

Functional Data Analysis in Cephalometric Tracing and Mandibular Examination

Emily M. Slade

University of Michigan, Ann Arbor

Table of Contents

| | |
|---|----|
| Acknowledgements..... | 3 |
| Chapter 1: Functional Data Analysis Overview..... | 4 |
| Chapter 2: Mandibular Analysis and Cephalometric Tracing..... | 8 |
| Chapter 3: Shape Analysis and Principal Component Analysis Methods..... | 12 |
| Chapter 4: Data and Hypothesis..... | 15 |
| Chapter 5: Landmark Analysis of Mandibles..... | 19 |
| Chapter 6: Functional Data Analysis of Mandibles..... | 24 |
| Chapter 7: Conclusions and Recommendation..... | 29 |
| References..... | 32 |
| Appendix A..... | 35 |
| Appendix B..... | 36 |

Acknowledgements

This study and analysis would not have been possible without the direction of a few incredibly knowledgeable and helpful individuals. First and foremost, the author would like to thank Dr. Edward Rothman for his supervision of this project and endless ideas for the application of functional data analysis. Dr. Geoffrey Gerstner has also given invaluable guidance through providing the data for this project as well as answering numerous questions about the basics and methods of cephalometry. Furthermore, Dr. Elizabeth Crane has been extremely helpful in teaching the basics of functional data analysis in R and providing some R code for the creation of functional data objects. Without all of these people, the pieces of this project could not have come together, and thus, an incredible amount of gratitude is owed to each of them.

Chapter 1: Functional Data Analysis Overview

Functional data analysis is a way to use the innately smooth nature of functions to gather and analyze information stored in these curves' derivatives. The methods involved in functional data analysis are especially useful in determining the location on and characteristics of a curve that differ the most from other generations of the same measure. Functional data analysis is a relatively new idea in the world of statistics with the first textbook published in just 1997. Since then, the idea has really caught on with applications extending to medicine, biostatistics, econometrics, and more (Gonzalez-Manteiga & Vieu, 2007). Providing a way to smooth and analyze even very complicated data sets makes it an extremely versatile tool in essentially unlimited areas of application.

One main use of functional data analysis is to interpolate the notion of "data points" to form an entire curve (Ramsay & Silverman, 2005). A user can control the smoothness of the created curve by specifying the number of smooth derivatives needed in their analysis (Ramsay & Silverman, 2005). Fitting a functional data object to maintain smoothness involves first establishing a basis set of functions, then creating a matrix of coefficients that represent forming a linear combination of these basis functions to fit the data (Ramsay, Hooker, & Graves, 2009). One could theoretically employ more and more basis functions to find a perfect fit. Thus, a user must specify the number of smooth derivatives needed in order to ensure that the generated curve is smooth enough for the necessary analysis (Ramsay & Silverman, 2005).

These basis functions are typically one of two types: spline or Fourier. The spline basis first defines a number of break points over an interval. Between break points, a polynomial of a predetermined degree is fit to the data. At each break point, the two polynomials on either side are smoothed together through a process requiring the user to set the number of matching

derivates for the two polynomials. The most popular spline basis is the B-spline basis system. In the B-spline system, the values of all of the spline functions at any given point in time sum to one at that point. The Fourier basis is made up of functions $\{1, \sin(\omega t), \cos(\omega t), \sin(2\omega t), \cos(2\omega t), \dots\}$ where $\omega = 2\pi/T$ with T representing the period. Due to the innate cyclical nature of sine and cosine functions, the Fourier basis is typically used for data that is periodic in nature. Since it is common for shape analysis to deal with closed-curve contours, Fourier bases typically underlie the functional data analysis. Figure 1 below shows four Fourier basis functions for simplicity, but many analyses involve more complex bases made up of dozens of basis functions. (Ramsay et al., 2009).

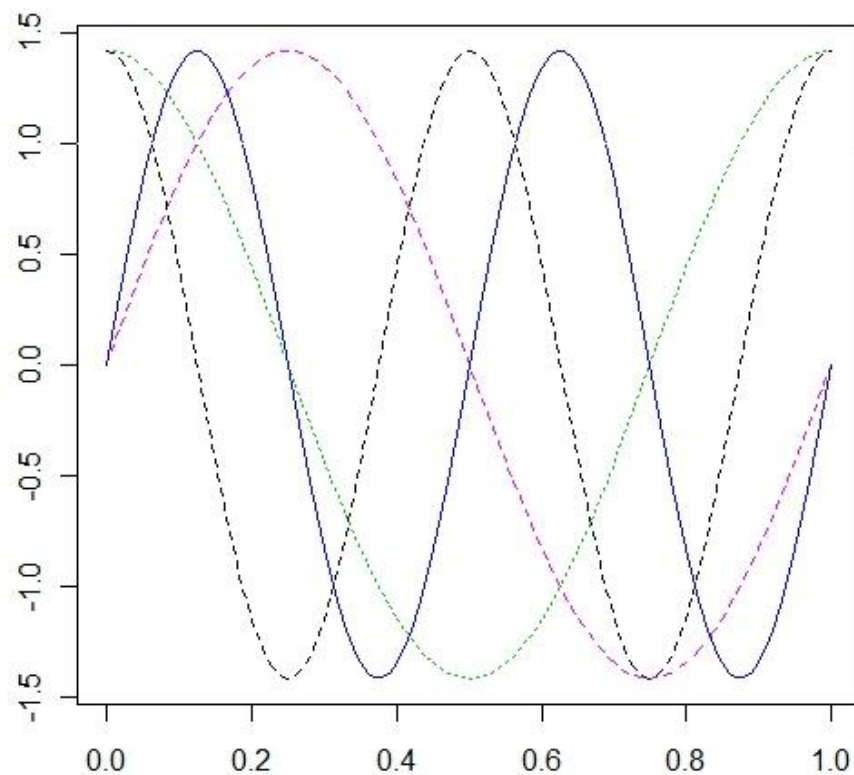


Figure 1: A visual representation of the first four Fourier functions.

Ramsay and Silverman (2005) have also developed methods for analyzing the data once it has been expressed in functional form. First off, registration is an important tool for lining up the functional objects to reduce unwanted variation in the data (Ramsay et al., 2009). Landmark registration uses a smooth, monotonic but non-linear function to warp the time axis in order to line up important features or landmarks across all functions (Ramsay et al., 2009). For an example to illustrate the need for registration, suppose we are interested in studying the pubertal growth spurt in girls. This spurt will happen at slightly different ages in different girls, so landmark registration will line up each function at the spurt so that a mean curve will be an accurate representation of the mean spurt over all girls (Ramsay & Li, 1998). After registration of curves, a common analysis to carry out on the functional data involves functional principal component analysis (Ramsay et al., 2009), but more about that method will be discussed in Chapter 3.

Likely due to its newness, functional data analysis has rarely been used in multiple dimensions or in image analysis applications (Epifanio & Ventura-Campos, 2011). One of the first two-dimensional analyses involved repeated samples of handwriting, with each sample being treated as a function (Ramsay, 2000). The location of the pen throughout the sample provides two-dimensional data in the form of 'X' and 'Y' coordinates, but a major part of this analysis actually separated the 'X' and 'Y' coordinates of the handwriting curves to be analyzed separately, reducing the dimensionality (Ramsay, 2000). Another early image analysis using functional data analysis occurred in 2008 when it was used to distinguish between cancerous and benign cells by treating the closed-curve contour of the cell as a function of arc length for functional data analysis (Nettel-Aguirre, 2008).

Since then, the idea of functional data analysis in image and shape analysis has expanded into even more applications including the analysis of contours of femur bones as well as hippocampal differences in patients with Alzheimer's disease (Epifanio & Ventura-Campos, 2011). Despite the very different nature of these studies, all of them are able to distinguish between two different groups through fitting functional data objects to the data, isolating the important information through principal component analysis, and using basic t-tests to identify the important principal component differences between the groups (Epifiano & Ventura-Campos, 2011). Ramsay and Silverman (2005) have developed methods that combine principal component analysis with functional data analysis, even extending this into multivariate principal component analysis, which they demonstrate through analyzing the aforementioned handwriting data with 'X' and 'Y' considered together as a multivariate function.

Chapter 2: Mandibular Analysis and Cephalometric Tracing

An area in which to consider applying functional data analysis is mandibular tracing and analysis. In orthodontics, doctors use facial profile radiographs, also known as lateral cephalograms, to view the shape and position of a patient's mandible. With this, orthodontists evaluate deformities or malocclusion as well as predict changes that will occur through orthodontic treatment (Boeck, Kuramae, Lunardi, Santos-Pinto, & Mazzonetto, 2010). Abnormalities such as mandibular retrognathism or prognathism (more commonly known as an overbite or underbite) can cause functional limitations such as inefficient chewing as well as esthetic insecurities for patients (Boeck et al., 2010). Numerous studies have shown that discrepancies between the position of the maxilla and mandible frequently lead to psychological issues such as poor self-image and possibly even job-related discrimination for patients (Hickey & Vergo, 1991). For these reasons, adjustments such as mandibular elongation and other surgical methods have been studied and performed for decades to correct these maxillary and mandibular abnormalities (Molina & Monasterio, 1995).

In order to make a cephalogram useful for an orthodontic analysis, tracing must first be done to identify important structures of the mouth such as the mandible. One way to do this tracing is to simply identify the contour by eye and trace it by hand (Little). It has been found that the mandible grows in a natural logarithmic spiral fashion, and methods have been developed that use predetermined sines and cosines to calculate these changes that will occur through surgery (Scolozzi, Link, & Schendel, 2007). Such methods use just two angles and one length measured by hand on a lateral cephalogram for the analysis (Scolozzi et al., 2007; Piroozmand, 2001). These measures are simply lengths and angles, not taking into account the

actual contour of the mandible which can be considered a closed curve (McNamara, Bookstein, & Shaughnessy, 1985).

