University of Alberta

FUNCTIONAL DATA ANALYSIS WITH APPLICATION TO MS AND CERVICAL VERTEBRAE DATA

by

Kate Yaraee

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science

in

Statistics

Department of Mathematical and Statistical Sciences

Fall 2011
Edmonton, Alberta

To my husband John for his great encouragement and support

Abstract

Functional data analysis (FDA) is a fast-growing area in statistics with the aim of estimating a set of related functions or curves rather than focusing on a single entity, like estimating a point, as in classical statistics. FDA has a wide range of applications in different fields such as biological sciences, geology, psychology, economics, image data, etc. In fact, discrete data points collected over a continuum can be thought of as a function that is assumed to be a reasonably smooth mechanism giving rise to those discrete observed points. This continuum is not necessarily a physical time point, but rather attributes such as age, spatial location, seasons, or temperature.

This thesis discusses the developed methods and techniques of functional data analysis and explores the versatile applications of FDA by employing those techniques for three sets of studies including Multiple Sclerosis (MS) data provided by clinicians from the field of neurological sciences, data taken from the field of anesthiology and Pain Medicine, and finally data collected from skeletal maturation study conducted by clinicians in the field of orthodontics.

Acknowledgements

My sincere thanks to my co-supervisor, Dr. Giseon Heo, without whose great support, wise guidance, and encouragement this thesis would not have been possible. I am truly grateful for her patience and enthusiastic help.

My great thanks to my supervisor, Dr. Douglas P. Wiens, for his thoughtful advice, constant support, and wisdom.

I would also like to thank Dr. Yunyan Zhang, and Dr. Wee Yong, Departments of Clinical Neurosciences and Oncology, University of Calgary; Dr. Bradley J. Kerr and his student Camille Olechowski, Department of Anesthesiology and Pain Medicine, University of Alberta; and Dr. Manuel Lagravère, Department of Dentistry, University of Alberta, for providing me with the datasets.

I would like to convey my thanks to the endMS Regional Research and Training Centres (RRTCs) for their financial support.

Finally, thanks to John for being understanding, supportive, and encouraging.

Kate

Contents

List of Figures

Chapter 1: Introduction

Functional data analysis (FDA) has been widely used across many disciplines and statisticians have shown a great interest in this area of study. The very beginning of its development can be traced back to around 1800 when Gauss and a French mathematician, Legendre, were trying to model and estimate the pathway of a comet that formed a curve. The usage of the term FDA was first developed by Ramsay and Dalzell (1991), and it is involved with a new approach which manifests the results mostly through graphical illustrations. Many of the methods used in classical statistics have their counterparts in the concept of FDA. Some methods are simply the extension of existing techniques in conventional statistics while others need more than exchanging the summation, used in discrete observation, to an integration, which is a continuum. We introduce some of the exploratory data analysis techniques adapted for functional data and explore the variability within and between curves using those tools.

The underlying theory of the FDA methods will be presented in this introductory chapter. The second and third chapters will explore the application of developed theory on the one dimensional curves applied to two sets of data. The first dataset

is related to the patients with Multiple Sclerosis (MS) disease and the second one is taken from the animal model with two groups: Experimental Autoimmune Encephalomyelitis (EAE) and Complete Freund's Adjuvant (CFA). The fourth chapter is allotted to the method of two dimensional data applied to the Cervical Vertebrae (CV) dataset, which is obtained from a clinical study in orthodontics. The detailed explanation of each set of data will be given in the related chapters. Each chapter ends with some conclusions and remarks explained in the body of that chapter, as well as recommendations for the future studies and research.

All the theories and methodologies explained in the subsequent chapters are adapted from the methodologies developed by Ramsay and Silverman (2002, 2005) and Ramsay, Hooker, and Graves (2009), unless otherwise specified. The [MATLAB] software, version [R2010b], Copyright © [2011], MATLAB(R), MATLAB(R) Compiler(TM), and other MATLAB family products and [R] software, version [2.12.2], Copyright © [2011] have been used to carry out all the analyses.

1.1 Goals and examples in FDA

A wide variety of applications of FDA methodologies can be illustrated through some examples in different disciplines. In the field of psychology, for example, researchers are interested in investigating how listening to a piece of music with both audio and video available to the audience has different emotional effects on people, compared to when they listen to a piece of music without seeing the performers or when they only see the performance without any audio available (Ramsay, 2004). Methods in FDA

can be employed to reveal any component of variability among these groups through Functional Principal Component Analysis (FPCA) techniques. Another example of this is taken from neuroimaging, where FDA is used to determine and study the connection between two components, brain and mind, over a continuum of time (Tian, 2010). Tian has discussed the methods of FDA employed in the field of brain imaging, where the dimensionality reduction approach and classification of spatial location using functional magnetic resonance imaging (fMRI) were used in particular. Its vast implementation is also extended to the field of behavioural sciences, namely, criminology, where each individual is traced for several years to keep the record of their criminal activities. In analyzing these types of data FDA plays a major role. Ongoing research is also exploring one of the most powerful techniques in FDA called dynamic models. An example of the application of FDA through dynamic models is given by Ramsay (2010), in the field of engineering. In his model, the amount of rainfall (in North Vancouver) and groundwater level are used as predictors for the amount of change in the level of groundwater; the model is known as a first order nonhomogeneous ordinary differential equation. Modeling the data by dynamic systems rather than Functional Linear Model (FLM) is more appropriate in two ways: first, FLM is unable to predict the rate of change in the groundwater level, which is the primary interest of the firm Bruce Geotechnical Consultants (BGC) engineers, a firm in geotechnical and water resources engineering and applied earth sciences, since the flooded places have to be evacuated as soon as the groundwater reaches a certain level. Second, the dynamic models are able to account for the lag between the time of rainfall and the increase in groundwater level, which, like most other processes in

the nature, happens with a delay. There are many other examples that illustrate the importance and appropriateness of dynamic modeling in the concept of FDA.

Our aim is to study one-dimensional functions (as in MS patients and animal models) and two-dimensional curves (CV data), to investigate any variability among replications in both human and animal models. The goals of using FDA methodologies are quite similar to other conventional statistical methods. As a first step in any statistical analysis, we visualize the data by conducting a functional version of descriptive statistics, which helps us to do further analysis by seeking any pattern or remarkable structures such as peaks and valleys in the data. We look for modes of variation within each replication using the input information, and also variations from curve to curve. We might also be interested in finding the source of those between- and within-curve variations. To do so, we first need to define a model for our data. This leads to the next section on how to model discrete observations as curves.

1.2 Modelling in FDA

Let y_j be the observed discrete data points for $j = 1, 2, ..., n$, collected over a continuum time, t. The time over which the measurements are taken does not necessarily need to be a physical time but rather any other continuum such as age, height, spatial location, frequency, etc. Further, assume that there is a reasonably smooth function $x(t)$ that gives rise to those discrete points. We define the model as

$$y_j = x(t_j) + \epsilon_j. \tag{1.2.1}$$

Function $x(t)$ is constructed using a system of basis functions (explained in Section 1.3), which is a linear combination of K independent basis functions denoted by $\phi_k(t)$ for $k = 1, 2, ..., K$. That is,

$$x(t) = \sum_{k=1}^{K} c_k \, \phi_k(t) = \mathbf{c}^T \boldsymbol{\phi}(t), \tag{1.2.2}$$

where $\mathbf{c}$ is the coefficients vector of length K and $\boldsymbol{\phi}(t)$ is the K-vector of basis functions. The vector format of formula (1.2.1) can be written as

$$\mathbf{y} = x(\mathbf{t}) + \mathbf{e}, \tag{1.2.3}$$

where $\mathbf{y}$, $\mathbf{t}$, and $\mathbf{e}$ are all column vectors of length n. The last term in (1.2.3) indicates the errors or disturbances that add up to the roughness of the measured data. Unlike classical statistics, where we usually assume that errors are independent and identically distributed with zero mean and constant variance, in functional data models often errors are correlated over time. We denote variance-covariance matrix of the errors by $\sum_e$.

Another way of representing the model in FDA is to write the discrete data, y, as measurements of the i^{th} observation at the j^{th} time point with $i = 1, 2, ..., N$ where N is the number of replications or curves under study. Thus, it has the form:

$$y_{ij} = x_i(t_j) + \epsilon_{ij}, \tag{1.2.4}$$

where y_{ij} are values measured as a set of discrete points including y_{i1},$y_{i2}, ..., y_{in}$ and

$x_i(t_j)$ is the i^{th} function that captures these discrete points at the j^{th} time point and these functions, x_i, can be estimated by a linear combination of K independent basis functions ϕ_k. The functions x_i are independent unless there is not enough sample, n, per curve for estimating a particular curve. In this case we are forced to borrow information from neighbouring curves to estimate a function which does not have adequate sample size. This implies dependency between curves x_i. The model is now formulated as:

$$x_i(t) = \sum_{k=1}^{K} c_{ik}\, \phi_k(t), \tag{1.2.5}$$

where c_{ik} denotes the coefficients of expansion that are estimated from the data. The next section comprises the details.

