, we can find the mass centroids based on population density. These centroids will allow us to draw a CVD over Washington State to produce congressional districts that are compact, contiguous and equal in energy. Once these districts are created, we may then examine how successful they are in achieving equal population. This method has not been carefully reviewed and there may be unexamined problems that affect our results.

6 Results

In this section, we will first review how successful the current congressional districts of Washington State are in meeting the criteria of compactness and equal population. Then we will examine the results of running our modified MacQueen's method three separate times starting with the same set of initial points for the same number of iterations. We will also examine the results of an alternate run of our modified MacQueen's method that begins with a set of initial points that are all placed in eastern Washington. Finally, we will compare these results to another run of the modified MacQueen's method starting with the same set of initial points but for more iterations.

Figure 7: 110th Congressional Districts of Washington [6]

6.1 The Current Congressional Districts

Figure 7 displays the current congressional districts overlaying county lines. These districts respect county lines whenever possible. Keeping counties together however, often makes it more difficult to keep the districts both equal in population and relatively compact since the large discrete sets of population in relatively non-compactly shaped counties cannot be divided. Also we note from Figure 7 that two small districts span over Puget Sound, giving the state a total of nine congressional districts. Overall the congressional districts of Washington State are drawn compactly and appear to be relatively free from gerrymandering unlike the controversial 110th congressional districts of Texas in Figure 8. However, we are testing our method on Washington State because of its relevance to the author.

Table 1 shows the total population of each district in addition to the percentage of the state population in the district and the difference of that percentage from the goal: $\frac{100}{9}\%$. We can see that the population deviation is relatively good. In absolute value, the largest deviation is 0.00775%. The population totals in this table are calculated by the Census Bureau with their "Sample" method as opposed to their "100%" method.

We observe in Figure 9 which districts in Washington State have lower population deviation in absolute value: the darker the district the higher the devi-

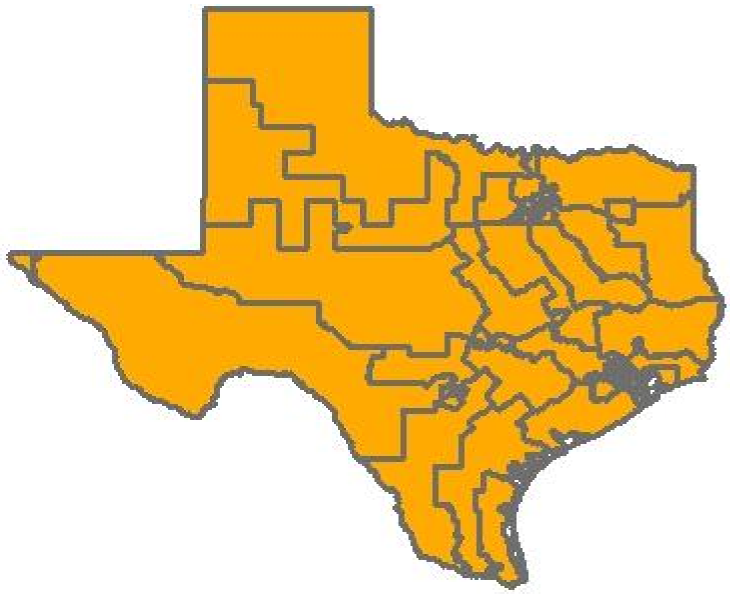

Figure 8: 110th Congressional Districts of Texas [6]

Total Population	Percentage of Total Population	Population Deviation in Percent
654799	11.109358%	-0.001753%
654984	11.112497%	0.001386%
654992	11.112632%	0.001521%
654851	11.110240%	-0.000871%
654935	11.111665%	0.000554%
655068	11.113922%	0.002811%
655016	11.113040%	0.001928%
655029	11.113260%	0.002149%
654447	11.103386%	-0.007725%