Studies show that positioning can be improved esthetically with changes in SNB, BNP, and PgNP, with a larger ANB leading to a greater mandibular advancement after surgery (Boeck et al., 2010). S refers to the sella, N is where the nasal and frontal bones meet along the curve of the bridge of the nose, and Pg is the lowest point on the chin (Little). Figure 2 shows the location of these important landmarks and lines. A and B are both arbitrary points, with A being the innermost point on the curve from the maxillary anterior nasal spine to the crest of the maxillary alveolar process and B being the innermost point on the curve from the chin to the alveolar junction (Little). As shown in Figure 2, all of the aforementioned angles and lines are

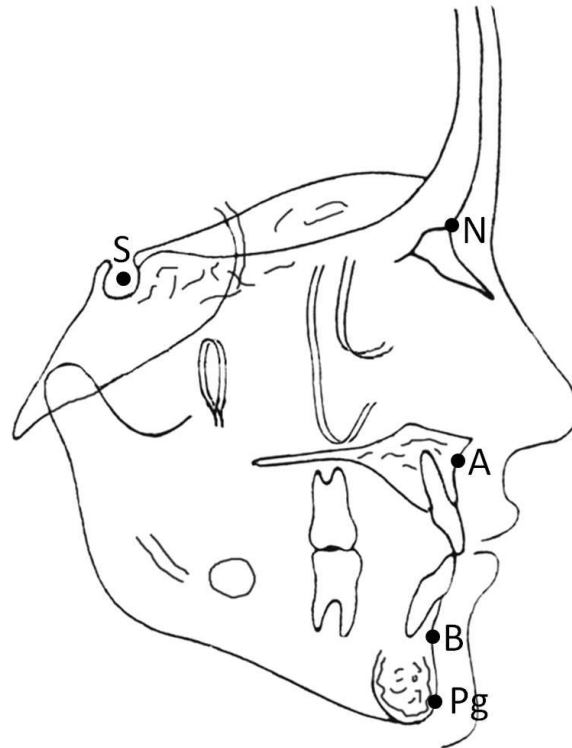


Figure 2: The location of the sella (S), nasion (N), pogonion (Pg), and arbitrary points A and B on a cephalogram tracing.

formed by connecting these landmark points (Little). ANB is commonly used when analyzing the length of the mandible because the line between A and N (maxilla and nose) is stable, but the position of B will change with changes in the position of the mandible. Thus, the angle ANB is a good way to capture information about the position of the patient's mandible with respect to the position of the maxilla.

A common program with which to digitize these cephalometric landmarks is called tpsDig, and this software can be downloaded online for free from Stony Brook University (Zelditch, Swiderski, & Sheets., 2012). To use this software for landmark image analysis, the user must click on each landmark by hand in the exact same order for each image in the analysis. This task could potentially be very tedious with many landmarks, many images, or both. If even one single landmark is accidentally not selected or selected in the wrong order, the digitized data is unusable for analysis.

Selecting landmarks by hand or tracing structures by hand has recently been updated through various computer programs. It can be difficult even for the trained eye to identify the necessary landmarks and structures on a cephalometric film radiograph, but Dolphin Imaging, a popular software used by orthodontists, can turn these radiographs into contours that are easy to read. A study using sixty-four cephalometric radiographs confirms that there is no difference between the superimpositions produced by Dolphin Imaging as compared to the hand-drawn standard (Huja, Grubaugh, Rummel, Fields, & Beck, 2009).

In a literature review concerning mandibular elongation, the authors find that in addition to surgical methods, functional appliances can also produce significant lengthening of the mandible (Cozza, Baccetti, Franchi, Toffol, & McNamara, 2006). One main caveat of this review is that the actual length of the mandible is not enough to holistically assess its length and

position in the mouth (Cozza et al., 2006). It is suggested that measures such as ANB as well as positioning of the teeth should also be considered. For this reason, more sophisticated methods should be developed that consider many of these measures together. All in all, cephalometric tracing and choice of key landmarks are important topics in cephalometrics that should be further studied, particularly for examination of the mandible.

Chapter 3: Shape Analysis and Principal Component Analysis Methods

The predetermined anatomical loci of a mandible can be thought of as landmarks in traditional shape analysis. A widely-accepted definition of shape is “all the geometrical information that remains when location, scale and rotational effects are filtered out from an object,” (Kendall, 1977; Zelditch et al., 2012). One way to think of shape is to consider these landmarks as a web or map of an image, providing more information by interpolating “semi-landmarks” between the anatomical landmarks (Zelditch et al., 2012).

One of the most basic ways to analyze shape is to use Bookstein shape coordinates which involves first identifying three landmarks on an image then rotating and scaling the triangle so that one landmark falls at (0,0) and another at (1,0) on a Cartesian coordinate grid (Zelditch et al., 2012). When multiple images have been prepared this way, two of the three coordinates of each image are standardized so that location differences in the third coordinate will reveal differences in shape between the two images (Zelditch et al., 2012). When there are more than three landmarks, multiple triangles can be analyzed to determine differences in shape (Zelditch et al., 2012).

Bookstein updated the field of cephalometric analysis using this triangle idea through the development of a method called the tensor method, using the sella (S), nasion (N), and anterior nasal spine (A) as the landmarks for the triangle (McNamara et al., 1985). In this study, these triangles were used to examine changes in subjects’ mandibular shape and growth using a Frankel appliance as compared to changes in mandibular length without the appliance, finding that the appliance does lead to significant lengthening of the mandible (McNamara et al., 1985).

Although less intuitive than the Bookstein shape coordinates, a more widely-used method of shape analysis is the Procrustes superimposition method (Zelditch et al., 2012). This process

can be summed up in three steps. First, center the landmarks of an image by subtracting the coordinates of the centroid (arithmetic averages of the landmarks' 'X' and 'Y' coordinates) from the coordinates of each landmark (Rohlf & Slice, 1990). Then, scale the shape by dividing the coordinates of each landmark by the centroid size (Rohlf & Slice, 1990). Since the shapes are centered, the centroid size is simply the square root of the sum of the squared coordinates (Rohlf & Slice, 1990). The last step is to rotate the shape to minimize the sum of the squared differences of the coordinates (Rohlf & Slice, 1990). This "registers" the images by reducing variation due to positioning as well as standardizing for size. When multiple images are lined-up in this way, each landmark can be averaged over the images to produce one mean image, and variation in one image can be assessed by comparing that image to the computed mean image (Rohlf & Slice, 1990).

Using these Procrustes mean shapes for mandibles, various analyses have been carried out in the field of cephalometry. One example uses Euclidean distance matrix analysis (EDMA) to differentiate between different classes of malocclusion (Singh, McNamara, & Lozanoff, 1998). Other analyses of Procrustes mean shapes use standard principal component analysis or basic comparison of important angles (e.g. ANB) when those shapes are mandibles (Epifiano & Ventura-Campos, 2011; McNamara et al., 1985).

Principal component analysis is a valuable tool in the field of shape analysis because it is designed to reduce the dimensionality of high-dimensional data, which shape analysis typically involves. Principal component analysis will break down the information into orthogonal, uncorrelated variables, and the amount of variation in the data set coming from each of these components is given (Jolliffe, 2004). In landmark analysis, when the first principal component explains a large percentage of the total variation, it can be concluded that there is a high

dependence between certain landmarks (Epifiano & Ventura-Campos, 2011). Depending on the type of data and desired analysis, principal component analysis can be used as a means for factor analysis, regression analysis, outlier detection, and more (Jolliffe, 2004).

Epifiano and Ventura-Campos (2011) offer an interesting take by combining functional data analysis, shape analysis, and principal component analysis. Using a sample of femura, the closed-curve outline of each bone is treated as a function parametrized by arc length (Epifiano & Ventura-Campos, 2011). With this data, principal component analysis is used in both landmark description and functional data analysis as a way to differentiate between arthritic and normal bones (Epifiano & Ventura-Campos, 2011). They find that functional principal component analysis is even better than landmark analysis because it produces fewer errors in their bone characterization (Epifiano & Ventura-Campos, 2011). This study provides an optimistic outlook for applications in other areas of shape analysis by demonstrating an elegant way to use functional data analysis for closed-curve shapes.

Chapter 4: Data and Hypothesis

This femoral analysis by Epifiano and Ventura-Campos provides the inspiration for the connection between functional data analysis and cephalometry. Like the femur, a lateral radiograph provides a mandibular contour that can be treated as a closed-curve function for the sake of functional data analysis. It is hypothesized that functional data analysis can be a tool that gives orthodontists all of the necessary and valuable information to examine and treat mandibular abnormalities using the traced contour. Further, the goal is that functional data analysis can provide more information about the mandible than the traditional landmark methods used in orthodontics today. Additionally, a validation of functional data analysis in cephalometrics could solve some current problems with the landmark approach including the inefficient and error-prone identification of mandibular landmarks by hand. Overall, a similar analysis to the one presented in Epifiano and Ventura-Campos' study (2011) will compare landmark analysis to functional data analysis using mandible contours.

Data for this analysis were collected by researchers at the University of Michigan in 2007 for a study concerning the relation between cephalometric structures and the facial esthetics during a smile. The data contain a wealth of length and angular measurements used by orthodontists in assessing these structures, including those presented in Figure 2 on page 9. These cephalometric measurements were produced electronically using the Dolphin Digital Imaging System, which has been proven to be comparable to the hand-drawn standard, as previously mentioned. These measurements were made from an image containing tracings of the important dental landmarks and contours, generated from a lateral radiograph from the right side of the face. An example of one such tracing from this data set in its entirety is provided in Figure 3 on the following page.

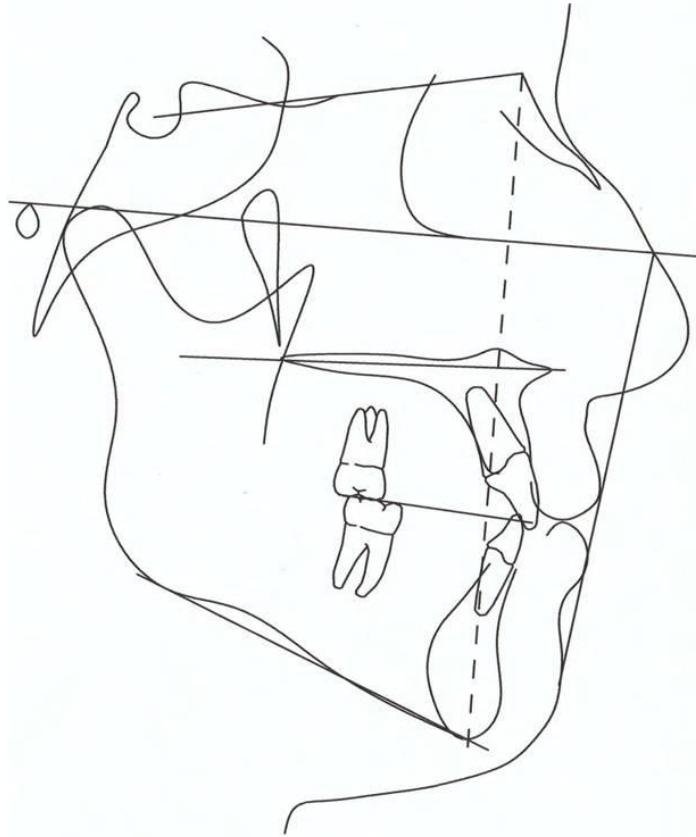


Figure 3: One complete tracing from the data set before any transformations have been applied.