1.3 Basis functions, splines, and B-splines

Basis functions are a set of mathematically independent functions, denoted by $\phi_k(t)$, that are combined and used to estimate any function, $x(t)$. There are several different types of basis functions that can be used to estimate a function, namely, polynomial, constant, Fourier, spline, and wavelet basis. The type of basis functions used to estimate a function is heavily dependent on the underlying behaviour of the data. For instance, Fourier basis functions are used for data with periodic nature, spline basis functions are more suitable for non-periodic data, and wavelet basis functions are used to estimate functions with discontinuities and so on. In any case, the closer the features of the basis functions are to those of the data, the better the estimation of the function $x(t)$ will be. In fact, we rely on the correct choice of basis functions

which are surmised to have close features to the true function, $x(t)$.

Our main focus will be on the spline basis functions, as all datasets we work with have non-periodic behaviour in nature. Splines themselves are defined to be a linear combination of basis functions. There are different types of basis functions for constructing a spline, such as M-splines, B-splines, truncated power functions, and natural splines. In defining a spline function, regardless of the type of basis functions that have been chosen, there are several components that need to be elucidated:

1. number of sub-intervals L;
2. number of knots (values at each breakpoint);
3. number of breakpoints with breakpoint sequence τ_l where $l = 0, 1, ..., L$; breakpoints are uniquely defined knot values.
4. interval over which a function is being estimated denoted by τ where $\tau = [T_1, T_2]$;
5. number of interior knots $L - 1$ (assuming that there is only one knot placed at each breakpoint);
6. order m (degree $m - 1$) of polynomials at each interval;
7. number of derivatives that need to be matched up at each junction, $m - 2$;
8. placement of breakpoints or knots.

The order m is the number of parameters used to define a function, and it is defined to be one more than the degree of a polynomial, where degree represents the highest degree in a polynomial. For example, a cubic spline has a degree of 3

and order 4. Also, there are different ways of placing the breakpoints. One could place them at each time point for which measurements are taken, but this is mostly done when the sample size is not too large. Another way of doing this is to place the knots at equally-spaced positions. This method is particularly useful if there are many sample points per curve and if the samples are taken at approximately equal time points. One could also place more knots where there are more curvatures in the data and fewer where there are sparse samples.

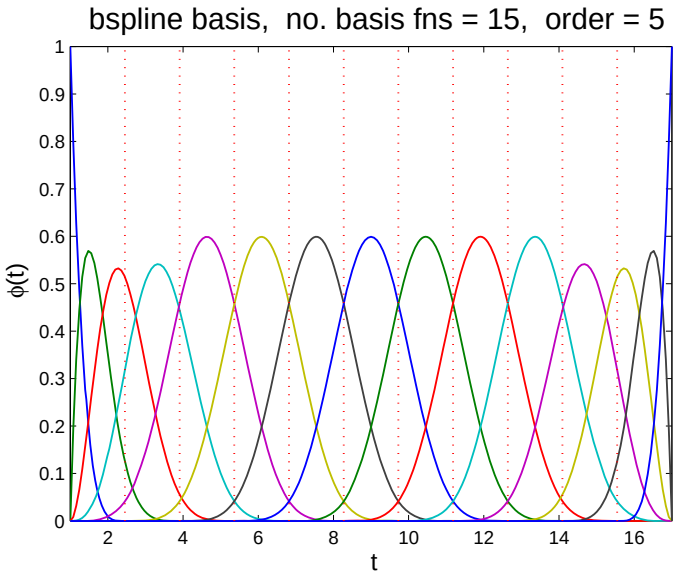

Figure 1.3.1: The fifteen B-spline basis of order 5 with equally-spaced knots.

We work with the most commonly used spline basis functions, B-splines. A spline function with B-spline basis functions has the form $S(t) = \sum_{k=1}^{m+L-1} c_k B_k(t, \tau)$ where B denotes B-spline basis functions at time t with knot sequence τ, m and L are defined as before, and the c_k indicate the coefficients of expansions. Note that every basis function, $B_k(t, \tau)$, is itself a spline function and that adding, subtracting, or multiplying several splines is still a spline. The efficient algorithm to compute B-spline

curves has been developed by de Boor (1977). A set of fifteen B-spline functions with order five and 10 interior knots is illustrated in Figure 1.3.1. In this figure, each curve is a B-spline basis function of order 5 with equally–spaced knots and the combination of all 15 B-spline functions forms a spline of order 5 with equally-spaced knots. Hastie, Tibshirani, and Friedman (2009) provide more details on B-spline basis computation.

1.4 Converting discrete points to functional data

In order to estimate a reasonably smooth function $x(t)$, which can be assumed to generate the discrete data points, we need to consider a smoothing method to reduce or ignore the unwanted errors. There are several methods to smooth a function in the literature. Kernel smoothing method has been discussed by Hastie, Tibshirani, and Friedman (2009). Local polynomial fitting method is another technique of approximating a function elaborated by Fan and Gijbels (1996). Each of these methods has its own advantages and disadvantages. The most familiar method among statisticians is the well-known method of Least Squares (LS) estimation. It minimizes the sum of squared errors and is formulated as Sum of Squared Errors (SSE) where $SSE(y|c) = \sum_{j=1}^{n}[y_j - \sum_{k=1}^{K} c_k\phi_k(t_j)]^2$. This can be written in a matrix form as $SSE(\mathbf{y}|\mathbf{c}) = (\mathbf{y} - \mathbf{\Phi c})^T(\mathbf{y} - \mathbf{\Phi c})$ where $\mathbf{c}$ is the coefficient vector of length K, $\mathbf{\Phi}$ is the $n \times K$ matrix including the values $\phi_k(t_j)$, and $\mathbf{y}$ is the n-vector of discrete data points. If errors are not assumed to be independent with zero mean and constant variance, we include weight $\mathbf{W}$ in the above formula. Thus, the formula has the form

$$SSE(y|c) = (\mathbf{y} - \mathbf{\Phi c})^T \mathbf{W} (\mathbf{y} - \mathbf{\Phi c}), \tag{1.4.1}$$

where $\mathbf{W}$ is the $n \times n$ symmetric positive definite weight matrix which is defined to be the inverse of the variance-covariance matrix of errors, $\sum_e^{-1}$. Weights are set to identity matrix if the errors are assumed to be independently and identically distributed (i.i.d).

1.5 Roughness penalty approach

The problem with LS method is that we do not have control over the level of smoothness. We seek a model which provides us with a better approach and allows us to control the smoothness of a function. Therefore, a more powerful method of smoothing called roughness penalty or regularization is used throughout the analysis. The basic idea of the roughness penalty approach is similar to LS estimation except that we add a penalty term multiplied by a smoothing parameter to the formula (1.4.1) which plays the role of penalizing the roughness in the estimated curve and yields a better result. The superiority of using the roughness penalty approach over the LS method has been shown through a simulation study of second derivatives of an estimated curve (Ramsay and Silverman, 2005, p. 90-91). The penalized sum of squared errors (PENSSE) is defined as

$$PENSSE_m(\mathbf{y}|\mathbf{c}) = (\mathbf{y} - \mathbf{\Phi c})^T \mathbf{W} (\mathbf{y} - \mathbf{\Phi c}) + \lambda \, PEN_m(x), \tag{1.5.1}$$

where $PEN_m(x)$ is the penalty term defined as the integral of the squared m^{th} derivative of the function $x(t)$, $\int [D^m x(s)]^2 ds$, and it measures the amount of variability and roughness in the function $x(t)$. The first term in (1.5.1) elucidates the fit to the data and λ in the second term is called the smoothing parameter– a positive value that controls the smoothness of the estimated function, $\widehat{x}(t)$. More on the role of smoothing parameter and how to choose one is developed in Section 1.6. $PENSSE$ is rewritten as

$$PENSSE_m(\mathbf{y}|\mathbf{c}) = (\mathbf{y} - \mathbf{\Phi c})^T \mathbf{W} (\mathbf{y} - \mathbf{\Phi c}) + \lambda \int [D^m x(s)]^2 ds. \tag{1.5.2}$$