Table 1: 2006 Population Estimates of the Current Congressional Districts

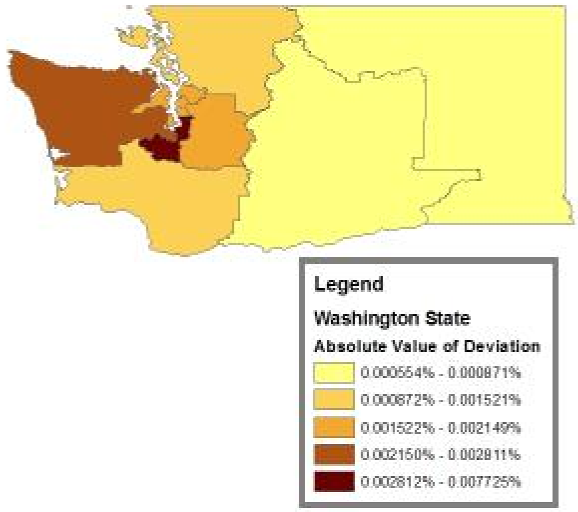

Figure 9: Absolute Value of the Deviation from Equal Population in Percent [6]

ation, the lighter the district the lower the deviation. Eastern Washington has the lowest and the area surrounding Puget Sound has higher deviation, especially to the west and south. The darkest district in Figure 9 has a negative deviation meaning its population is less than the correct proportion. The second darkest district has a positive deviation. Figure 10 compares the absolute value of the population deviation in Washington's districts to the rest of the Continental United States. The districts with higher deviation are shown in darker colors. Some of the states of the districts with higher deviations are Idaho, Nevada, New Mexico, Lousiana, Alabama, Arkansas, the Carolinas, New Hampshire, Maryland and West Virginia.

A further examination of this data yields some interesting questions about how the Census Bureau calculates these population estimates. First, the sum of the population estimates of all the districts is $5,894,121$, while the Census Bureau estimates the total state population to be $6,395,798$. The difference, about $500,000$, is very significant. Another curiosity is that the sum of the 2006 county population estimates of the counties within the easternmost district is $702,268$ while the Bureau's 2006 population estimate for that district is $654,935$. This district is comprised of only these counties and completely encloses these counties except for a small portion of Adams County. However, the population of that particular portion cannot explain the difference of almost $50,000$ because the 2006 total county population estimate of Adams County is only 16887. So even if the whole population of Adams County was in that small portion outside of the district, there would still be a difference of almost $35,000$ in the population

District Number	Roeck Ratio
1-Northern Seattle	0.293089
2-Bellingham	0.546939
3-Vancouver	0.261801
4-Central Washington	0.300806
5-Eastern Washington	0.126157
6-Olympic Peninsula	0.234736
7-Seattle	0.365203
8-Seattle Outgrowth	0.463537
9-Olympia	0.300019

Table 2: Roeck Ratio of the Current Congressional Districts

estimates for the district. These large inconsistencies in the data suggest further inquiry into the population estimate techniques of the Census Bureau is needed.

Now we consider the compactness of these current districts. We will do so by using the Roeck ratio which is the ratio of the area of the district over the area of the smallest circle that completely encloses the district. The closer the ratio is to one, the more compact the district. Table 2 lists the Roeck ratios for each district.

As we look at these ratios, we should consider that for a district to have a Roeck ratio of 1, it must be a perfect circle which would mean very low Roeck ratios for the surrounding districts. A more fitting comparison for our districts is the perfect square that has a Roeck ratio of $2/\pi = 0.6366$. Also, we must remember that these districts are being drawn within a relatively non-compactly shaped state. Figure 11 displays the districts according to their Roeck ratio: the darker the district means the more compact it is. The central districts have the highest Roeck ratios while, not surprisingly, one of the districts that spans over Puget Sound has the lowest.

Given all of these observations on the current congressional districts, we will now compare them to three different district plans drawn using CVDs with centroids found by our modified MacQueen's method.

6.2 Centroidal Voronoi Diagram-Based Congressional Districts

From three independent runs of the modified MacQueen's method, each for 150 million iterations starting with the same set of initial points and each subsequent to a different set of random numbers (step 2 of MacQueen's), we have three different sets of centroidal centers for Washington State. Figures 12, 13 and 14 display the corresponding CVDs with the corresponding centroidal points. Fig-