The full data set consists of cephalometric tracings from 24 subjects. These 24 subjects are all healthy female dental students from the University of Michigan in Ann Arbor. Their ages range from 22 to 28 with a mean of 24.3 years old. The average height and weight are 1.7 meters and 58.2 kilograms, respectively. The data was collected for the purpose of examining smiles, but the presence of mandible tracings makes it an excellent data set for the purposes of this project.

To prepare the data for this analysis, the first step was to digitize the images using a standard office scanner. Next, the digitized images were uploaded into Adobe Photoshop software. Using the “ruler” tool in this software, the images were rotated to align to a standard

horizontal line in each image. This line, commonly called the Frankfurt horizontal, is a standard reference in cephalometrics and was created to be as close as possible to parallel to the ground when a human holds their head naturally (Piroozmand, 2001). In Figure 3 on page 16, this is the nearly-horizontal line that passes from the far left side of the image to the far right side of the image. Next, the “polygonal lasso” tool in Adobe Photoshop was used to isolate the tracing of just the mandibular contour, eliminating all other unnecessary lines from the image. The tracing that is left represents a closed planar curve of the boundary of the mandible, and all 24 tracings in this data set are visible in Appendix A.

For landmark analysis, each cephalometric image was uploaded in tpsDig, and 10 landmarks were selected along the mandible by hand for each image. These landmarks were chosen by the fact that they correspond to the anatomical curves of a mandible. Figure 4 on page 18 is an example mandible that shows the location of these 10 landmarks. The coordinates for each landmark were saved through tpsDig for analysis. For functional data analysis, a free raster-vector software called WinTopo Freeware, available at wintopo.com, was used to translate the mandibular contour into over 1000 Cartesian coordinates for use in R. As assessed by eye and visible in Figure 5 on page 18, this process introduces no substantial variation in the analysis. This image shows a magnified section of the mandible to display that the software fits a vector so close to the contour that the rasterized contour is visible on both sides of the vector, meaning that we can assume this software is accurately vectorizing each mandibular contour.

The primary analyses for this study were carried out using R, a free statistical software available at r-project.org. Ramsay and Silverman (2005) provide a number of functions accessible through the ‘fda’ library in R that are necessary for this analysis. The details of each

R analysis will be presented separately for landmark and functional analyses in Chapters 5 and 6, respectively. All R code used in this analysis is available in Appendix B.

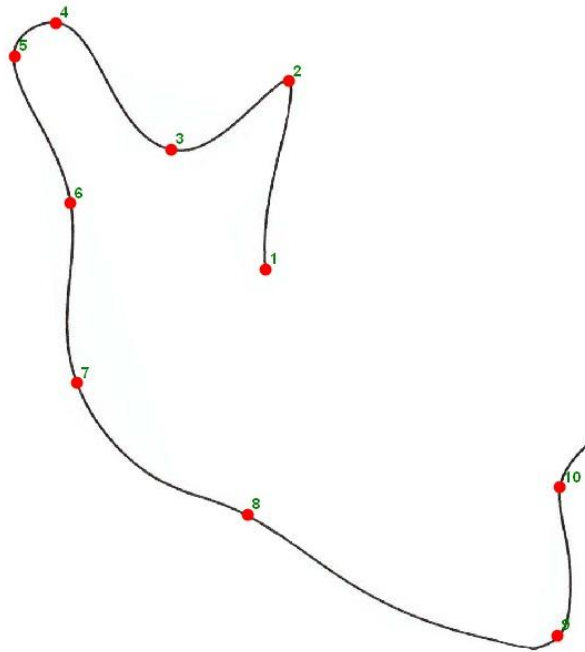


Figure 4: One isolated mandibular tracing from the data set demonstrating the position of the 10 landmarks selected for this analysis.

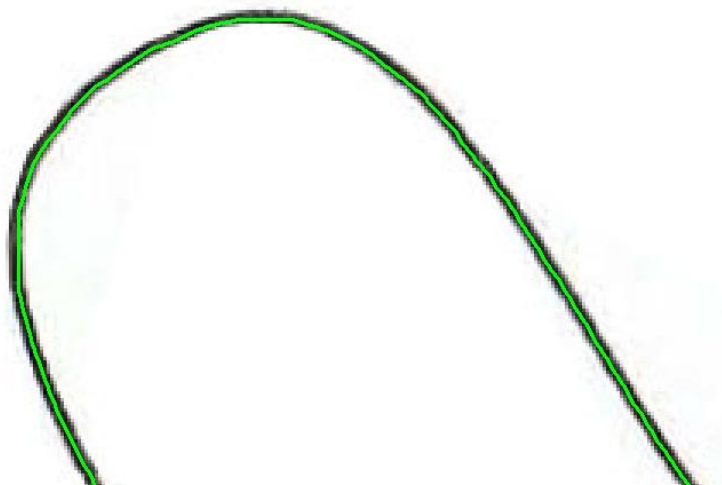


Figure 5: A magnified look at the calculated vector (green) as compared to the actual contour (black) of a mandible.

Chapter 5: Landmark Analysis of Mandibles

The current standard for mandibular analysis makes assessments using the anatomical landmarks present on a lateral cephalogram. As aforementioned, the angle ANB is used in examining the positioning and length of the mandible. Our data set includes the ANB for each subject, ranging from -2.0° to 6.5° with a mean of 2.52° . Of the 24 subjects, two subjects have an ANB less than 0° which is typically classified as an underbite, and six subjects have an ANB greater than 4° which is typically classified as an overbite. However, the boxplot of ANB in Figure 6 shows that there are no outliers in this data set. This finding is not surprising because all of the subjects in the study are considered healthy, normal individuals.

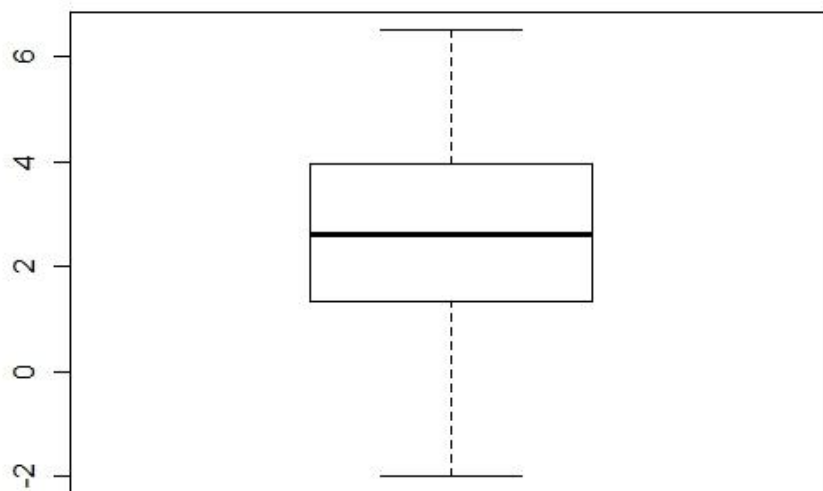


Figure 6: A boxplot of ANB angle measures for all subjects shows that there are no outliers in the set.

This data only tells us about the length of the mandible, not necessarily the shape of the mandible. To assess the shape of the mandibles in our data set using landmark analysis, the same procedure was followed that was carried out in Epifiano and Ventura-Campos' study (2011) on femur contours. First, the shapes are rotated and scaled using the Procrustes method

through the 'shapes' library in R. Next, a generalized Procrustes analysis is performed using the 'procGPA' function in R. This first uses the landmarks to register each image so that only the necessary shape data are present. The generalized Procrustes analysis also calculates a mean shape based on these 24 standardized contours. The Procrustes registration is shown for all 24 mandibles in Figure 7, and the Procrustes mean shape with labeled landmarks is shown in Figure 8 on page 21. A noticeable effect of the Procrustes registration is that the mandibles are rotated approximately 90° clockwise.

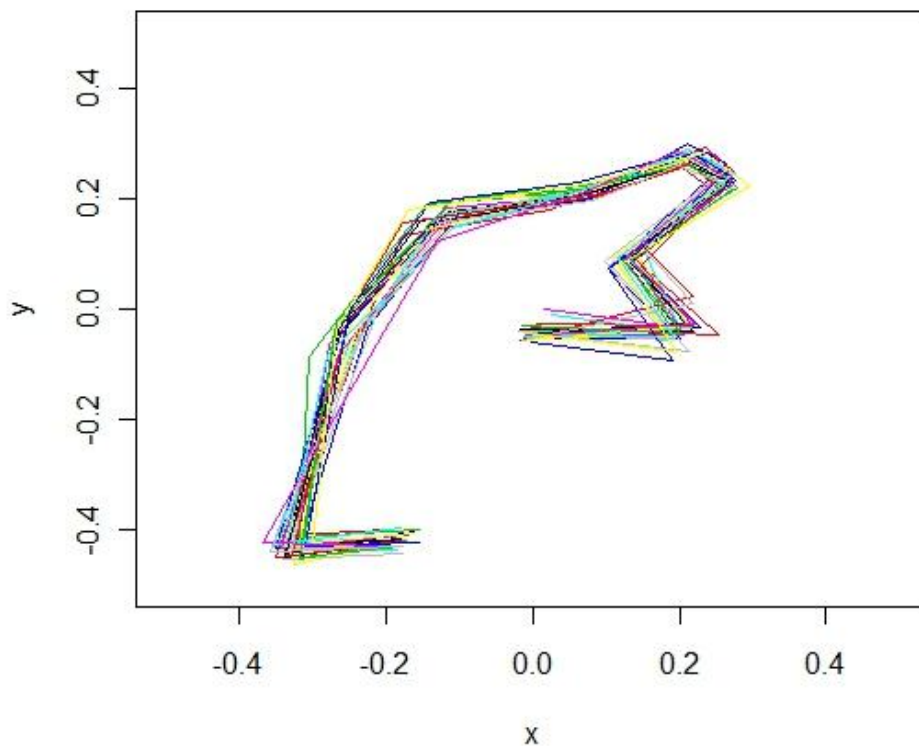


Figure 8: The tracings of all 24 mandibles after landmark selection and Procrustes rotation and scaling.