Since $x(s)$ is equal to $\mathbf{c}^T \boldsymbol{\phi}(s)$, after applying some linear algebra we obtain:

$$PENSSE_m(\mathbf{y}|\mathbf{c}) = (\mathbf{y} - \mathbf{\Phi c})^T \mathbf{W} (\mathbf{y} - \mathbf{\Phi c}) + \lambda\, \mathbf{c}^T \mathbf{R c}, \tag{1.5.3}$$

where $\mathbf{R}$ is called the penalty matrix of order K, defined as the integral of products of m^{th} derivatives of the basis functions, $\mathbf{R} = \int D^m \boldsymbol{\phi}(s)\; D^m \boldsymbol{\phi}^T(s)\; ds$. Putting it all together and taking the derivative of $PENSSE_m(\mathbf{y}|\mathbf{c})$ with respect to $\mathbf{c}$ results in $\widehat{\mathbf{c}} = (\mathbf{\Phi}^T \mathbf{W} \mathbf{\Phi} + \lambda \mathbf{R})^{-1} \mathbf{\Phi}^T \mathbf{W} \mathbf{y}$**,** where $\widehat{\mathbf{c}}$ is the estimated coefficients vector of length K and $\mathbf{\Phi}$ and $\mathbf{W}$ are defined as before. Consequently, $\widehat{\mathbf{y}} = \mathbf{\Phi}\widehat{\mathbf{c}} = \mathbf{\Phi}(\mathbf{\Phi}^T \mathbf{W} \mathbf{\Phi} + \lambda \mathbf{R})^{-1} \mathbf{\Phi}^T \mathbf{W} \mathbf{y}$. Denoting $\mathbf{\Phi}(\mathbf{\Phi}^T \mathbf{W} \mathbf{\Phi} + \lambda \mathbf{R})^{-1} \mathbf{\Phi}^T \mathbf{W}$ by $\mathbf{S}_{\phi,\lambda}$, $\widehat{\mathbf{y}}$ is written as $\mathbf{S}_{\phi,\lambda}\mathbf{y}$ and $\mathbf{S}_{\phi,\lambda}$ is called smoothing matrix or sub-projection operator matrix. This smoothing matrix has the role similar to the hat matrix in classical multiple linear regression model but with somehow different properties. For example, idempotency property does not hold for

the smoothing matrix, $\mathbf{S}_{\phi,\lambda}$, that is, $\mathbf{S}_{\phi,\lambda}\mathbf{S}_{\phi,\lambda} \neq \mathbf{S}_{\phi,\lambda}$.

1.6 More on smoothing parameter

The smoothing parameter, λ, controls a compromise between the fit to the data and the variability in the function. It is usually chosen by a data-driven method called Generalized Cross-Validation (GCV) which is known to give the optimal solution, according to Gu (1992). The method was developed by Craven and Wahba (1979) and is defined to be

$$\begin{aligned} GCV\,(\lambda) &= \frac{n\ trace(\mathbf{Y}^T[\mathbf{I}-\mathbf{S}_{\phi,\lambda}]^{-2}\mathbf{Y})}{(trace[\mathbf{I}-\mathbf{S}_{\phi,\lambda}])^2} \\ &= (\frac{n}{n-df(\lambda)})(\frac{SSE}{n-df(\lambda)})\ , \end{aligned} \tag{1.6.1}$$

where $df(\lambda) = trace(\mathbf{S}_{\phi,\lambda})$, $\mathbf{Y}$ is the $n\ \times N$ data matrix. The first term on the right hand side in (1.6.1) is always greater than one and the second term is an unbiased estimate of σ^2 which is increased by dividing SSE by $n - df(\lambda)$ rather than n. The term σ^2 is the variability of errors in the model $y_j = x(t_j) + \epsilon_j$ denoted as $var(\epsilon_j)$. The quantity in formula (1.6.1) is referred to as being "twice-discounted mean squared error measure" (Ramsay and Silverman 2005, p. 97).

In formula (1.5.1), for small values of λ, the estimated curve becomes more variable since it is being penalized less for its roughness. In the case when $\lambda \to 0$, the curve fits the discrete points exactly at all sampling points, that is, $y_j = x(t_j)$ for all j. The

curve fitting problem in this case becomes an interpolation problem. On the other hand, as $\lambda \to \infty$, variability in function $x(t)$ become so small that the fitted curve approaches standard linear regression ($PEN = 0$).

The graphical method to choose λ includes plotting $log(\lambda)$, for some chosen λ, against the values of $GCV\,(\lambda)$ in formula (1.6.1). If $GCV\,(\lambda)$ does not change much with different choices of $log(\lambda)$, then this is an indication of lack of information in the data to help us choose the more suitable smoothing parameter. In this case, we rely on our judgement; we use different values for λ and see which one results in a closer estimated curve to the raw data. This way of choosing λ seems to be subjective but there is no universally good smoothing parameter even through GCV method.

1.7 Other form of roughness penalty

In formula (1.5.2) we penalized the m^{th} derivative of function $x(t)$, shown by $D^m x$. An alternative is to penalize a linear differential operator denoted by Lx such that $Lx = \beta_0 x + \beta_1 Dx + ... + \beta_{m-1} D^{m-1} x + D^m x$ where coefficients $\beta_0, \beta_1, ..., \beta_{m-1}$ can be constants or functions. Thus, the penalty in (1.5.1) can be expressed as

$$PEN_L(x) = \int [(Lx)^2](s)ds = \| \; Lx \; \|^2 \; .$$

It has been shown that using Lx penalty of order m results in a smaller Integrated Mean Squared Error ($IMSE$) than using $D^m x$ (Wahba 1990), where $IMSE$ is defined to be $IMSE(\widehat{x}) = \int E[\widehat{x}(t) - x(t)]^2 dt$.

1.8 Functional mean and variance

As a part of descriptive statistics, the functional version of ordinary mean is defined to be

$$\overline{x}(t) = \frac{1}{N}\sum_{i=1}^{N} x_i(t), \tag{1.8.1}$$

for $i = 1, 2, ..., N$ where N is the number of curves or replications and $x_i(t)$ is each curve or function evaluated at time t. Similarly, we define the variance of a function as

$$var[x(t)] = \frac{1}{N-1}\sum_{i=1}^{N}[x_i(t) - \overline{x}(t)]^2. \tag{1.8.2}$$

The quantities defined in formulas (1.8.1) and (1.8.2) are the counterparts of the descriptive statistics in functional data.

Chapter 2: Application of FDA to MS data

2.1 Background of the data

Ten patients with Relapsing-Remitting Multiple Sclerosis (RRMS) have been followed up for three years. A written consent has been obtained from all patients and the approval of the Calgary research ethics board has been acquired. The criteria to include these patients in this study were based on: a) being between 18-50 years old; b) having Extended Disability Status Scale (EDSS) between 0-5.5; and c) having had at least 2 relapses during the last two years. The time intervals at which measurements were taken are not equally spaced. At the beginning of the study, there were 10 consecutive monthly measurements of the number of MS lesions counts and volume of lesions, at times -3 (three months before the treatment), -2 (2 months before the treatment), -1 (a month before the treatment), 0 (baseline), and $1, 2, 3, 4, 5$, and 6 months after the treatment. Afterwards, the patients were followed up on an annual basis. Thus, the last three observations were taken at time points 12 (a year after the

treatment), 24 (two years after the treatment), and 36 months (three years after the treatment). Overall, there are 13 time points at which measurements are taken. The $T2$-weighted MRI images were used to take these measurements.

2.2 Missing values

Occurrence of missing values is an inevitable part of many processes and it happens frequently, especially in clinical studies. Our MS datasets do not constitute an exception. There are 42 missing values among the observed data, 21 in the variable counts and 21 in the variable volume. In general, in the case of having large sample sizes, we can discard the subjects with missing values; otherwise we can use the imputation methods to estimate the missing values. In order to do the imputation, we rely on the assumption of Missing At Random (MAR) or Missing Completely At Random (MCAR); the latter is a stronger assumption and it ensures that the method used for imputation is well-founded. Assumption of MCAR implies that no pattern in the missing observations can be found (Hastie, Tibshirani, Friedman, 2009, p. 332-333). The MS datasets that we are analysing in this chapter are MCAR. Some existing methods such as the interpolation (before/ after/ linear), the substitution (zero/ mean/ median), and an ad hoc method of matching cases within imputation classes were tested. The results of many of the above mentioned methods were inspected graphically.