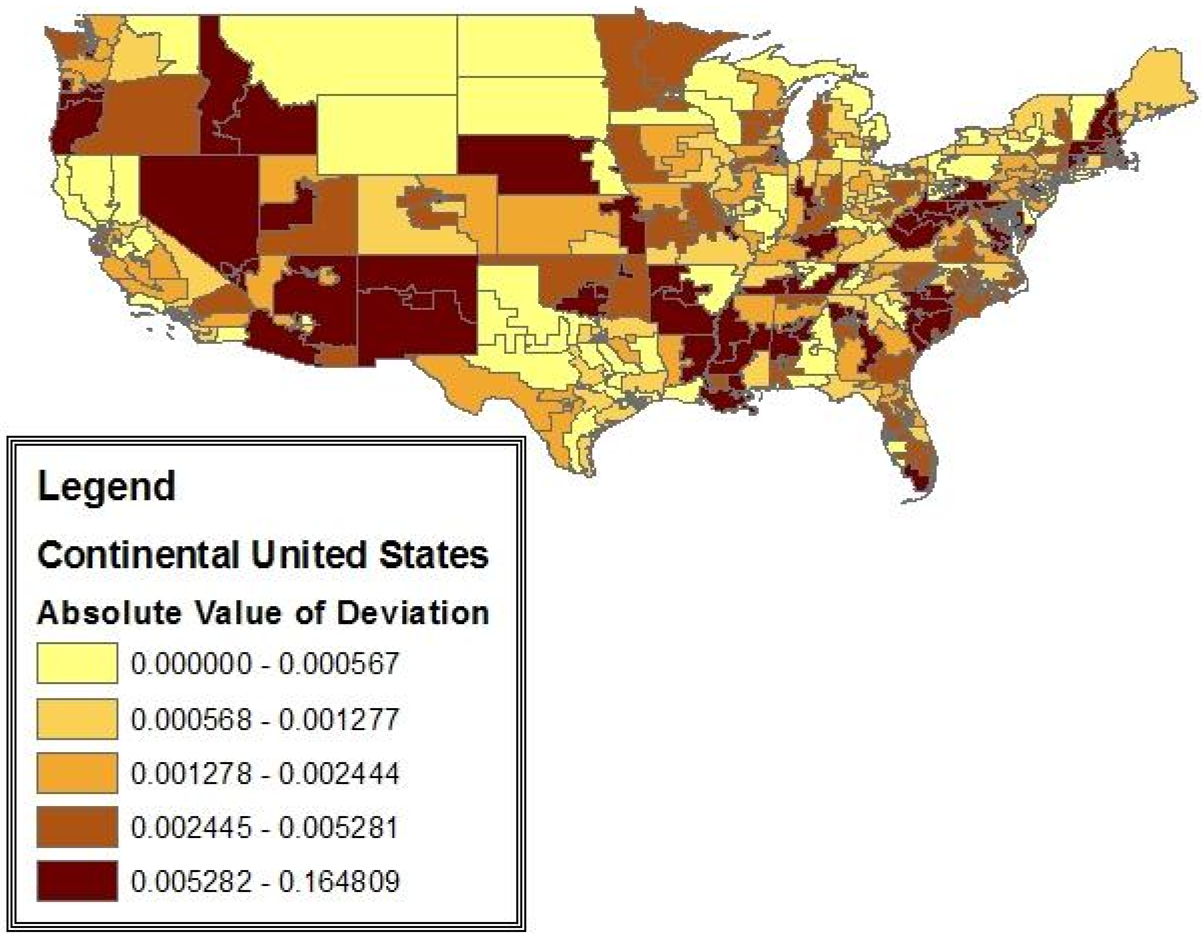

Figure 10: Absolute Value of the Deviation from Equal Population in Percent of the Continental United States [6]

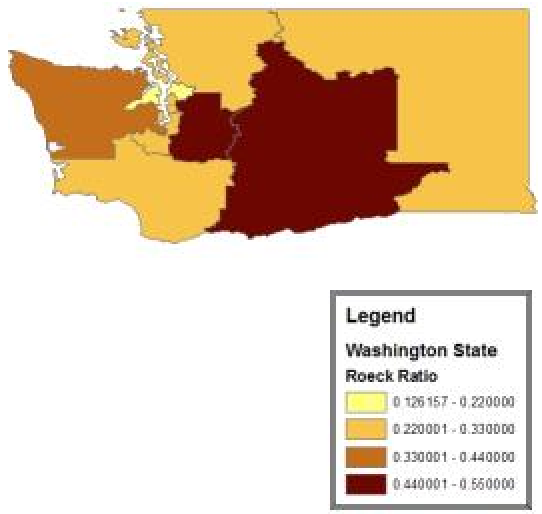

Figure 11: Roeck Ratios of the Current Congressional Districts of Washington [6]

ure 15 shows the three diagrams together revealing a large degree of instability within the results. This lack of stability might be resolved by simply running the modified MacQueen's method for more iterations, or using a Monte Carlo method to find the initial points, or using a continuous population density function instead of a discrete function.

