Evolution of Gregorian Chant

Howard Liu, Michelle Lu

June 14, 2018

**Abstract:**

During the Middle Ages, the Roman Catholic Church had a huge influence on its people. The Church had locations widespread throughout Europe which still exist today in modern day. Gregorian chants were considered to be the sacred songs of the Roman Catholic Church.

Utilizing various Gregorian Chants, we intend to analyze each chant for its degree of variability both within the chant and its church locations. We hope to use this variability to find a correspondence between chant locations and the geographic distances of Europe. Looking at a two-dimensional plot of the chant variations, the locations and distances do seem to be related. Although, not perfectly matched, variations in chants seem to be a good estimate of distances of the actual locations on a modern-day map of Europe.

**Introduction:**

Gregorian Chants were developed between the 9th to 10th centuries in Western and Central Europe. Each chant is a sacred song of the Roman Catholic Church typically sung by choirs of men and boys. Because of this, Gregorian Chants are often referred to as Roman Chants as well. These chants were passed orally by these choirs to other areas of Europe. As the chants spread, they were modified to incorporate local style and to accommodate the changing practices of the Church. Around the 15th century, the old melodies went through a series of revisions that completed distorted the original essential qualities of the chants. This was due to the rise in popularity of organ music.

Eventually, these chants were written onto paper for further purpose in passing them down. This music was typically written in an early music notation called neume. This music notation later became the five-line staff which we now use in modern music.

With these modified chants on paper and in hand, our goal is to perform an analysis on these chants to understand how regional differences affect the variations of each chant. By doing so we may be able to plot out each chants’ locations similarly to a layout of modern Europe.

**Design:**

We used a sample of 12 different chants each with 10 different locations spread throughout Europe, including areas in France, Rome, Germany, and Italy. These locations per chant were not necessarily the same. Some of chants may have one specific location while others would not. These chants can be found in the book “The Emergence of Gregorian chant : a Comparative Study of Ambrosian, Roman, and Gregorian chant.” Each set of pages shows a single chant and its various notations per location in neume. The following are the chants we used: Lex Domini, A Summo Caelo, Deus in Adjutorium, Dicit Dominus, Victricem, Benedictus Dominus, Rorate, Factus Est, Dominus Dixit, Sicut Oculi, Benedicite Dominum, and Nomine Domini.

In order to analyze the distances between the chants, the chants had to be converted from music to data. To do so, we looked at the sheet music and translated the notes in neume to that of our musical notation today. The music reads the same as our modern notation, but the main difference is the clef and the 4-line format. Then, in order to convert musical notes into numerical values, we counted the number of half steps away from middle C and used that value. Even though Gregorian Chant music consists of very little ornaments, aside from the occasional B flat or B natural, we still counted the “black keys” in between. For chords, the music is sung from bottom to top and so the bottom note is recorded first.



Figure 1: Excerpt from Benedicite Dominum

From Figure 1 above, the first note would be translated as a G in modern music. G is 5 half steps down from C and thus this note would be recorded as -5. The entire measure would be translated as so: -5, -5, -3, -3, -3, 0, 0 , -1, -3. To make sure, the distances were more accurate, we also aligned similar phrases. All of these converted values were placed into a spreadsheet which became our datasets of twelve chants.

**Distance Methods:**

Inspired by studies involving the relationships between genetic differences and geographic locations in Europe, we used similar methods to conduct our analysis. All of our analyses were done in R.

The first step of our analysis was to take in the chant data and formulate a distance metric between each chant location. Five different distance metrics were used for comparison: Euclidean distance, Manhattan distance, Canberra distance, Levenshtein (edit) distance, and Mean Squared Error. Table 2 gives a description of the general concepts of each distance metric.

|  |  |
| --- | --- |
| **Distance Metric** | **Equation** |
| Mean Squared Error | Dmse = 1/n \* (∑(xi - yi)2) |
| Canberra | Dcanberra = ∑ (|xi - yi| / (|xi| + |yi|)) |
| Manhattan | Dmanhattan = ∑|xi - yi| |
| Euclidean | Deuclidean  = square root (∑(xi - yi)2) |
| Edit | Dedit = ∑(number of adjustments from string(xi) and string(yi)) |