The linear interpolation method was chosen to impute the missing values which resulted in fairly reasonable substitutions. It was done in [R] software using func-

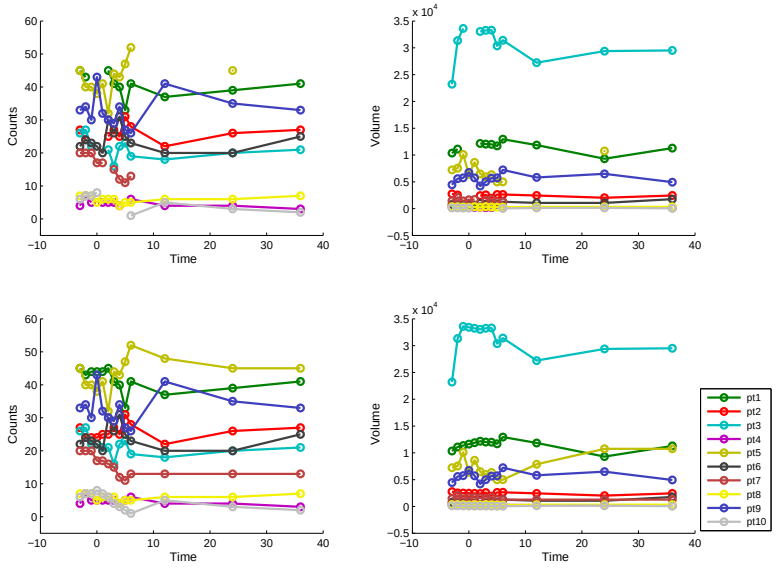

Figure 2.2.1: Missing values (top) estimated missing values (bottom). These plots show the observed values at each patient visiting time. The lines between the observations over time are added to ease graphical presentation.

tion 'interpNA' in library (fSeries). This method uses the time index in the data, regardless of the length of the intervals between the two consecutive time points, and interpolates the missing values based on the previous and the next observed values linearly, performed within each subject. Note that if there are any missing values in the last time points, they will be replaced by the previous values. Figure 2.2.1 shows the raw data, counts and volume of the MS lesions, with and without missing values in the first and second rows respectively, over the course of the study period.

2.3 Converting raw data to smoothed curves

In this case study we decided to work with the log transformation of the data since: a) it has nicer mathematical properties; b) it results in a lower variability; and c) it eliminates the possibility of having negative estimations of the smoothed curves. Plots of the log transformed data of volume and counts lesions are shown in Figure

2.3.1. The model can be written as in the formula (1.2.1) or (1.2.4). To make the notations clearer, we consider all formulas for one curve using formula (1.2.1). Let y_j denote the discrete points where $j = 1, 2, ..., n = 13$. The model is then defined as $y_j = x(t_j) + \epsilon_j$.

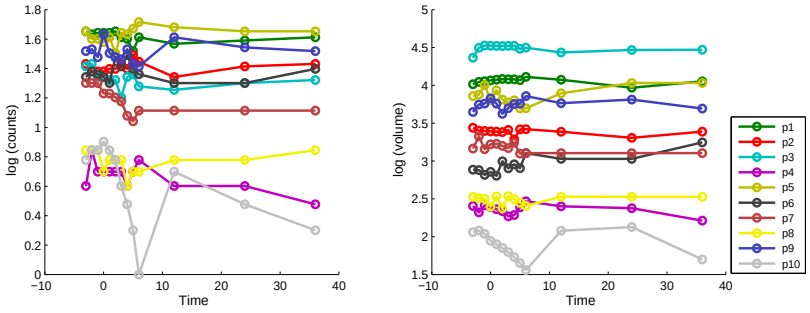

Figure 2.3.1: The log transformed data for the variables volume (right) and counts (left).

The spline smoothing method has been employed to estimate the smoothed curves. We work with the 15 B-spline basis functions of order 4, called cubic splines, with knots placed at each time point. Also, we penalize the second order derivative or alternatively a differential equation of order 2. The penalized second order of the derivative ensures that the estimated curve itself is smooth. This knot placement method works well since with a small sample size (samples per curve) it ensures that there is at least one point at each interval to estimate the smoothed curve. According to a theorem by de Boor (2002), this setting results in an efficient way of estimating the curve, $x(t)$. This theorem indicates that a cubic spline with knots placed at each time point will result in the minimum fitting criterion,

$$PENSSE_\lambda(x|\mathbf{y}) = [\mathbf{y} - x(\mathbf{t})]^T \mathbf{W} [\mathbf{y} - x(\mathbf{t})] + \lambda \times PEN_2(x).$$

One way to choose λ is to use the GCV method in formula (1.6.1). But as explained in Section 1.6 of Chapter 1, if λ_{GCV} tends to oversmooth or undersmooth the estimated curves, we need to choose the most suitable λ according to our own judgement through trial and error. We select $\lambda = 4$ to estimate both volume and counts curves. The smoothing parameter reported by the GCV approach was 1. This discrepancy happens when there is a correlation between or within the smoothed curves, since the GCV method heavily relies on the independence of the data. We examined other settings for smoothing the curves but no dramatic changes with higher orders of the B-splines or different knot placement methods were observed.

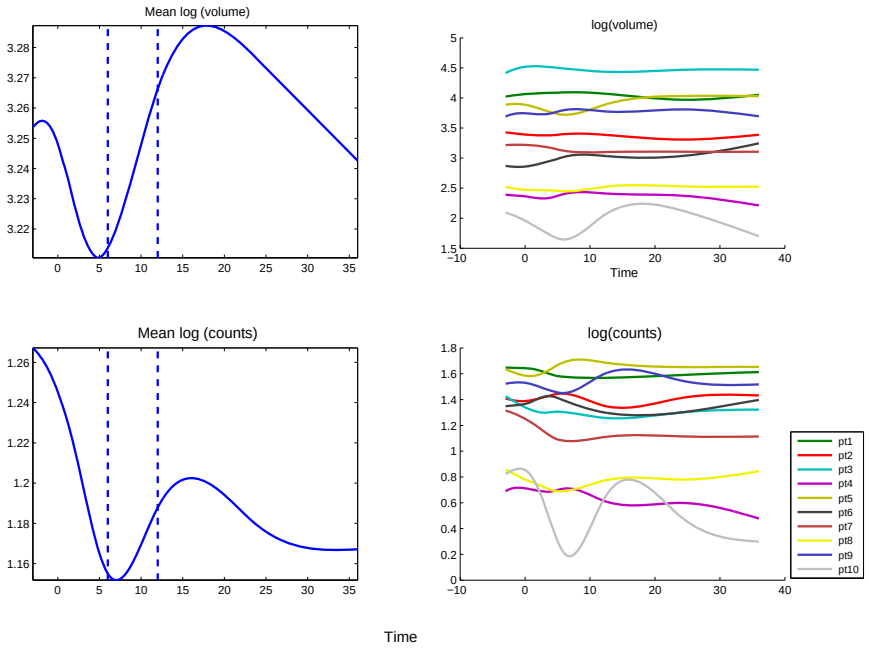

Figure 2.3.2: Mean of curves (left) and smoothed curves for each subject (right). The important phases at which the monthly follow-up ends and the annual patients monitoring starts is shown by the blue dashed lines on the mean plots.

The results of smoothing the curves are illustrated in the right panel of Figure 2.3.2. From these plots patient 10 seems to be an outlier. We carried out the analysis

of the mean functions without patient 10; It showed that the mean of volume and counts are both influenced by this patient. The heterogeneity characteristics among MS patients under the study make it hard to find any common pattern in the way the volume and the lesion counts curves change over time. Plot of the mean volume and counts functions are shown in the left panel of Figure 2.3.2. The blue dashed lines in these plots indicate the important phases at which the monthly follow-up ends and the annual monitoring starts. The individual plots of the smoothed volume and counts curves, shown in Figures 2.3.3 and 2.3.4, make it easier to observe and follow the peaks and valleys which are hard to see from the superimposed plots of the smoothed curves in Figure 2.3.2.

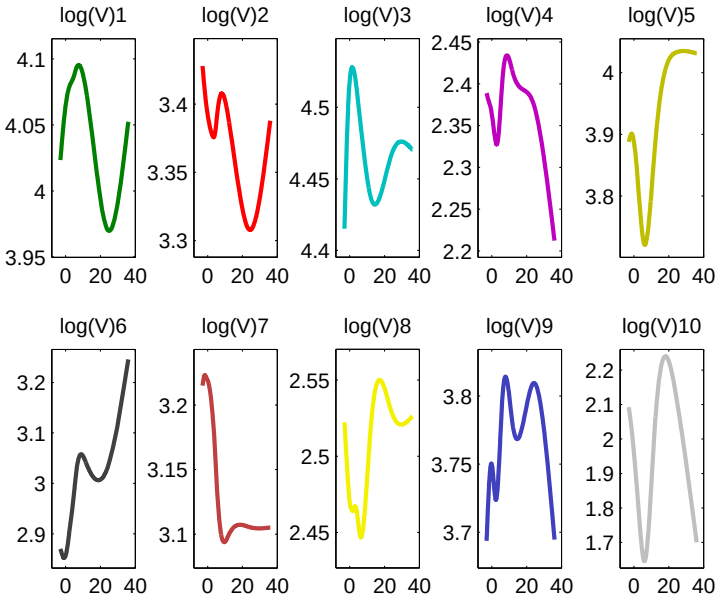

Figure 2.3.3: The log (volume) curves over time for each subject.