To calculate population estimates for the districts drawn from CVDs, we use the population data of the census tracts. The population of each census tract is assigned to the district whose centroid is closest to the census tract's geographic center. This method only gives a rough estimation. Furthermore, the census tract population estimates are from the year 2004 because they were not readily available for the years 2005 and 2006. Table 3 compares the 2006 current district population estimates to our rough 2004 population estimates of our CVD-based districts. The last row of the table gives the difference between the largest and smallest district population of each district set.

Given the questionability of the current district population estimates, the rough estimates based on census tracts and the difference of 2 years in the data, the results, while not impressive, are also not conclusive. The large deviation in the populations of the CVD-based districts might be cleared by more reliable data or resolved by solutions mentioned above: more iterations, using a Monte Carlo method or a continuous population function.

However, there remains a fundamental problem with the CVD: equal energy instead of equal population. The CVD, similar to the uncapacitated warehouse-

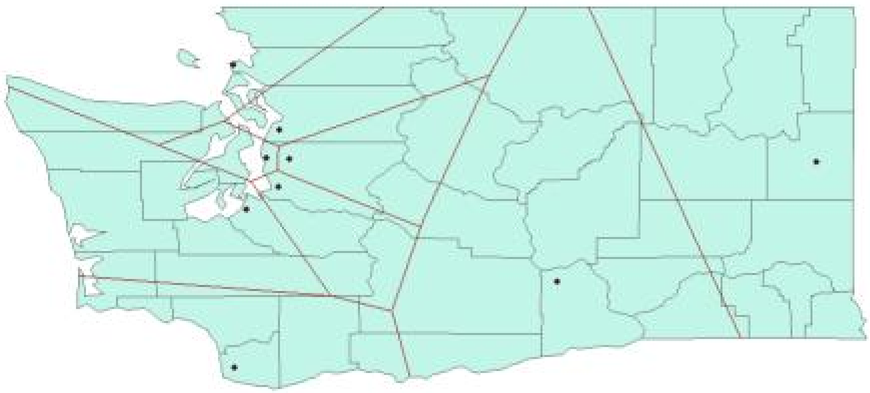

Figure 12: First Set of the CVD-Based Districts of Washington [6]

Figure 13: Second Set of the CVD-Based Districts of Washington [6]

Current	First CVD	Second CVD	Third CVD
654799	615634	539531	610259
654984	684289	429944	511327
654992	854434	400906	929671
654851	787900	990066	718284
654935	636725	501469	500615
655068	686268	755819	854239
655016	481501	735582	610492
655029	570619	909689	863893
654447	853830	906405	568304
621	372329	589160	429056

Table 3: Comparison of District Population Estimates

Figure 14: Third Set of the CVD-Based Districts Washington [6]

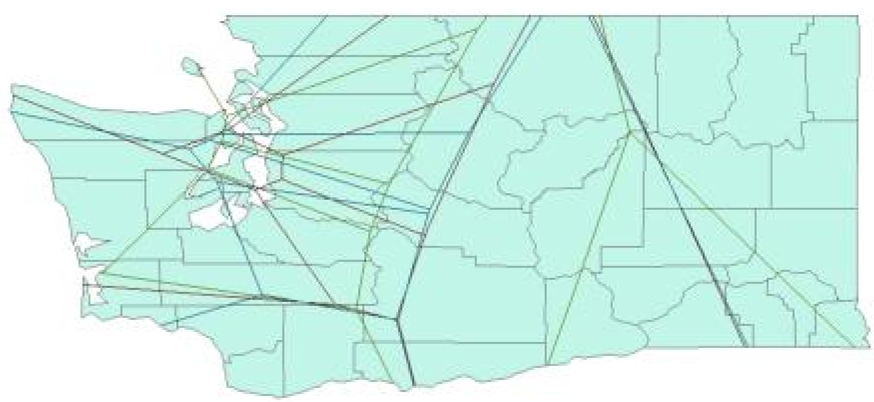

Figure 15: The Three CVD-Based District Sets of Washington Overlaying Each Other [6]