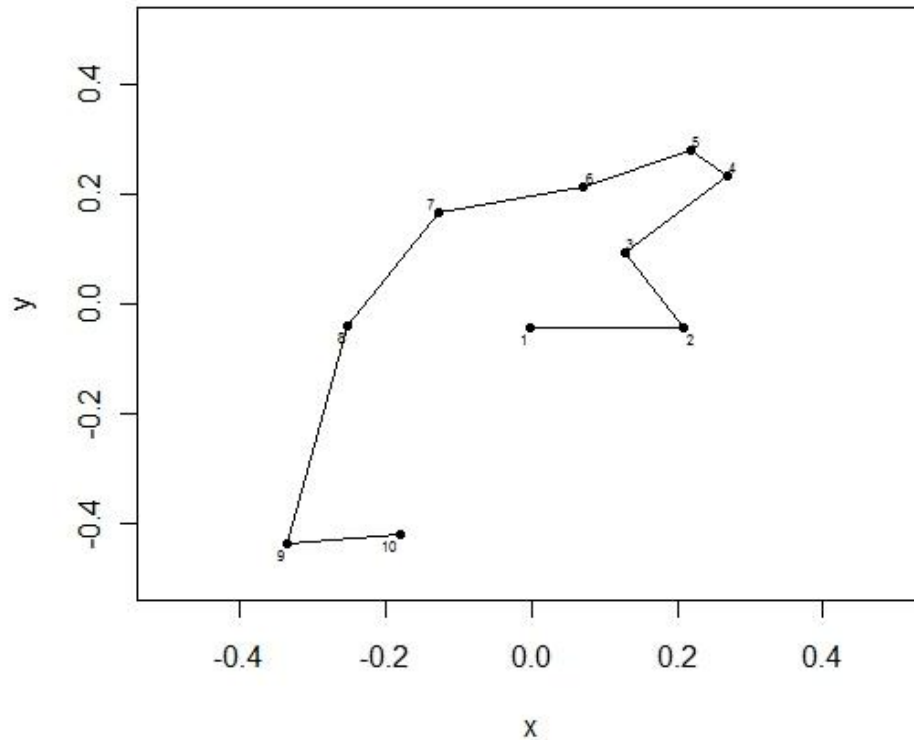


Figure 9: The Procrustes mean mandible for the data set with labeled landmarks.

The generalized Procrustes analysis in R also performs a principal component analysis. Figure 10 provides a visualization of the first four principal components which account for 29.6%, 21.0%, 15.4%, and 12.7% of the variation in the data, respectively, for a total of 78.7% of variation. Each image shows the coordinates of the mean mandible with vectors along each respective principal component (Epifiano & Ventura-Campos, 2011). The fact that the variation is very spread out amongst many principal components indicates that there is not a high dependence between landmarks (Epifiano & Ventura-Campos, 2011). This may be the reason that these principal components in Figure 10 appear to be complex even with this visualization.

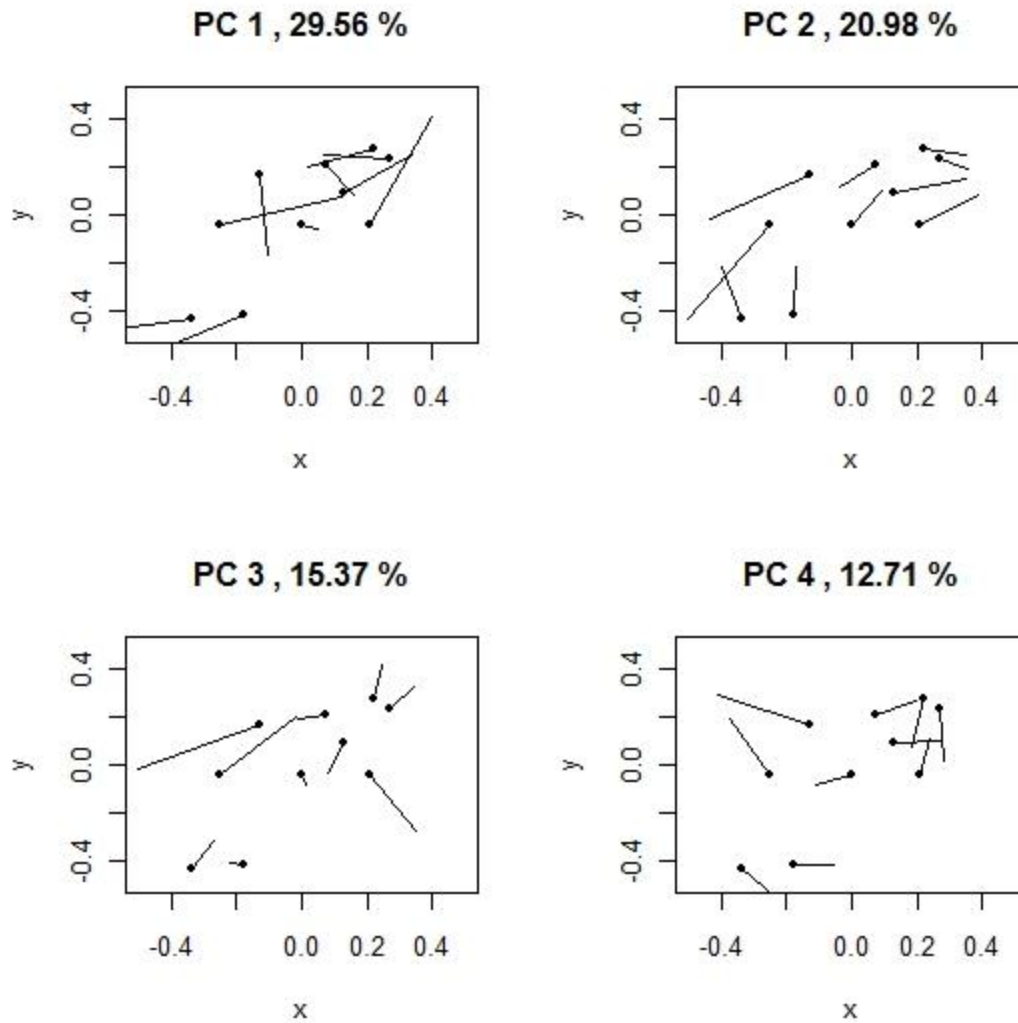


Figure 10: Each image shows the coordinates of the Procrustes mean mandible with vectors in the direction of the respective principal component for the first four principal components.

In this type of analysis, the shape variability in the data is assessed through the root mean square of the full Procrustes distance (Dryden & Mardia, 1998). In our data, this root mean square is 0.07 which indicates very little shape variability (Dryden & Mardia, 1998). The Procrustes distance for each shape is a way to indicate how close that shape is to the Procrustes mean shape (Dryden & Mardia, 1998). The boxplot of Procrustes distances for all mandibles in

Figure 11 shows that there are no outliers in this set, meaning that none of the mandibles' landmarks are significantly different than those of the mean mandible. Again, this is not surprising because the data comes from normal, healthy individuals.

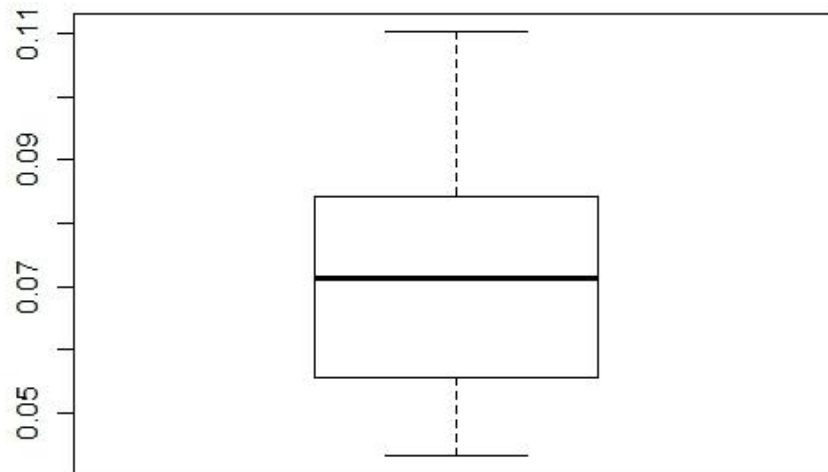


Figure 11: A boxplot of all Procrustes mean differences, indicating that there are no outliers in the set.

Chapter 6: Functional Data Analysis of Mandibles

In landmark analysis, the information in shape of the mandible is reduced to just 10 landmarks, but functional data analysis retains much more information. After using the raster-vector software, each mandible contour is made up of over 1000 coordinates, providing information about every millimeter of the curve. Like in landmark analysis, it is necessary for each mandible to be represented by the same number of points. However, because the data points are so dense in this method, we do not lose very much information by eliminating every few points along the curve as necessary to force each coordinate set to be the same size.

Once all tracings have the same number of coordinates, functional data objects can be created. Because we are treating this as closed-curve data, the cyclical Fourier basis is used to define our functional objects. Selecting too many basis functions will result in introducing extra noise into the data, but selecting too few could over-smooth (over-simplify) the data. It is reasonable to detect both extremes of this problem by eye through plotting the data, and 45 Fourier basis functions have been chosen for the purposes of this analysis. Once this basis is defined through the ‘create.fourier.basis’ function in R, the ‘smooth.basis’ function can be used to create a linear combination of the bases for each replicate in the data set (Ramsay et al., 2009). For multivariate data like the ‘X’ and ‘Y’ data in this analysis, these functions have been built to consider the variables together, providing a better analysis than separating the data (Ramsay et al., 2009; Ramsay, 2000). This is possible by considering the position of the curve in space at equidistant points along the arc length of the curve.

Like the Procrustes method in landmark analysis, it is necessary to line up these functional data objects as well. Landmark registration and continuous registration are both processes developed for functional data analysis to solve this problem. Landmark registration

involves specifying predetermined peaks, zeroes, or other defining features, and the time axis of each curve will be warped to meet those landmarks at the same point in time (Ramsay & Silverman, 2005). Continuous registration prevents the need for this tedious landmark selection process by considering the whole curve and warping each curve to match a specified target curve (Ramsay & Silverman, 2005). Because we have data on the whole curve, the ‘register.fd’ function was used to continuously register our 24 curves. Even though the multivariate data is being considered together, the ‘X’ and ‘Y’ registered curves are shown separately in Figures 12 and 13 on page 26, respectively, for simplicity.