These individual plots help us to examine the quality of the estimated curves graphically and see whether the smoothed curves are able to successfully capture the

features appearing in the raw dataset (shown in Figure 2.3.1). Also, each individual plot shows the unique pattern over time in volume and counts of the lesions clearly. From the individual plots, each smoothed curve seems to fit the data well without any danger of overfitting. This characteristic reflects the fact that there is a compromise between the goodness of the fitted curves and their roughness in the formula (1.5.1).

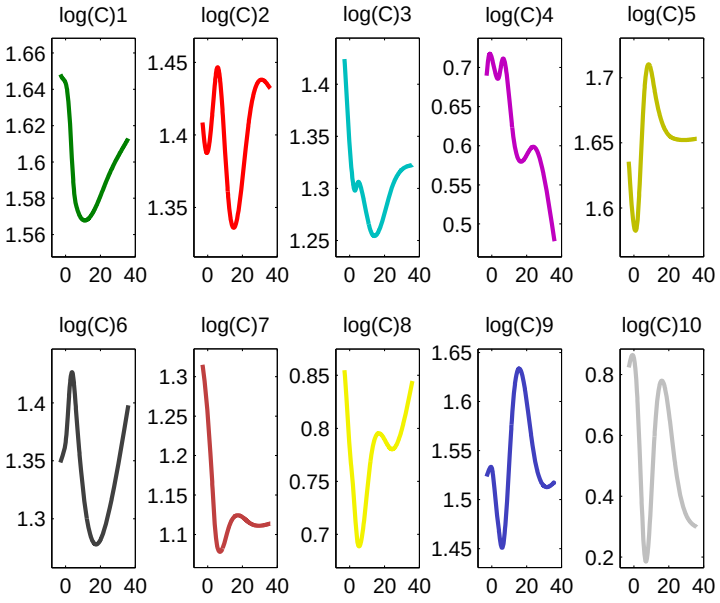

Figure 2.3.4: The log (counts) curves over time for each subject.

The plot of the functions of the transformed data, log (volume) versus log (counts) in Figure 2.3.5, shows a linear relationship between the two functions. This linearity is more pronounced in the plot of the log transformed data compared to the plot of the raw data (not shown here). This might be considered as another possible advantage of working with the log transformed data besides the ones mentioned at the beginning of this section.

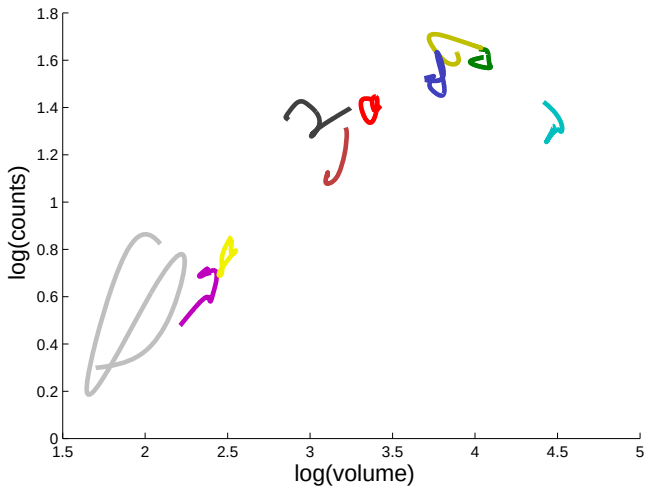

Figure 2.3.5: Smoothed curves log(volume) vs log(counts).

To interpret the individual plots of log(volume) vs log(counts) in Figure 2.3.6 we select a patient, for instance the first patient, and follow the pattern of the simultaneous changes in the observed measurements over time. Note that the times are not exact and are used as a tool to trace the approximate time at which a noticeable event happens. For the first patient, before starting the treatment (from time -3 to time 0) the number of lesion counts is at its highest, while the amount of lesions volume is at a medium level. Afterward, the counts decreases very slowly, whereas the volume remains almost the same; this pattern continues up to the end of the monthly follow-up, month 6. Then, volume decreases as counts increases at a slow rate for about a year. For the last year of the study period, both volume and counts go up sharply. At month 36, the volume is about the same as the volume at time -3, while the count is stabilized at about medium. Unfortunately, there is no information between the annual measurements to precisely explain these patterns.

Analysis of the simultaneous changes in the means of the two functions is done similarly. The result is illustrated in Figure 2.3.7, where the pattern in the mean

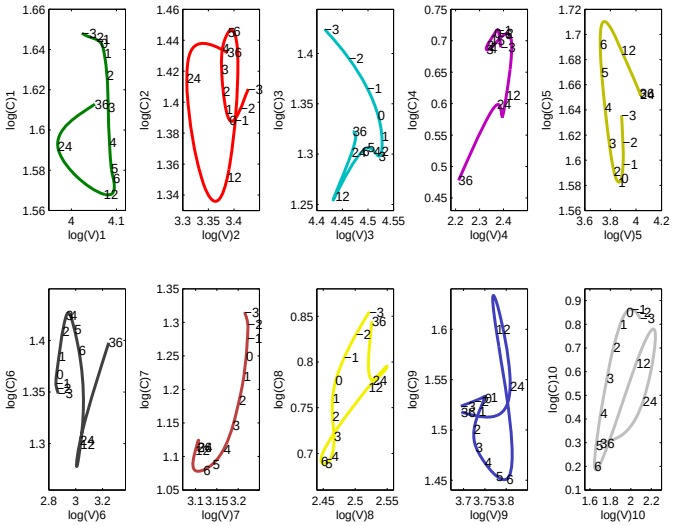

Figure 2.3.6: Individual log (volume) vs log (counts) curves. Approximate times are shown along each curve.

functions plotted against each other does not resemble any of the individual plots; this might be due to the facts that between-patients variability is high and that patients are very heterogeneous. In this plot we can see that on average, the number of lesions counts is at its highest level before the treatment starts. Both volume and counts decline till the end of the monthly treatment, then both increase until about 18 months after the treatment. For the remaining 1.5 years both of the mean functions decrease until the end of the three-year study period.

In the analysis of derivatives the important matters to consider are: a) the size of the loops or their radii; b) the location of the loops; and c) the overall shape of the loops. The plot in Figure 2.3.8 displays the dynamics in the MS lesion counts and how the function log (counts) is related to its rate of change. The x-axis shows the function log (counts) and the y-axis shows the first derivative of the function log

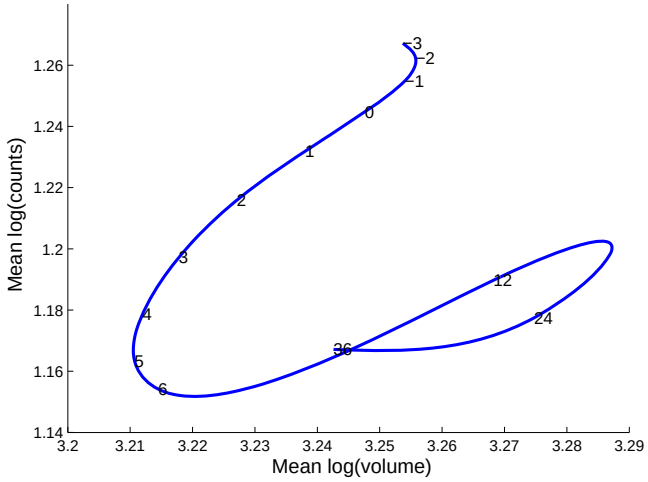

Figure 2.3.7: Plot of the mean log (volume) vs log (counts) with the time points shown along the mean curve.

(counts) denoted by Dlog (counts).

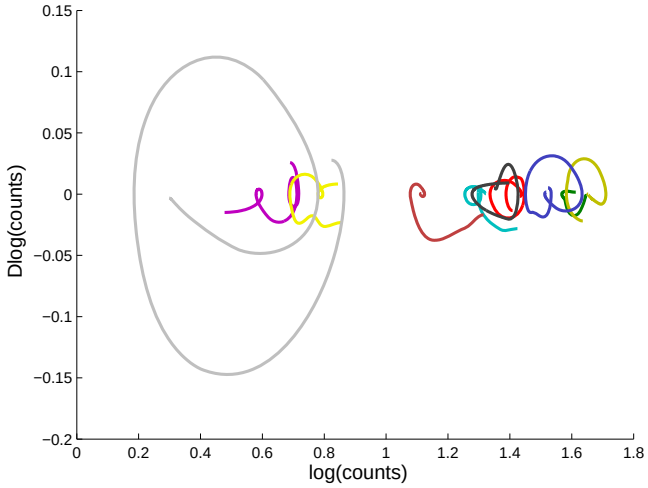

Figure 2.3.8: The phase-plane plot of log (counts).