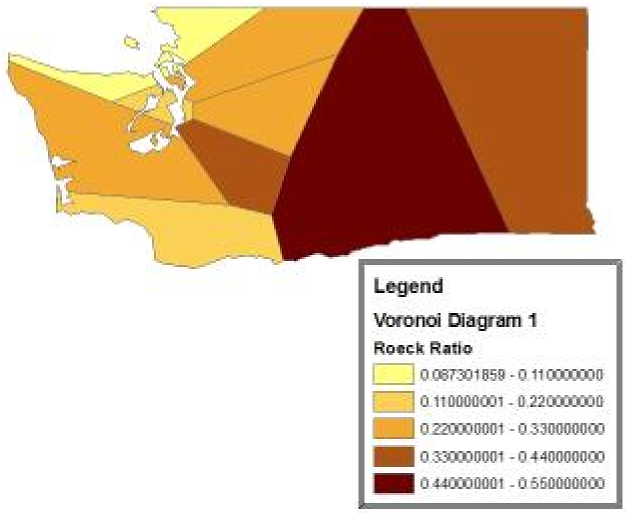

Figure 16: Roeck Ratio over First CVD [6]

location model (the model without the supply constraint), equalizes the energy that each district must expend to bring its total mass to the center. In terms of the warehouse-location model, it minimizes the total square of the distance over which goods are delivered to consumers from factories. The end result is that a sparsely populated district may receive one factory while a smaller-sized district with a higher population density and perhaps higher total population will only receive one factory. The total distance that the goods must travel in each district would be equal, but the number of consumers in each district would not necessarily be equal. In terms of redistricting, the CVD is composed of districts whose total distance between its centroid and each member of its population is equal to that of other districts. If we were designing districts to equalize the total distance traveled by voters to a voting center in each district, then the CVD would be extremely useful on its own. However, while this diagram lacks any constraint on the total mass or total population in each of its polygons, it cannot solely be relied upon to draw districts of equal population. The bias of this method is towards the rural population who would receive disproportionately more representation. It is interesting to note that this bias has been the case historically, given the ongoing urbanization of the U.S. population and the subsequent need to update district boundaries for the changing populations.

Table 4 lists the Roeck ratios for the current districts in addition to the three CVD-based districts. The last row of the table gives the average Roeck ratio for each set of districts. As we examine this table, we can compare the Roeck ratios for different sets of districts to each other and to the Roeck ratio for a perfect

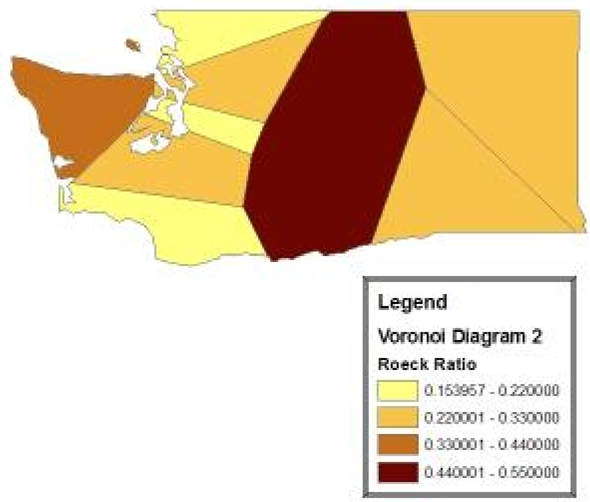

Figure 17: Roeck Ratio over Second CVD [6]