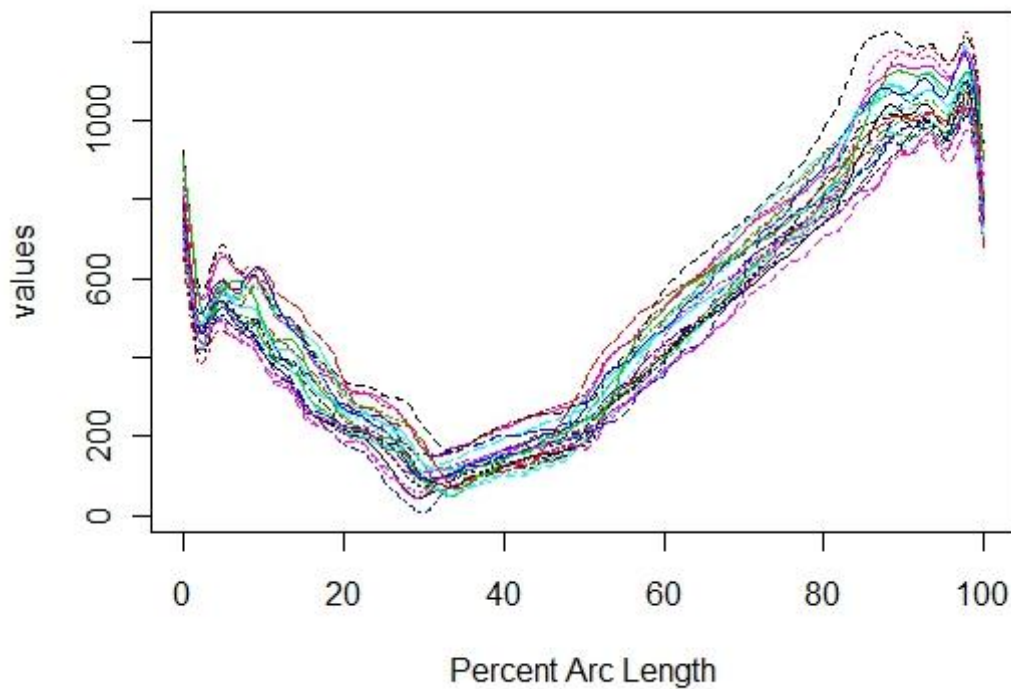


Figure 12: The ‘X’ values for all mandibles as a function of arc length after registration has been performed.

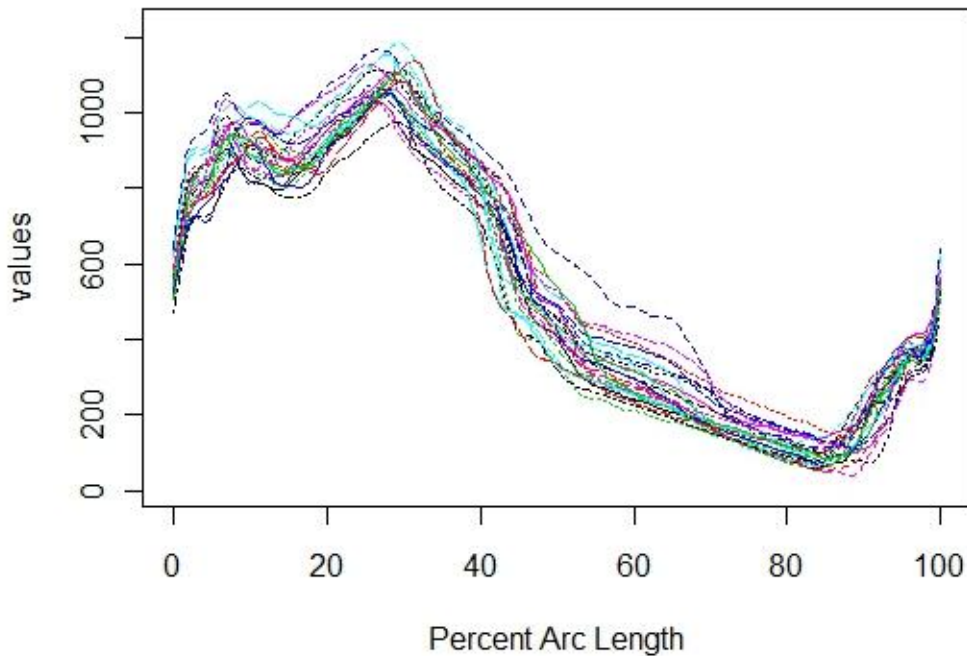


Figure 13: The ‘Y’ values for all mandibles as a function of arc length after registration has been performed.

Now that the data has been smoothed and registered, functional principal components analysis can be carried out to discern the important information from the data. The function ‘pca.fd’ is used to perform this analysis for our multivariate functional data, and the first three principal components are displayed in Figures 14 through 16 on page 27. The first three principal components account for 40.6%, 28.7%, and 10.5% of the variation, respectively, for a total of 79.8% of the overall variation. Again, the ‘X’ and ‘Y’ data are separated here for the ease of visualization, but they are considered together in the analysis. In each of these figures, the mean curve is shown in the middle, and the effect of the given principal component is shown by the curves on either side of the mean curve. This allows for visualization of the exact location on the curve where there is the most variation, as opposed to forcing this on to just a few predetermined landmarks.

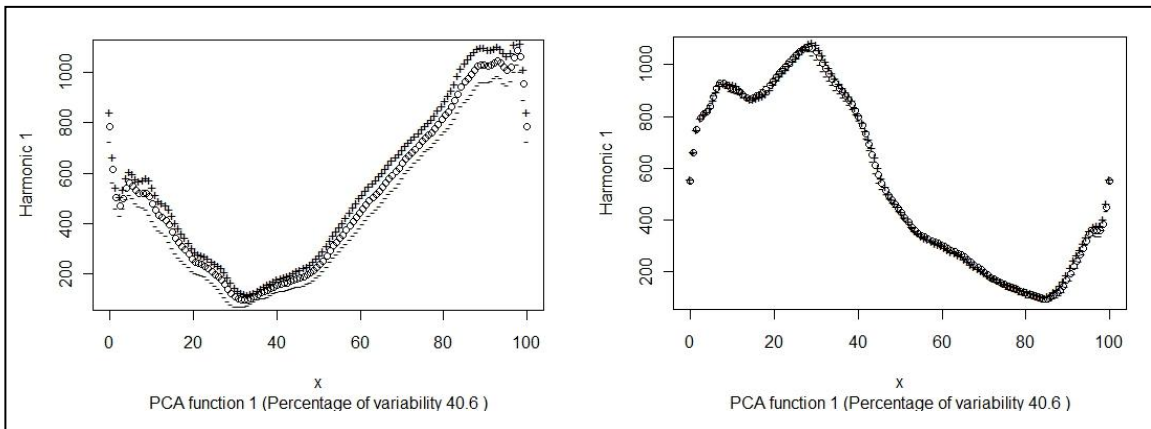


Figure 14: The effect of the first principal component on the ‘X’ values (left) and ‘Y’ values (right) of the mean function.

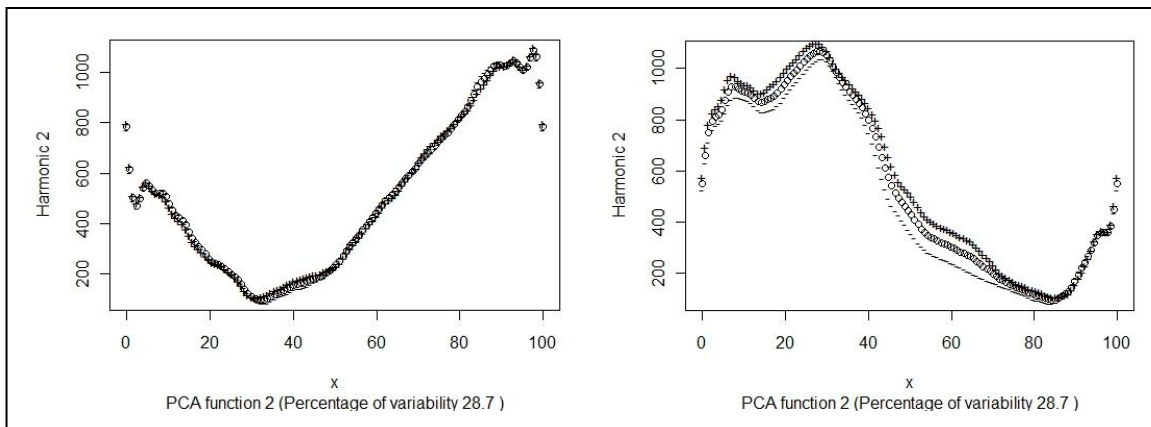


Figure 15: The effect of the second principal component on the ‘X’ values (left) and ‘Y’ values (right) of the mean function.

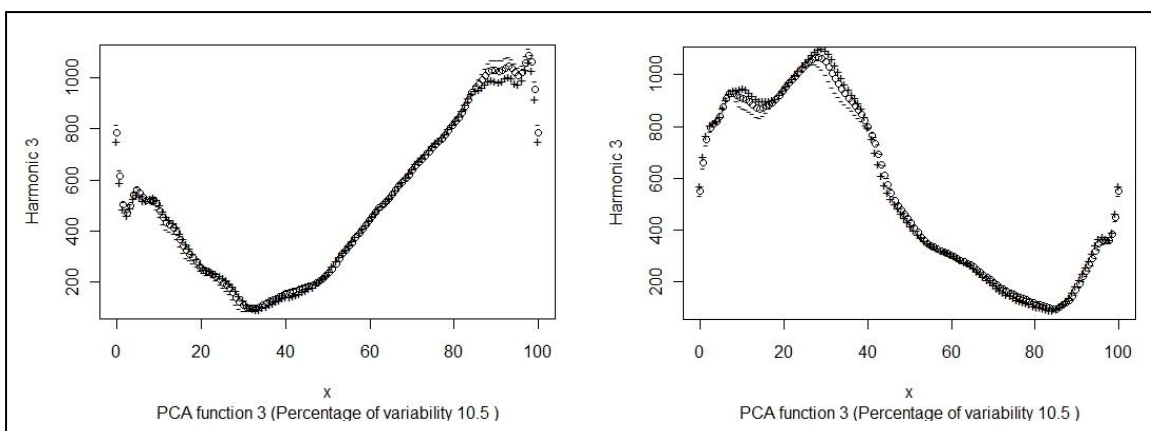


Figure 16: The effect of the third principal component on the ‘X’ values (left) and ‘Y’ values (right) of the mean function.