From the superimposed plot of the function log (counts) against its derivative we can see 10 loops, one for each patient. Patient 10 in grey color has the largest loop which is associated with the most intense changes in the log (counts), whereas the first patient, color coded as green, has the smallest changes in the log (counts) as it has a tiny loop. We note that the velocity of patient 10 is larger than the other

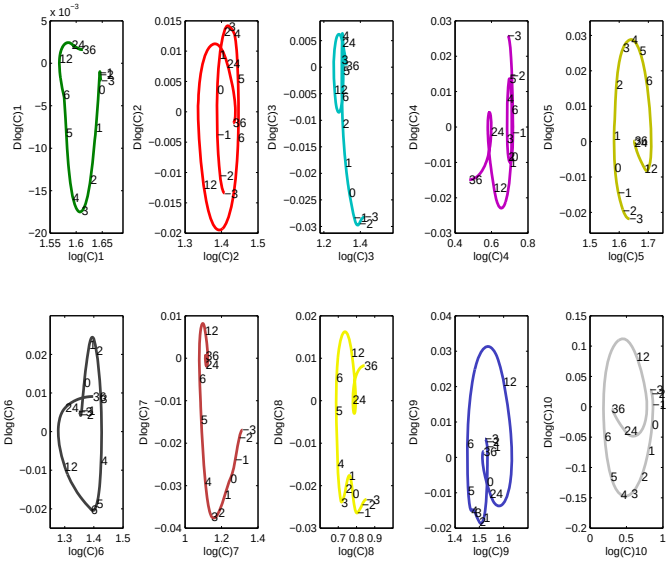

Figure 2.3.9: The individual phase-plane plot of log (counts) with approximate times shown along the curves.

patients, which is because the log (count) is smaller than the other 9 patients. If we remove this patient, we see that the intensity of change in the log (counts) is almost the same for all the remaining 9 patients and the process seems to be stable, deduced from the cyclic shape of the loops. Individual plots of the log (counts) versus Dlog (counts) have been drawn in Figure 2.3.9 for more readability.

We take a similar approach to the log (volume) function and its first derivative. Figure 2.3.10 shows the phase-plane plot of the log (volume) process. The loop with the largest radius belongs to patient 10 and it is located on the left side of the plot meaning that the most changes in the volume of MS lesions occur in this patient. The velocity of patient 10 is larger than the other 9 patients and this is due to the small log (volume) for this patient. The analysis with patient 10 removed confirms this result since all the other 9 patients have almost the same loop size. The individual

plots of the phase-plane plots of the function log (volume) in Figure 2.3.11 provide us with the details of the pattern of these changes.

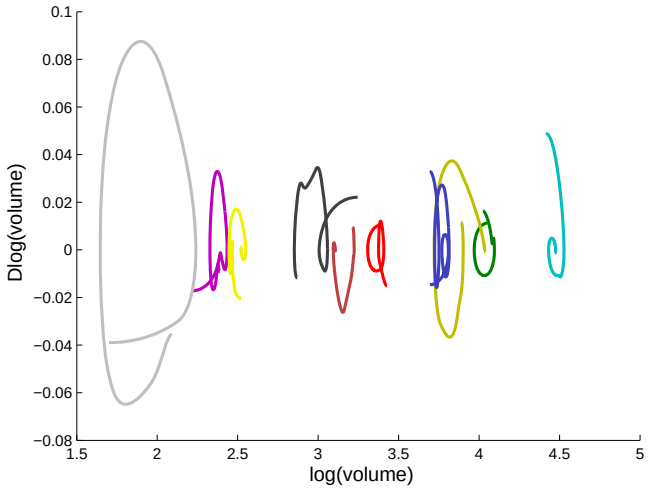

Figure 2.3.10: The phase-plane plot of log (volume).

The plot in Figure 2.3.12 right panel shows the interplay between the mean of the log (counts) function and its first derivative Dlog (counts) with times in months indicated along the curve. There are three changing points in this plot occurring at times 2-4, 6-12, and 12-24. The occurrence of these cusps is similar to what we see in the phase-plane plot of the mean of log (volume) function in Figure 2.3.12 left panel except for the first cusp which happens at 1-2 months after the treatment. From the comparison of these two phase-plane plots in Figure 2.3.12, we see that the process is more stable for the log (volume) function and it is consistent throughout the study period.

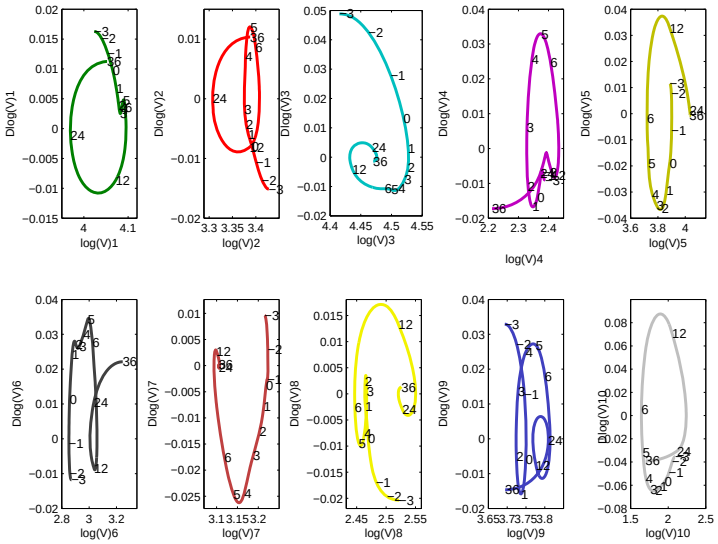

Figure 2.3.11: The individual phase-plane plot of log (volume) with the approximate times shown along the curves.

2.4 Covariance and correlation functions

The covariance function is used to investigate the dependence of a function over time. It is formulated by formula (2.4.1).

$$cov[x(t), x(s)] = \frac{1}{N-1} \sum_{i=1}^{N} [x_i(t) - \overline{x}(t)][x_i(s) - \overline{x}(s)]. \tag{2.4.1}$$

In the case of having pairs of observed functions, cross-covariance function is calculated, being defined as shown in formula (2.4.2).

$$cov[x(t), y(s)] = \frac{1}{N-1} \sum_{i=1}^{N} [x_i(t) - \overline{x}(t)][y_i(s) - \overline{y}(s)], \tag{2.4.2}$$

where the mean functions in (2.4.1) and (2.4.2) are defined as in formula (1.8.1).

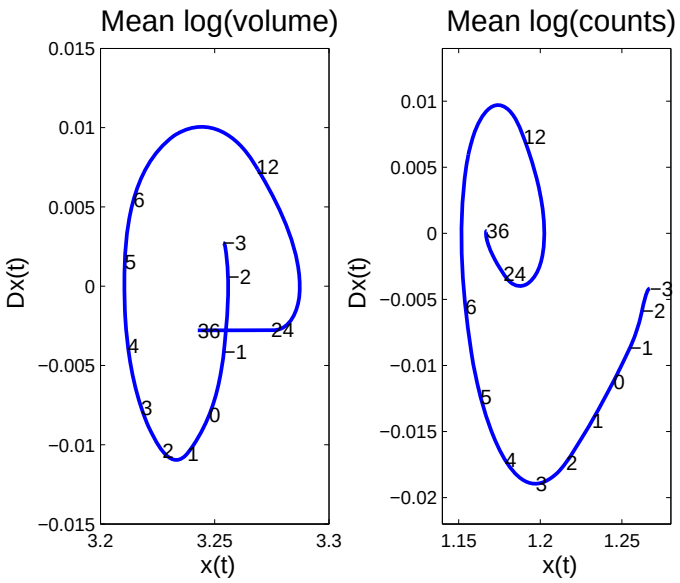

Figure 2.3.12: The phase-plane plot for the mean of log (counts) (right) and log (volume) (left) along the approximate times.

The quantities in formulas (2.4.1) and (2.4.2) are the counterparts of the descriptive statistics in functional data. The plots in Figure 2.4.1 illustrate the covariance and cross-covariance of two functions volume and counts in 2-D and 3-D images. The complete information is obtainable by reading two plots together; the surface plot shows an overall view of the way two functions covary or a single function varies over time. To detect the time points at which these features occur, the contour plot is used. Most of the variability in the function counts happens 4-6 months after the treatment. For the volume this occurrence starts at the early stage of the treatment between months 1 and 6. The intensity of each variability can be read from the colormap bar located below the pictures.