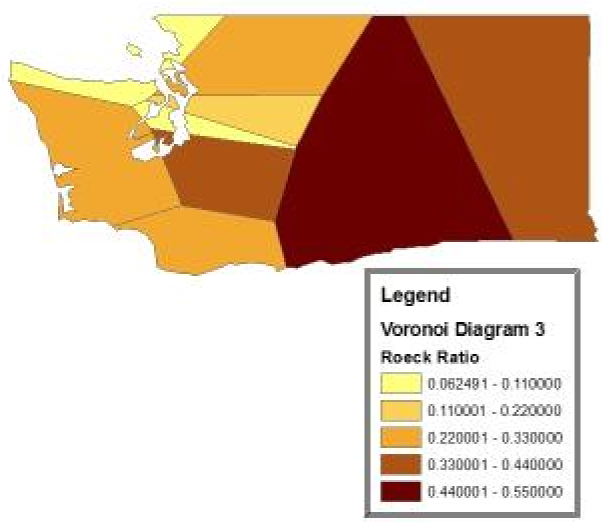

Figure 18: Roeck Ratio over Third CVD [6]

Current	First CVD	Second CVD	Third CVD
0.293089	0.414687	0.294886	0.407779
0.546939	0.494973	0.289332	0.484481
0.261801	0.296180	0.482301	0.263783
0.300806	0.173535	0.235988	0.131780
0.126157	0.331262	0.201280	0.062491
0.234736	0.255775	0.286379	0.330565
0.365203	0.087302	0.189124	0.251200
0.463537	0.256457	0.392799	0.305724
0.300019	0.159989	0.153957	0.104363
0.321365	0.274462	0.280672	0.260241

Table 4: District Roeck Ratios

square, $2/\pi = 0.6366$. The average Roeck ratios for the CVD-based districts are all lower than the current districts average. The ranges of each set are roughly similar. The Roeck ratios are also displayed for each CVD-based district set in Figures 16, 17 and 18.

In regards to compactness, these three CVD-based districts do not offer any improvement over the current districts. Any of the solutions mentioned previously (more iterations, a Monte Carlo method, etc.) may produce more districts with a higher Roeck ratio average and a smaller range of the Roeck ratio.

6.3 Geographic Concentration of Initial Points

To better judge the stability of this method, we may examine two other CVDs that both underwent 150 million iterations but each had different sets of initial points than the first three CVDs. Figure 19 displays a diagram that had an initial set of points equally spaced out along a line running through central Washington. This diagram only assigns the Puget Sound region one district, while giving roughly three districts to eastern Washington and keeping three to four districts within central Washington. Given that a large part of Washington's population is concentrated in the Puget Sound region and that eastern Washington is less densely populated, this diagram obviously is not an improvement upon the current districts. However, considering that it began with a set of initial points that were so intentionally concentrated, it is very plausible that after more iterations, this diagram would begin to resemble the first three more closely.

The diagram in Figure 20 was derived from an initial set of points equally spaced along a line running north-south through the city of Spokane in eastern Washington. This diagram assigns less districts to western Washington and more districts to eastern Washington than did the last diagram. When the ini-

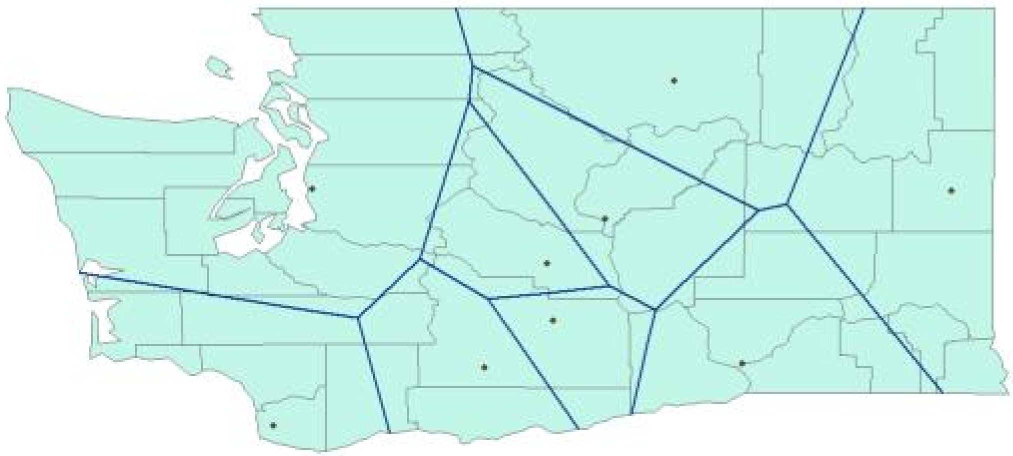

Figure 19: CVD Derived from Initial Points Located in Central Washington [6]

tial points, such as in the this case, are all placed in a low-density area furthest away from the highest-density area, it will take even more iterations to find the centroids because MacQueen's method has a very slow convergence rate. Since the Puget Sound has the largest collection of high population-density census tracts, when the first of those tracts is randomly selected it will move the nearest of the nine points closer to itself. The next Puget Sound census tract that is selected, will likely move the same point closer to itself. With each move, this point's counter, j_i, will increase by one each time and limit the point's movement for the next time it is selected. Our case of eastern Washington exaggerated this effect.