For this analysis, only three principal components are used because after these large initial few, the logarithm of eigenvalues decreases with a linear trend, as shown in Figure 17 (Ramsay et al., 2009). Because the first three are well above the linear trend, we can assume that they are the most important (Ramsay et al., 2009). With just these three principal components, we are able to see exactly where almost 80% of the variation lies on the mandible. Functional data analysis provides a tool to employ as much information about the curve as possible while still breaking down the variation into three simple principal components.

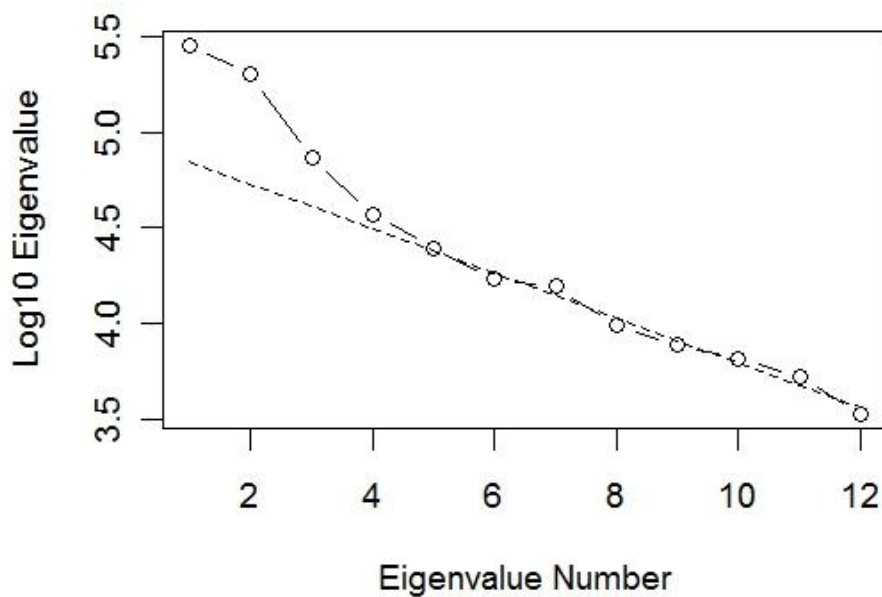


Figure 17: The log eigenvalues for each subsequent principal component.

Chapter 7: Conclusions and Recommendation

Functional data analysis has emerged as a valuable tool in many areas of application, and this list can be extended to cephalometry. Current measures used in cephalometric analysis are limited in the information they contain. Distances or angles between anatomical structures are the current basis of analysis, but this completely ignores the multi-dimensional relationships of structures like the contour of the mandible. Suppose a patient has a mandible of normal length, but the condyloid or coronoid process is deformed. The measurement of mandibular length as well as the ANB would be normal in the cephalometric data, and the deformed shape could go unnoticed. Landmark analysis helps solve this problem by characterizing shape in addition to these lengths and angles. However, landmark analysis has its limitations, too. Identifying landmarks is a process done by hand, using computer software to click each landmark in the exact same order for every image. Additionally, once this tedious and error-prone process is complete, the actual shape to analyze has been reduced to simply the number of landmarks chosen. This could potentially ignore a great deal of shape information.

Functional data analysis can solve these problems. It adds the multi-dimensional shape information that traditional cephalometric analysis is lacking, and it is much more user-friendly than landmark analysis. Treating each coordinate along the curve as a semi-landmark, functional data analysis provides hundreds or thousands of landmarks rather than just a dozen. Additionally, this digitization is automatic and does not require hand-selection that is prone to human error. Functional data analysis retains a very large amount of information, but it can still be summarized in a small number of principal components through the methods developed by Ramsay. In our study, the landmark analysis contained just 10 landmarks per image, and four principal components were necessary to account for 78.7% of the total variation. In comparison,

the functional data analysis contained 1000 semi-landmarks per image, and only three principal components were necessary to account for 79.8% of the total variation. Thus, a greater amount of information can be considered, an easier method can be used for computation, and an even more elegant summary of the information can be provided through the use of functional data analysis in cephalometrics.

Despite this optimistic suggestion of the role of functional data analysis in cephalometry, further analysis is necessary to confirm its efficacy. The data set used in this study consisted of only 24 healthy, normal cephalograms. Further research should analyze a wider range of mandibles, and an investigation should be pursued to find if functional principal component analysis can differentiate between different types of mandibles such as those that need mandibular elongation and those that do not. This type of analysis was not possible with the given data set, but the success of differentiation using functional principal component analysis of contours in related studies suggests that this would likely be possible to replicate with mandibles (Epifiano & Ventura-Campos, 2011; Nettel-Aguirre, 2008).

While it is an excellent step to begin considering two-dimensional shapes over one-dimensional lengths, another limitation of this study is that it does not include even more dimensions. The currently available data consist of two-dimensional lateral cephalograms, but this does not take into account the possibility of the mandible being offset to the left or right of the patient's face, something that would only show up in three dimensions. Because the functional principal component methods have already been developed to deal with high dimensions, it would be an interesting and likely simple extension to perform this analysis in three dimensions, as long as three-dimensional data were available.

All in all, this extension of functional data analysis could bring easier and more accurate examination methods of the mandible to the field of orthodontics. As long as the contour of the mandible has been traced, all of the methods for fitting functional objects and analyzing the principal components of the contour are free and easy to execute. Once a larger data set allows for a more accurate mean mandible to be calculated, subsequent mandibles can be compared to this standard through functional data analysis to standardize mandibular shape examination rather than relying on human judgment. While this analysis focuses on the shape of the mandible, orthodontists can further this idea by using functional data analysis to assess shape in even more aspects of a cephalogram, leading to easier and more informative assessments of potential orthodontic problems.

References

- Boeck, E. M., Kuramae, M., Lunardi, N., Santos-Pinto, A., & Mazzonetto, R. (2010). Cephalometric evaluation of surgical mandibular advancement. *Brazilian Oral Research*, 24(2), 189-196.
- Cozza, P., Baccetti, T., Franchi, L., DeToffol, L., & McNamara, J. A. (2006). Mandibular changes produced by functional appliances in Class II malocclusion: A systematic review. *American Journal of Orthodontics and Dentofacial Orthopedics*, 129(5).
- Dryden, I. L. & Mardia, K. V. (1998). *Statistical shape analysis*. West Sussex, England: John Wiley & Sons Ltd.
- Epifanio, I. & Ventura-Campos, N. (2011). Functional data analysis in shape analysis. *Computational Statistics and Data Analysis*, 55(9).
- Gonzalez-Manteiga, W. & Vieu, P. (2007). Statistics for functional data. *Computational Statistics and Data Analysis*, 51(1). 4788-4792.
- Hickey, A. & Vergo, T. J. (1991). Prosthodontic consideration in the treatment of patients with maxillary and mandibular deficiencies. *Journal of Prosthetic Dentistry*, 66(5), 645-649.
- Huja, S. S., Grubaugh, E. L., Rummel, A. M., Fields, H. W., & Beck, F. M. (2009). Comparison of hand-traced and computer-based cephalometric superimpositions. *Angle Orthodontist*, 79(3). 428-435.
- Jolliffe, I. T. (2004). *Principal component analysis*. (2nd ed.). New York, NY: Springer Science.
- Kendall, D. G. (1977). The diffusion of shape. *Advances in Applied Probability*, 9, 428-430.
- Little, R. M. (n.d.) Interpretation of cephalometric data. *University of Washington Ortho 521 Cephalometric Handbook*. Retrieved from [https://wiki.umms.med.umich.edu/download/attachments/127078836/Interpretation of Cephalometric Data.pdf](https://wiki.umms.med.umich.edu/download/attachments/127078836/Interpretation%20of%20Cephalometric%20Data.pdf)

- McNamara, J. A., Bookstein, F. L., & Shaughnessy, T. G. (1985). Skeletal and dental changes following functional regulator therapy on Class II patients. *American Journal of Orthodontics*, 88(2). 91-110.
- Molina, F. & Monasterio, F. O. (1995). Mandibular elongation and remodeling by distraction: A farewell to major osteotomies. *Plastic and Reconstructive Surgery*, 96(4). 825-840.
- Nettel-Aguirre, A. (2008). Nuclei shape analysis, a statistical approach. *Image Analysis and Stereology*, 27(1). 1-10.
- Piroozmand, F. (2001). A simple method to predict changes of mandibular landmarks following superior repositioning of the maxilla. *International Journal of Adult Orthodontics and Orthognathic Surgery*, 16(1). 72-75.
- Ramsay, J. O. (2000). Functional components of variation in handwriting. *Journal of the American Statistical Association*, 95(449). 9-15.
- Ramsay, J. O., Hooker, G., & Graves, S. (2009). *Functional data analysis with R and MATLAB*. New York, NY: Springer Science.
- Ramsay, J. O. & Li, X. (1998). Curve registration. *Journal of the Royal Statistical Society*, 60(2). 351-363.
- Ramsay, J. O. & Silverman, B. W. (2005). *Functional data analysis*. (2nd ed.). New York, NY: Springer Science.
- Rohlf, F. J. & Slice, D. E. (1990). Extensions of the Procrustes method for the optimal superimposition of landmarks. *Systematic Zoology*, 39, 40-59.
- Scolozzi, P., Link, D. W., & Schendel, S. A. (2007). Computer simulation of curvilinear mandibular distraction: Accuracy and predictability. *Plastic and Reconstructive Surgery*, 120(7). 1975-1980.

Singh, G. D., McNamara, J. A., & Lozanoff, S. (1998). Procrustes, Euclidean, and cephalometric analyses of the morphology of the mandible in human Class III malocclusions. *Archives of Oral Biology*, 43(7). 535-543.

Zelditch, M. L., Swiderski, D. L., & Sheets, H. D. (2012). *Geometric morphometrics for biologists: A primer*. (2nd ed.). Oxford, UK: Elsevier Inc.