The corresponding correlation function in formula (2.4.1) is defined as

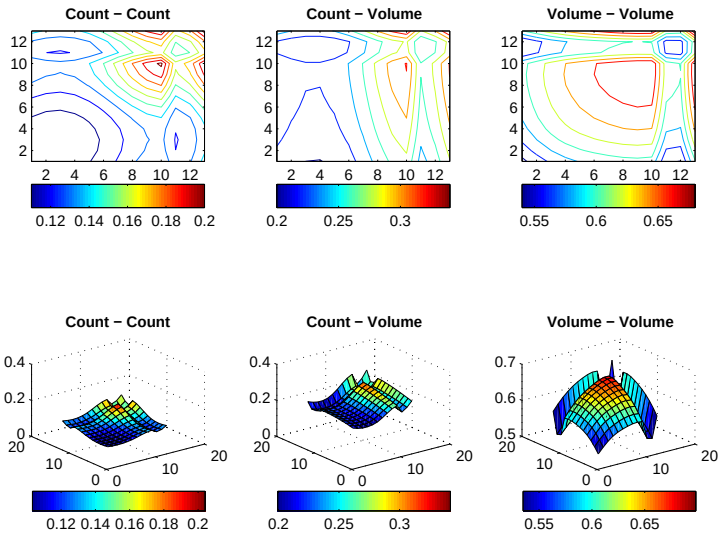

Figure 2.4.1: Variance-Covariance contour (above) and surface plots (below) of variables volume and counts.

$$corr[x(t), x(s)] = \frac{\sum_{i=1}^{N}[x_i(t) - \overline{x}(t)][x_i(s) - \overline{x}(s)]}{\sqrt{\sum_{i=1}^{N}[x_i(t) - \overline{x}(t)]^2 * \sum_{i=1}^{N}[x_i(s) - \overline{x}(s)]^2}}. \tag{2.4.3}$$

The cross-correlation function, when having pairs of observed functions, corresponding to the formula (2.4.2) is defined in formula (2.4.4):

$$corr[x(t), y(s)] = \frac{\sum_{i=1}^{N}[x_i(t) - \overline{x}(t)][y_i(s) - \overline{y}(s)]}{\sqrt{\sum_{i=1}^{N}[x_i(t) - \overline{x}(t)]^2 * \sum_{i=1}^{N}[y_i(s) - \overline{y}(s)]^2}}. \tag{2.4.4}$$

The range of correlation for each variable volume and counts over time are reported as 0.97-0.99 and 0.92-0.99, respectively. The minimum range for the volume occurs at times 6 and 24 months after the treatment and the maximum is at times one and two months before the baseline. Also, the minimum range for variable counts takes place at times a month before the treatment and 6 month after the treatment and

for the maximum range it happens at the baseline and a month before the baseline. The variables volume and counts are positively correlated and the cross-correlation for these two variables ranges between 0.82-0.91. The minimum range happens at times two months after the treatment for variable counts and at baseline for variable volume and the maximum range happens at the end of the study period, month 36.

We are interested in knowing the overall correlation between these two functions. An exploratory technique called Functional Canonical Correlation Analysis (FCCA) is used to explore this, which is the topic of the next section.

2.5 Functional Canonical Correlation Analysis (FCCA)

In this section, we analyse the correlation between the two functions volume and counts through FCCA. We know that the volume of lesions and the number of lesion counts in a specific part of the brain are correlated but we need to know more precisely what the strength of this correlation is and at what stage of the disease in MS patients the link between the two functions is stronger across 10 patients. This method is the extension of the Canonical Correlation Analysis (CCA) in classical statistics. FCCA is used to detect how variability between the functions volume and counts are correlated over time. To formulate the problem, let (X_i, Y_i) be the N pairs of observed functional variables log (volume) and log (counts), respectively for $i = 1, 2, ..., N$. Assume that these observations are available for the time points $t \in \tau$ where τ is some finite interval, over which all integrals are taken. Also, consider the centred version of these functions. The penalized squared sample correlation is then formulated as

$$
\begin{aligned}
& R^2(\boldsymbol{\xi}, \boldsymbol{\eta}) \\
= & \frac{[cov(\int \xi(t)X_i(t)dt, \int \eta(t)Y_i(t)dt)]^2}{[var(\int \xi(t)X_i(t)dt) + \lambda||D^2\boldsymbol{\xi}(t)||^2]\,[var(\int \eta(t)Y_i(t)dt) + \lambda||D^2\boldsymbol{\eta}(t)||^2]} \\
= & \frac{[\sum_{i=1}^{N}(\int \xi(t)X_i(t)dt)(\int \eta(t)Y_i(t)dt)]^2}{[\sum_{i=1}^{N}(\int \xi(t)X_i(t)dt)^2 + \lambda||D^2\boldsymbol{\xi}(t)||^2]\,[\sum_{i=1}^{N}(\int \eta(t)Y_i(t)dt)^2 + \lambda||D^2\boldsymbol{\eta}(t)||^2]},
\end{aligned}
\tag{2.5.1}
$$

where $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are called canonical variate weight functions corresponding to the functions volume and counts, respectively, $\int \xi(t)X_i(t)dt$ and $\int \eta(t)Y_i(t)dt$ are called canonical variates, and λ is a positive value chosen to regularize the variances of the canonical variates. We maximize the quantity in the formula (2.5.1) with respect to $\boldsymbol{\xi}(\mathbf{t})$ and $\boldsymbol{\eta}(\mathbf{t})$ subject to the constraint $var(\int \xi(t)X_i(t)dt) + \lambda||D^2\boldsymbol{\xi}(t)||^2 = var(\int \eta(t)Y_i(t)dt) + \lambda||D^2\boldsymbol{\eta}(t)||^2 = 1$. The weight functions $\boldsymbol{\xi}(\mathbf{t})$ and $\boldsymbol{\eta}(\mathbf{t})$ are estimated using the basis system with the $M = 20$ B-spline basis functions where 20 basis functions is considered enough to estimate the canonical variates. The choice of λ can be made either by the Cross Validation (CV) method or it can be selected subjectively. We used $\lambda = 10$ as the smoothing level for all functions $X_i(t)$, $Y_i(t)$, $\boldsymbol{\xi}(t)$ and $\boldsymbol{\eta}(t)$. There was no remarkable changes in the weight functions when we used different λ. The above procedure is called smoothed canonical correlation analysis.

The first k largest $R^2(\boldsymbol{\xi}, \boldsymbol{\eta})$ are $R_1^2 > R_2^2 > ... > R_k^2$ where k is the $min(N, M)$. Each of these R_k^2 corresponds to a pair (ξ_k, η_k). The two conditions for the weight

functions have to be satisfied: $\int \xi_1(t)\xi_2(t)dt = 0$ and $\int \eta_1(t)\eta_2(t)dt = 0$, etc. As a result, each component of variation unveils a new aspect of the correlation between the two components of variations.

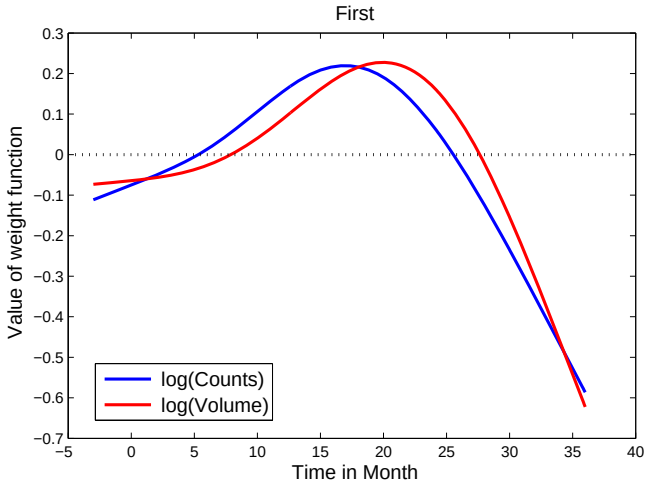

Figure 2.5.1: The weight functions over time.

The interaction between the two functions volume and counts is expressed in terms of $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$, which indicate the components of variation in two curves. The value of the weight functions $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are plotted in Figure 2.5.1. Note that the values shown in this figure are normalized, so that $\int \xi^2(t) = \int \eta^2(t) = 1$. From this plot, we see that the two functions log (volume) and log (counts) are correlated at any particular time because the weight functions are so similar. We can also see how the two functions are correlated; at the beginning of the study, log (volume) variability occurs first, however the order of this occurrence is reversed at the end of the study. The variability in counts starts earlier than volume, approximately a month after the baseline time. The extreme between the variability of the two functions $X_i(t)$ and $Y_i(t)$ takes place approximately at the end of the study period.