Given that the placing of the initial points can have such a large effect on the number of iterations needed to find the centroids, it is helpful to place the initial points strategically in urban areas as we did for the first three diagrams. A Monte Carlo method however, would find a set of intial points by a stochastic method based on population distribution. So while our method attempts to imitate the end result of the Monte Carlo method, it still lacks the randomized aspect.

6.4 Multiple Runs

Another way to correct for the slow convergence rate of MacQueen's method is to run the method on the initial set of points for a set number of times, then take the results and rerun the method on the results such that at the start of this second run, all of the counters j_i for $i = 1, 2, ..., 9$, will each be reset to one. Whereas all the previous diagrams have been drawn from centroids that

Figure 20: CVD Derived from Initial Points Located in Eastern Washington [6]

underwent 150 million iterations under one run of the method, this new diagram that we will call the fourth diagram, underwent a total of one billion iterations from six runs of our method. The number of iterations in millions in each run is: 300, 150, 150, 150, 150 and 100. At the end of each run, the resulting points were used as the initial points for the start of the next run. Now we will be able to judge the difference in compactness and population equality between the original three diagrams and one that has underwent six runs of the method for a total of a billion iterations. Figure 21 displays the districts of the fourth diagram which on first glance resemble the original three district, but a closer examination shows that the districts in western Washington are more developed so that they are more compact. Table 5 lists the 2006 population estimates provided by the Census Bureau for the current districts, alongside the approximations, derived from the 2004 population estimates for census tracts, for the CVD-based districts. The bottom row is the difference in population between the maximum and minimum. The fourth diagram, with a range of 375, 173, does not show any improvement over the current districts. In regards to the first three diagrams, its range is just above the smallest range of the three, even after roughly six times more iterations. However, Figure 22 and Table 6 have more positive findings. Figure 19 displays the Roeck ratios by district. There is an overall improvement in the Roeck ratios in this diagram from the original three diagrams that each only had one district in the most compact category (0.44-0.55) and only one or two in the next highest category (0.33-0.44). Figure 19 however, has two in the highest category, two or three in the next, two each in the next two categories and none in the lowest category (0-0.11). Table 6 lists the Roeck ratios and the average of the Roeck ratios for each district set. The fourth diagram has a Roeck ratio average that not only is higher than the first three diagrams, but also is higher than the current. These findings comfirm the earlier discussion of

Figure 21: CVD After Six Runs of the Modified MacQueen's Method [6]

Current	First CVD	Second CVD	Third CVD	Fourth CVD
654799	615634	539531	610259	606541
654984	684289	429944	511327	736856
654992	854434	400906	929671	872513
654851	787900	990066	718284	512892
654935	636725	501469	500615	824469
655068	686268	755819	854239	715028
655016	481501	735582	610492	860934
655029	570619	909689	863893	497340
654447	853830	906405	568304	546918
621	372329	589160	429056	375173

Table 5: Comparison of District Population Estimates

how centroidal Voronoi diagrams divide to equalize energy, not necessarily to equalize mass. Otherwise, we would expect greater improvement of the fourth diagram in regards to population equality over the first three.