Table 2: Distance Metric

In order to use the Edit distance metric, we had to convert our data into strings. To do so, we manually converted our numbers into our strings of choice. Table 3 is an example of our conversion sequence. By using the table shown, we converted each chant into a long string of letters. From there the edit distance is calculated by totalling the number of insertions, deletions, and substitutions needed to make the two strings identical.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Number** | 0 | -1 | -2 | -3 | -5 | -7 | -8 | -10 | -12 | 2 | 4 | 5 |
| **String** | “a” | “b” | “c” | “d” | “e” | “f” | “g” | “h” | “i” | “j” | “k” | “l” |

Table 3: Table used for Edit Distance

**Visualization Methods:**

In order to visualize the produced distance matrices, we originally created multiple heatmaps, one for each chant. This would allow us to get an idea of how the spread of variation looked between locations. However, the different heatmaps between different distance metrics were hard to compare to one another - harder to determine which one is “better”. Therefore, we continued on with other statistical methods.

We used two different dimensionality reduction techniques: Multi-Dimensional Scaling (MDS) and Kernel Principal Component Analysis (KPCA). MDS is a non-linear form of dimensionality reduction. Because each location has distances with all the other locations, we have n-location dimensions. MDS is a means of visualizing the levels of similarity of each distance between every location. Thus the n-dimensions can be plotted onto a two-dimensional plot where each object is an assigned coordinate. The benefit of using MDS is that it reduces these dimensions in a way that preserves the distances between these n-locations.

Similarly, we used KPCA to reduce our distance metrics from n-location dimensions to two dimensions. Regular Principal Component Analysis is often used on data to linearly separate it into a smaller number of dimensions. Kernel PCA allows us to do the exact same but the data is extracted into some non-linear feature space where the data is then further reduced into a smaller number of dimensions. This analysis is more useful in finding patterns and allows us to indicate potential relationships between locations.

After looking at the individual plots of distances, we determined that these plots did not give us a good visual idea of the locations. Originally, we worked on every chant individually and this may have had too much noise . Instead, we combined the chants for each distance metric before proceeding to the analysis to reduce noise. We combined all the chants by summing all the previous matrices by shared values. Thus if two locations had many shared chants, the final value in the cell of the combined matrix would be their sum. However, if two locations did not share a single chant together the cell would be left empty or be “NA”. To keep the scaling of the matrix consistent, we took both the mean and median (separately).

To further clean the dataset, we wanted to have no missing values in the large distance matrix. This means we only kept locations where they all share a chant(s) with one another. If one location was known to share missing values with multiple other locations, that one would be removed. For our project, we ended with only 11 locations. This is the case because we wanted to make sure that each location appeared in a chant together at least once with the 10 other locations.

For clarity on how we eliminated some locations for our final dataset, look at the table below (values in the table represent distances) :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Location A** | **Location B** | **Location C** | **Location D** |
| **Location A** | 0 | 2 | 3 | 3 |
| **Location B** | 2 | 0 | NA | NA |
| **Location C** | 3 | NA | 0 | 4 |
| **Location D** | 3 | NA | 4 | 0 |

Table 4: Location Reduction Example

In the case of the table above, we would utilize the locations of A, C, and D because A, C, and D all share a complete set of distances. We could also remove C and D, allowing us to retain A and B but we want to maximize the number of locations being used.

In order to compare our produced coordinate plots with that of a geographical map of Europe, we conducted Procrustes Analysis. Procrustes takes in two inputs, the data that we have and the data of something we are trying to match up to. It then finds the best transformation of our data through scaling, translating, and rotating to minimize the sum of square error between the data after the transformation and the data we are trying to match up to. For our project, we had to find the longitude and latitude of the actual locations of the chants in Europe. We then ran Procrustes Analysis on the actual locations and our coordinates from KPCA and MDS for each distance metric on mean and median. After comparison, we found that the MSE distance fit the data the best.

**Results:**

Below are our figures for predicted vs. actual and the actual locations on a geographical map of Europe.