Appendix A

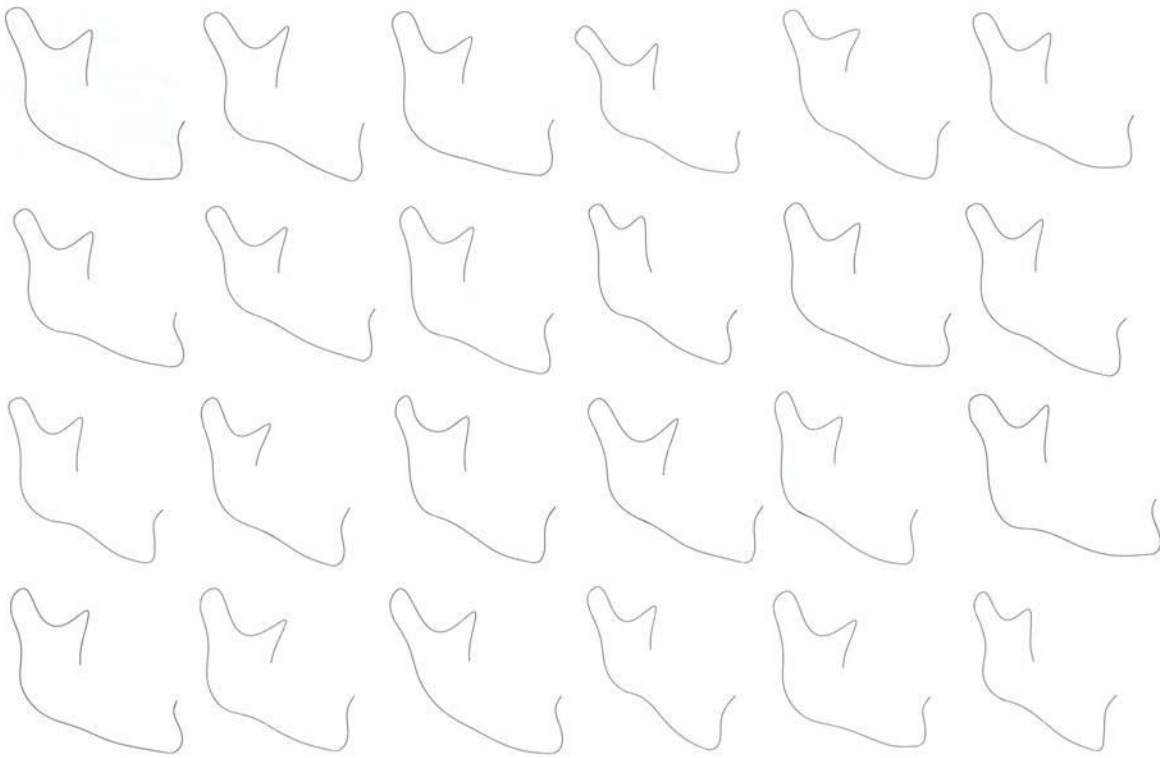


Figure 18: All 24 cephalometric tracings of the mandible used in this analysis.

Appendix B

The following is the full code used to generate images and perform all aforementioned analysis

using the free statistical software, R.

```
#####
#####
#### simple ANB analysis ####
#####
#####

par(mfrow=c(1,1))
data=read.csv("subjraceanb.csv",sep=" ",header=FALSE)
colnames(data)=c("subj","race","anb")

boxplot(data[,3]) #no outliers for anb

#####
#####
#### TPS landmark data - PCA ####
#####
#####

#load packages/libraries
install.packages("calibrate")
library(calibrate)
install.packages("RGraphics")
library(RGraphics)
install.packages("shapes")
library(shapes)

#initialize data array
dataarray=array(0,dim=c(10,2,24))

#ceph1
data=read.csv("tps01",header=TRUE)
data=do.call(rbind, strsplit(as.matrix(data), " "))
data=data[1:(nrow(data)-1),]
data=matrix(as.numeric(data),nrow=nrow(data),ncol=2)
dataarray[,,1]=data
```

```
#ceph2
data=read.csv("tps02",header=TRUE)
data=do.call(rbind, strsplit(as.matrix(data), " "))
data=data[1:(nrow(data)-1),]
data=matrix(as.numeric(data),nrow=nrow(data),ncol=2)
dataarray[, , 2]=data

#ceph3
data=read.csv("tps03",header=TRUE)
data=do.call(rbind, strsplit(as.matrix(data), " "))
data=data[1:(nrow(data)-1),]
data=matrix(as.numeric(data),nrow=nrow(data),ncol=2)
dataarray[, , 3]=data

#ceph4
data=read.csv("tps04",header=TRUE)
data=do.call(rbind, strsplit(as.matrix(data), " "))
data=data[1:(nrow(data)-1),]
data=matrix(as.numeric(data),nrow=nrow(data),ncol=2)
dataarray[, , 4]=data

#ceph5
data=read.csv("tps05",header=TRUE)
data=do.call(rbind, strsplit(as.matrix(data), " "))
data=data[1:(nrow(data)-1),]
data=matrix(as.numeric(data),nrow=nrow(data),ncol=2)
dataarray[, , 5]=data

#ceph6
data=read.csv("tps06",header=TRUE)
data=do.call(rbind, strsplit(as.matrix(data), " "))
data=data[1:(nrow(data)-1),]
data=matrix(as.numeric(data),nrow=nrow(data),ncol=2)
dataarray[, , 6]=data

#ceph7
data=read.csv("tps07",header=TRUE)
data=do.call(rbind, strsplit(as.matrix(data), " "))
data=data[1:(nrow(data)-1),]
data=matrix(as.numeric(data),nrow=nrow(data),ncol=2)
dataarray[, , 7]=data

#ceph8
data=read.csv("tps08",header=TRUE)
data=do.call(rbind, strsplit(as.matrix(data), " "))
data=data[1:(nrow(data)-1),]
data=matrix(as.numeric(data),nrow=nrow(data),ncol=2)
```

```
dataarray[, , 8]=data

#ceph9
data=read.csv("tps09",header=TRUE)
data=do.call(rbind, strsplit(as.matrix(data), " "))
data=data[1:(nrow(data)-1), ]
data=matrix(as.numeric(data), nrow=nrow(data), ncol=2)
dataarray[, , 9]=data

#ceph10
data=read.csv("tps10",header=TRUE)
data=do.call(rbind, strsplit(as.matrix(data), " "))
data=data[1:(nrow(data)-1), ]
data=matrix(as.numeric(data), nrow=nrow(data), ncol=2)
dataarray[, , 10]=data

#ceph11
data=read.csv("tps11",header=TRUE)
data=do.call(rbind, strsplit(as.matrix(data), " "))
data=data[1:(nrow(data)-1), ]
data=matrix(as.numeric(data), nrow=nrow(data), ncol=2)
dataarray[, , 11]=data

#ceph12
data=read.csv("tps12",header=TRUE)
data=do.call(rbind, strsplit(as.matrix(data), " "))
data=data[1:(nrow(data)-1), ]
data=matrix(as.numeric(data), nrow=nrow(data), ncol=2)
dataarray[, , 12]=data

#ceph13
data=read.csv("tps13",header=TRUE)
data=do.call(rbind, strsplit(as.matrix(data), " "))
data=data[1:(nrow(data)-1), ]
data=matrix(as.numeric(data), nrow=nrow(data), ncol=2)
dataarray[, , 13]=data

#ceph14
data=read.csv("tps14",header=TRUE)
data=do.call(rbind, strsplit(as.matrix(data), " "))
data=data[1:(nrow(data)-1), ]
data=matrix(as.numeric(data), nrow=nrow(data), ncol=2)
dataarray[, , 14]=data

#ceph15
data=read.csv("tps15",header=TRUE)
data=do.call(rbind, strsplit(as.matrix(data), " "))
```

```
data=data[1:(nrow(data)-1),]
data=matrix(as.numeric(data),nrow=nrow(data),ncol=2)
dataarray[, ,15]=data

#ceph16
data=read.csv("tps16",header=TRUE)
data=do.call(rbind, strsplit(as.matrix(data), " "))
data=data[1:(nrow(data)-1),]
data=matrix(as.numeric(data),nrow=nrow(data),ncol=2)
dataarray[, ,16]=data

#ceph17
data=read.csv("tps17",header=TRUE)
data=do.call(rbind, strsplit(as.matrix(data), " "))
data=data[1:(nrow(data)-1),]
data=matrix(as.numeric(data),nrow=nrow(data),ncol=2)
dataarray[, ,17]=data

#ceph18
data=read.csv("tps18",header=TRUE)
data=do.call(rbind, strsplit(as.matrix(data), " "))
data=data[1:(nrow(data)-1),]
data=matrix(as.numeric(data),nrow=nrow(data),ncol=2)
dataarray[, ,18]=data

#ceph19
data=read.csv("tps19",header=TRUE)
data=do.call(rbind, strsplit(as.matrix(data), " "))
data=data[1:(nrow(data)-1),]
data=matrix(as.numeric(data),nrow=nrow(data),ncol=2)
dataarray[, ,19]=data

#ceph20
data=read.csv("tps20",header=TRUE)
data=do.call(rbind, strsplit(as.matrix(data), " "))
data=data[1:(nrow(data)-1),]
data=matrix(as.numeric(data),nrow=nrow(data),ncol=2)
dataarray[, ,20]=data

#ceph21
data=read.csv("tps21",header=TRUE)
data=do.call(rbind, strsplit(as.matrix(data), " "))
data=data[1:(nrow(data)-1),]
data=matrix(as.numeric(data),nrow=nrow(data),ncol=2)
dataarray[, ,21]=data

#ceph22
```

```
data=read.csv("tps22",header=TRUE)
data=do.call(rbind, strsplit(as.matrix(data), " "))
data=data[1:(nrow(data)-1),]
data=matrix(as.numeric(data),nrow=nrow(data),ncol=2)
dataarray[, ,22]=data

#ceph23
data=read.csv("tps23",header=TRUE)
data=do.call(rbind, strsplit(as.matrix(data), " "))
data=data[1:(nrow(data)-1),]
data=matrix(as.numeric(data),nrow=nrow(data),ncol=2)
dataarray[, ,23]=data

#ceph24
data=read.csv("tps24",header=TRUE)
data=do.call(rbind, strsplit(as.matrix(data), " "))
data=data[1:(nrow(data)-1),]
data=matrix(as.numeric(data),nrow=nrow(data),ncol=2)
dataarray[, ,24]=data

#perform generalized procrustes analysis
x=procGPA(dataarray, scale=TRUE, reflect=FALSE, eigen2d=TRUE,
          toll=1e-05, tol2=1e-05, proc.output=FALSE, distances=TRUE,
          pcaoutput=TRUE, alpha=0, affine=FALSE)
#x

par(mfrow=c(1,1))
#plot mean shape, label landmarks, connect points
plot(x$mshape,type='o',pch=20,xlim=c(-0.5,0.5),ylim=c(-
      0.5,0.5),xlab="x",ylab="y",main="Procrustes Mean Mandible")
textxy(x$mshape[,1],x$mshape[,2],1:10)

#plot all procrustes mandibles on top of each other
plot(x$rotated[, ,1],type='l',col=1,xlim=c(-0.5,0.5),ylim=c(-
      0.5,0.5),xlab='x',ylab='y',main='All Procrustes Mandibles')
for(i in 2:x$n)
{
  lines(x$rotated[, ,i],type='l',col=i)
}

#plot first four PCs
par(mfrow=c(2,2))
for(a in 1:4)
{
```



```

plot(x$mshape,type='p',pch=20,xlim=c(-0.5,0.5),ylim=c(-
  0.5,0.5),xlab="x",ylab="y",main=paste("PC",as.character(a),
  ",",as.character(round(x$percent[a],2)),"%"))
for(i in 1:x$k) #i is landmark number
{
  tempx=0; tempy=0
  for(j in 1:x$n) #j is ceph number
  {
    tempx=tempx+(x$stdscores[j,a]*x$rotated[i,1,j])
    tempy=tempy+(x$stdscores[j,a]*x$rotated[i,2,j])
  }

  segments(x$mshape[i,1],x$mshape[i,2],tempx+x$mshape[i,1],te
    mpy+x$mshape[i,2])
}
}

#investigate differences in shape
par(mfrow=c(1,1))
x$rmsdl #root mean square of full Procrustes distance is 0.07
  which means shape variability in the data is quite small
boxplot(x$rho) #no outliers

#####
#####
##### FDA and PCA #####
#####
#####

#load data into array
dataarray=array(0,dim=c(1000,24,2))

data=read.csv("coord01.3.asc",header=FALSE)[,2:3]
diff=nrow(data)-1000
n=floor(nrow(data)/diff)
data=data[-(seq(n,to=n*diff,by=n)),]
dataarray[,1,1]=data[,1]
dataarray[,1,2]=data[,2]

data=read.csv("coord02.asc",header=FALSE)[,2:3]
diff=nrow(data)-1000
n=floor(nrow(data)/diff)
data=data[-(seq(n,to=n*diff,by=n)),]
dataarray[,2,1]=data[,1]