7 Conclusion

Given the compact nature of CVDs, they present intriguing possibilities for redistricting. Though our modified MacQueen's method produced population centroids that created districts with substantial population inequality, there are several possible changes that could improve our method so that a CVD could at least serve as the starting map for a state legislature's redistricting committee. From that point, a committee could make adjustments to achieve population equality and any other goals or concerns of the particular state, such as respect-

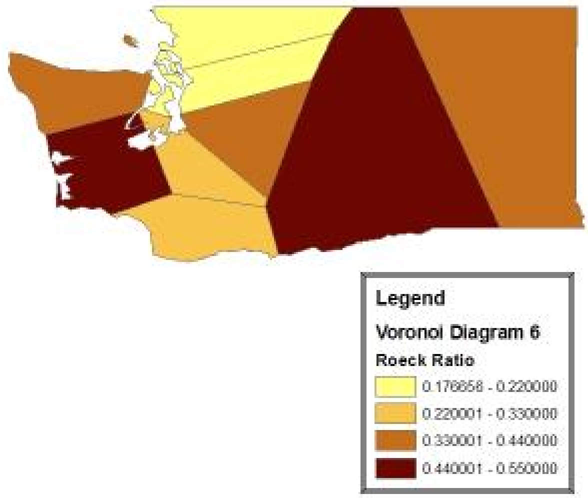

Figure 22: Roeck Ratios for CVD After Six Runs of the Modified MacQueen's Method [6]

Current	First CVD	Second CVD	Third CVD	Fourth CVD
0.293089	0.414687	0.294886	0.407779	0.356896
0.546939	0.494973	0.289332	0.484481	0.194746
0.261801	0.296180	0.482301	0.263783	0.176656
0.300806	0.173535	0.235988	0.131780	0.349716
0.126157	0.331262	0.201280	0.062491	0.226212
0.234736	0.255775	0.286379	0.330565	0.533446
0.365203	0.087302	0.189124	0.251200	0.255510
0.463537	0.256457	0.392799	0.305724	0.495053
0.300019	0.159989	0.153957	0.104363	0.415854
0.321365	0.274462	0.280672	0.260241	0.333788

Table 6: District Roeck Ratios

ing county lines, preserving communities of interest, etc. While these adjustments would detract from the compactness of the CVD, we would still expect the outcome to retain some of the compactness of the starting diagram. Some of the possible changes to our derivation of a CVD include running our modified MacQueen's method for more iterations, using a Monte Carlo method to find the initial points, using a continuous population density function, determining a measure of convergence within the modified MacQueen's method and putting a population constraint on the CVD. Also, an examination of the population estimation methods of the Census Bureau for congressional districts, counties and census tracts, would strengthen conclusions surrounding these data. Another possibility for further investigation is additional testing and analysis on the number of runs versus the number of iterations within our modified MacQueen's method so that the otherwise very slow convergence rate of MacQueen's method might be increased. Yet another possibility to consider is using Lloyd's method as it is described by Gunzburger, Du and Faber, Vance in their article "Centroidal Voronoi Tessellations: Applications and Algorithms" [3] in place of our modified MacQueen's method. With some of these suggested changes, CVDs could greatly contribute to the creating of unbiased congressional districts.

References

[1] Cain, Bruce E. *The Reapportionment Puzzle*. Berkeley and Los Angeles: University of California Press. 1984.

[2] Du, Qiang. "Qiang's Research Gallery" from Penn State University. http://www.math.psu.edu/qdu/Res/Pic/gallery3.html. 23 April 2007.

[3] Gunzburger, Max, Du, Qiang and Faber, Vance. "Centroidal Voronoi Tessellations: Applications and Algorithms." *SIAM Review* 21 (1999): 637-676.

[4] Hojati, Mehran and Bector, C.R. "Near-Optimal Political Redistricting." University of Saskatchewan and University of Manitoba. 2002.

[5] Ju, Lili, Du, Qiang and Gunzburger, Max. "Probabilistic Methods for Centroidal Voronoi Tessellations and their Parallel Implementations." *Parallel Computing*. 28 (2002): 1477-1500.

[6] Miller, Stacy. ArcGIS graphic. April 2007.

[7] Miller, Stacy. KolourPaint graphic. March 2007.

[8] Okabe, Atsuyuki, Barry Boots, Kokichi Sugihara, and Sung Nok Chiu. *Spatial Tessellations: Concepts and Applications of Voronoi Diarams, 2nd Ed.* Chichester, West Sussex, England: John Wiley & Sons, 1999.

[9] Schueller, Albert. "A Nearest Neighbor Sweep Circle Algorithm for Computing Discrete Voronoi Tessellations." Whitman College, Walla Walla, WA.

[10] U.S. Census Bureau. www.census.gov

[11] Williams, Justin C. "Political Redistricting: A Review." *Papers in Regional Science*. 74: 13-40.