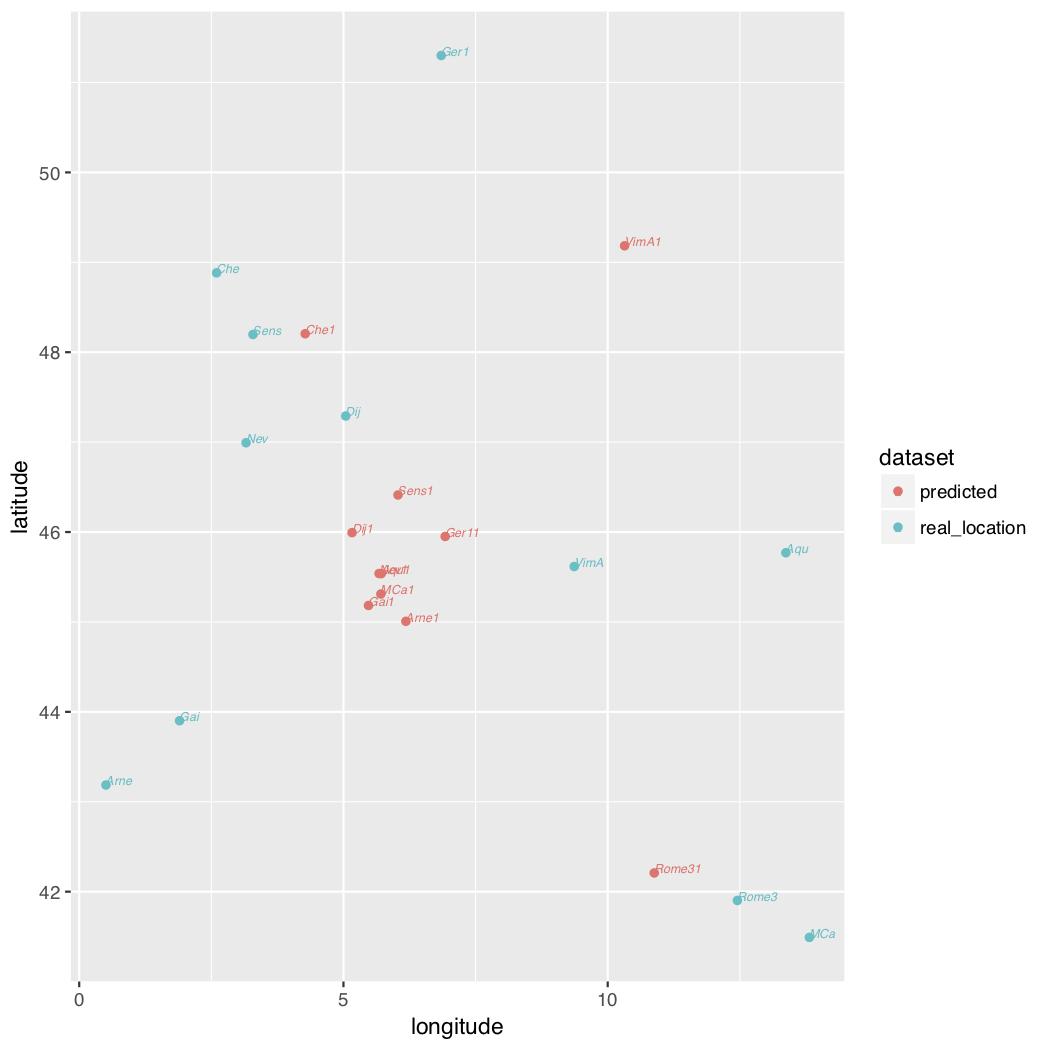
****

Figure 2: Plot of the MSE Distance after performing Procrustes Analysis

****

Figure 3: Geographical map of Europe with actual locations plotted

Our resulting figures bears some resemblance to the geographical map of Europe, however, not perfectly. The fit was not as great as we predicted. In some cases our points are approximately close, however there are some points that are far spread apart between the predicted and the actual. Unfortunately there are also some points that just fail to be close to their actual locations.

There could have been a variety of reasons for large discrepancies between points. If given more time, we may want to make some adjustments to fix this. Aspects of the experiment to consider in the future may include coding the data differently, utilizing other distance metrics, and working with a larger sample size. We began our experiment by coding our data with the number of half steps away from C as described previously. However, we think that there may have been a better way to transcribe the data. We did not experiment with other approaches, but considering that there is only one accidental in Gregorian chant, B flat, there may be a better way to code the data. The distance metrics we used were pretty similar and we think that there may be other distance metrics we did not consider that would more accurately demonstrate the differences between the chants. Our sample size of 12 is not ideal. If we had used all 40 chants provided in the book we used, there may have been better results.

**References:**

Werf, Hendrik van der. *The Emergence of Gregorian Chant a Comparative Study of Ambrosian,*

*Roman, and Gregorian Chant*. Hendrik Van Der Werf, 1983.