```

```
dataarray[,2,2]=data[,2]

data=read.csv("coord03.asc",header=FALSE)[,2:3]
diff=nrow(data)-1000
n=floor(nrow(data)/diff)
data=data[-(seq(n,to=n*diff,by=n)),]
dataarray[,3,1]=data[,1]
dataarray[,3,2]=data[,2]

data=read.csv("coord04.asc",header=FALSE)[,2:3]
diff=nrow(data)-1000
n=floor(nrow(data)/diff)
data=data[-(seq(n,to=n*diff,by=n)),]
dataarray[,4,1]=data[,1]
dataarray[,4,2]=data[,2]

data=read.csv("coord05.asc",header=FALSE)[,2:3]
diff=nrow(data)-1000
n=floor(nrow(data)/diff)
data=data[-(seq(n,to=n*diff,by=n)),]
dataarray[,5,1]=data[,1]
dataarray[,5,2]=data[,2]

data=read.csv("coord06.2.asc",header=FALSE)[,2:3]
diff=nrow(data)-1000
n=floor(nrow(data)/diff)
data=data[-(seq(n,to=n*diff,by=n)),]
dataarray[,6,1]=data[,1]
dataarray[,6,2]=data[,2]

data=read.csv("coord07.asc",header=FALSE)[,2:3]
data=rbind(data[7:nrow(data),],data[1:6,])
diff=nrow(data)-1000
n=floor(nrow(data)/diff)
data=data[-(seq(n,to=n*diff,by=n)),]
dataarray[,7,1]=data[,1]
dataarray[,7,2]=data[,2]

data=read.csv("coord08.asc",header=FALSE)[,2:3]
diff=nrow(data)-1000
n=floor(nrow(data)/diff)
data=data[-(seq(n,to=n*diff,by=n)),]
dataarray[,8,1]=data[,1]
dataarray[,8,2]=data[,2]

data=read.csv("coord09.asc",header=FALSE)[,2:3]
diff=nrow(data)-1000
```

```
n=floor(nrow(data)/diff)
data=data[-(seq(n,to=n*diff,by=n)),]
dataarray[,9,1]=data[,1]
dataarray[,9,2]=data[,2]

data=read.csv("coord10.asc",header=FALSE)[,2:3]
diff=nrow(data)-1000
n=floor(nrow(data)/diff)
data=data[-(seq(n,to=n*diff,by=n)),]
dataarray[,10,1]=data[,1]
dataarray[,10,2]=data[,2]

data=read.csv("coord11.asc",header=FALSE)[,2:3]
diff=nrow(data)-1000
n=floor(nrow(data)/diff)
data=data[-(seq(n,to=n*diff,by=n)),]
dataarray[,11,1]=data[,1]
dataarray[,11,2]=data[,2]

data=read.csv("coord12.asc",header=FALSE)[,2:3]
diff=nrow(data)-1000
n=floor(nrow(data)/diff)
data=data[-(seq(n,to=n*diff,by=n)),]
dataarray[,12,1]=data[,1]
dataarray[,12,2]=data[,2]

data=read.csv("coord13.asc",header=FALSE)[,2:3]
diff=nrow(data)-1000
n=floor(nrow(data)/diff)
data=data[-(seq(n,to=n*diff,by=n)),]
dataarray[,13,1]=data[,1]
dataarray[,13,2]=data[,2]

data=read.csv("coord14.asc",header=FALSE)[,2:3]
diff=nrow(data)-1000
n=floor(nrow(data)/diff)
data=data[-(seq(n,to=n*diff,by=n)),]
dataarray[,14,1]=data[,1]
dataarray[,14,2]=data[,2]

data=read.csv("coord15.asc",header=FALSE)[,2:3]
diff=nrow(data)-1000
n=floor(nrow(data)/diff)
data=data[-(seq(n,to=n*diff,by=n)),]
dataarray[,15,1]=data[,1]
dataarray[,15,2]=data[,2]
```

```
data=read.csv("coord16.asc",header=FALSE)[,2:3]
diff=nrow(data)-1000
n=floor(nrow(data)/diff)
data=data[-(seq(n,to=n*diff,by=n)),]
dataarray[,16,1]=data[,1]
dataarray[,16,2]=data[,2]
```

```
data=read.csv("coord17.asc",header=FALSE)[,2:3]
diff=nrow(data)-1000
n=floor(nrow(data)/diff)
data=data[-(seq(n,to=n*diff,by=n)),]
dataarray[,17,1]=data[,1]
dataarray[,17,2]=data[,2]
```

```
data=read.csv("coord18.asc",header=FALSE)[,2:3]
diff=nrow(data)-1000
n=floor(nrow(data)/diff)
data=data[-(seq(n,to=n*diff,by=n)),]
dataarray[,18,1]=data[,1]
dataarray[,18,2]=data[,2]
```

```
data=read.csv("coord19.asc",header=FALSE)[,2:3]
diff=nrow(data)-1000
n=floor(nrow(data)/diff)
data=data[-(seq(n,to=n*diff,by=n)),]
dataarray[,19,1]=data[,1]
dataarray[,19,2]=data[,2]
```

```
data=read.csv("coord20.asc",header=FALSE)[,2:3]
diff=nrow(data)-1000
n=floor(nrow(data)/diff)
data=data[-(seq(n,to=n*diff,by=n)),]
dataarray[,20,1]=data[,1]
dataarray[,20,2]=data[,2]
```

```
data=read.csv("coord21.asc",header=FALSE)[,2:3]
diff=nrow(data)-1000
n=floor(nrow(data)/diff)
data=data[-(seq(n,to=n*diff,by=n)),]
dataarray[,21,1]=data[,1]
dataarray[,21,2]=data[,2]
```

```
data=read.csv("coord22.asc",header=FALSE)[,2:3]
diff=nrow(data)-1000
n=floor(nrow(data)/diff)
data=data[-(seq(n,to=n*diff,by=n)),]
dataarray[,22,1]=data[,1]
```

```

dataarray[,22,2]=data[,2]

data=read.csv("coord23.asc",header=FALSE)[,2:3]
diff=nrow(data)-1000
n=floor(nrow(data)/diff)
data=data[-(seq(n,to=n*diff,by=n)),]
dataarray[,23,1]=data[,1]
dataarray[,23,2]=data[,2]

data=read.csv("coord24.asc",header=FALSE)[,2:3]
diff=nrow(data)-1000
n=floor(nrow(data)/diff)
data=data[-(seq(n,to=n*diff,by=n)),]
dataarray[,24,1]=data[,1]
dataarray[,24,2]=data[,2]

#create functional objects
fdarange=c(0,100)
fdabasis=create.fourier.basis(fdarange,45)
fdatime=seq(0,100,len=1000)
fdafd=smooth.basis(fdatime,dataarray,fdabasis)$fd
fdafd$fdnames[[1]]="Percent Arc Length"
fdafd$fdnames[[2]]="Replications"
fdafd$fdnames[[3]]=list("X","Y")
plot(fdafd)

#perform continuous registration
reglist=register.fd(fdafd$meanfd,fdafd)
names(reglist)
plot(reglist$regfd)
#plot(reglist$warpdf)

nharm=4
fdapcaList=pca.fd(reglist$regfd,nharm)
plot.pca.fd(fdapcaList) #plots first 4 PCs

plot(fdapcaList$meanfd) #plots registered mean X and registered
mean Y

#eigenvalue information
fdaeig=fdapcaList$values
neig=12
x=matrix(1,neig-nharm,2)
x[,2]=(nharm+1):neig
y=log10(fdaeig[(nharm+1):neig])
c=lsfit(x,y,int=FALSE)$coef

```

```
par(mfrow=c(1,1),cex=1.2)
plot(1:neig,log10(fdaeig[1:neig]),"b",xlab="Eigenvalue
      Number",ylab="Log10 Eigenvalue")
lines(1:neig,c[1]+c[2]*(1:neig),lty=2)
```