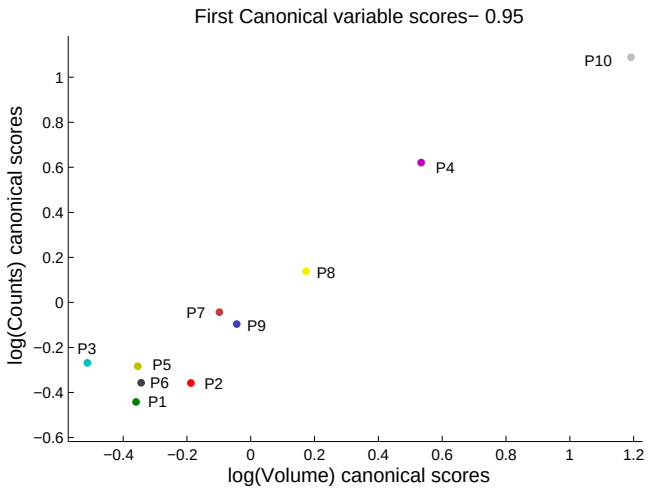

Figure 2.5.2: Plot of the canonical scores.

Figure 2.5.2 illustrates the Canonical Correlation (CC) scores, $\int \xi(t)X_i(t)dt$ and $\int \eta(t)Y_i(t)dt$. It can be used to answer the question of how strong the correlation between the two functions is. There is a nearly perfect linear correlation between the first pair of canonical variates $\int \xi_1(t)X_i(t)dt)$ and $\int \eta_1(t)Y_i(t)dt)$ with $R_1 = 0.95$. The first five canonical correlations are reported as 0.95, 0.76, 0.31, 0.16, and 0.04.

2.6 Conclusions and future study

In conclusion, the heterogeneity of the MS patients was confirmed through the phase-plane analysis and the analysis of the simultaneous plot of the two functions shown in Figure 2.3.5. We explored a correlation between the functions volume and counts of the MS lesions with the magnitude of 0.95 through FCCA.

The possibility of having the adequate dynamic models with different orders of the Ordinary Differential Equations (ODE) were also explored. The result was not satisfactory in that the model did not work well in predicting the rate of change in the

function, which was considered as a response, with the function itself as a predictor. One of the researchers' interests is to detect any differences in volume and counts of MS lesions over time. As our future work, this problem will be answered by employing two methods: paired permutation test used in functional data and a method known as change point detection.

Chapter 3: Application of FDA to animal models (MS data)

3.1 Background of the data

Forty mice are followed up for five days, day $2, 4, 6, 8$, and 10. Among these 40 mice, 10 have moderate Experimental Autoimmune Encephalomyelitis (EAE), 8 suffer from strong EAE, 9 have mild EAE, and 13 are the control group called Complete Freund's Adjuvant (CFA). An openfield test, test of anxiety and fatigue, has been performed on these 40 subjects. On each day mice are watched for 4 minutes and four measurements including activity, crossings, rearing, and grooming scores are taken on each mouse. Almost, equal activity scores for the EAE and CFA mice are expected at the beginning of the study by the researchers; as mice get sick, their scores drop down. This pattern is slightly different in crossings, rearing, and grooming scores in that the healthy mice have always higher scores than the sick ones. Activity scores are the measurements taken for each mouse out of 12, as total score. The rest of the scores count the number of times a mouse has a specific behaviour like crossings, rearing, or grooming. MS

researchers are interested in comparing onset of pain behaviour with CFA control and EAE mice. Particularly, they are interested in knowing whether the scores of activity, crossings, rearing, and grooming for EAE mice are decreased over time. Also, they want to know the day at which sick mice start to differ from the control mice. Moreover, they are interested in investigating any pain behaviour within EAE groups including mild, moderate, and strong EAE mice.

There are 19 missing values among the observed scores on day 10. Activity, rearing, and grooming scores include 5 missing values each, and crossings includes 4. The linear interpolation method, explained in Chapter 2, has been used to estimate the missing values.

As a first step in our data analysis, we visualize the raw data and detect any unusual behaviour or possible outliers. Figure 3.1.1 shows the raw data including activity, crossings, rearing, and grooming scores for all 40 mice over the course of five days. Each plot has been detected visually through individual plotting (see appendix A). From the plots, most curves differ in amplitude and there are a few variations in phase. No curve alignment technique (see Section 4.2) has been used in this analysis.

3.2 Converting raw data to smoothed curves

We model discrete data as functions or curves that are assumed to give rise to discrete points using the spline smoothing method. We smoothed each dataset, activity, crossings, rearing, and grooming separately since we analysed each dataset independently. Seven basis functions of the B-spline basis system of order 4, cubic B-spline

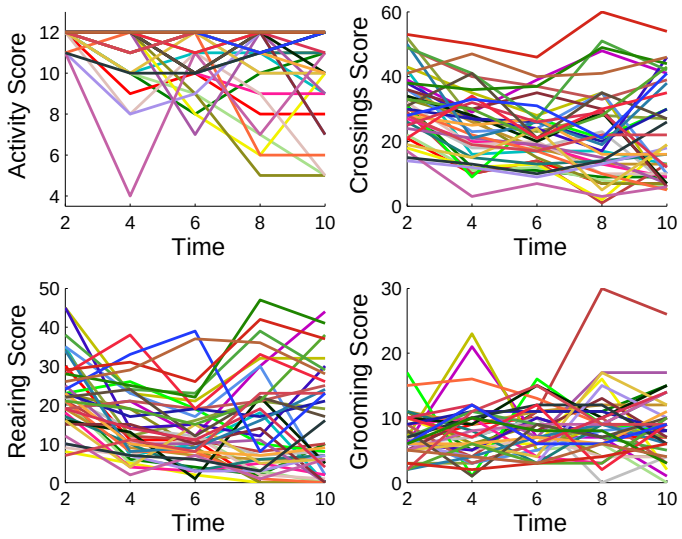

Figure 3.1.1: Scores over time of each mouse. The lines between observations are drawn to help visualization.

basis functions which provides the optimum result, with knots placed at each time point is used and we penalized the first order derivative.

The choice of B-spline basis among all other types of basis functions is made based on the non-periodic behaviour of the data. The level of smoothness is the only entity chosen differently for each variable– 0.05 for activity and grooming, 0.005 for crossings, and 0.2 for rearing. The GCV method, explained in Chapter 1, reported different smoothing parameters which are higher than what was used in our smoothing procedure. The smoothing parameters chosen by the GCV method for activity, crossings, rearing, and grooming are 3.16, 1, 1, and 10, respectively. This difference in reported λ based on GCV might be due to the violation of the independence of data (see Section 2.3 of Chapter 1). Figure 3.2.1 shows the result of smoothing each set of raw data corresponding to the plots in Figure 3.1.1.

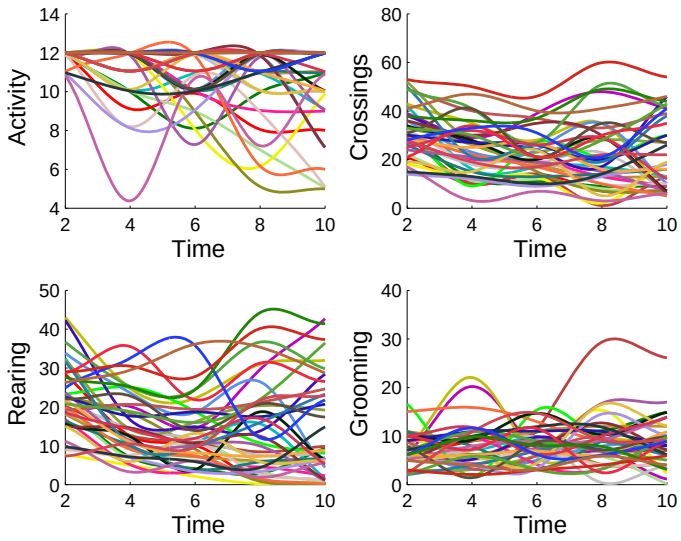

Figure 3.2.1: Smoothed curves for four variables activity, crossings, rearing, and grooming.

3.3 Functional T-test

Functional T-test is an extension of classical T-test, where the t-statistic is a function of time. The t-statistic is then the maximum of all values of $T(t)$. The formula is denoted by:

$$T(t) = \frac{|\overline{x}_1(t) - \overline{x}_2(t)|}{\sqrt{\frac{1}{n_1} var[x_1(t)] + \frac{1}{n_2} var[x_2(t)]}}, \tag{3.3.1}$$

with the $\overline{x}(t)$ and $var[x(t)]$ introduced as in Chapter 1. Functional T-test is used to compare the means of two groups over time. To conduct this, we use a pointwise 0.05 critical value using the permutation test and the maximum 0.05 critical value as reference lines. Then we compare them with the result from the observed t-statisitc. The procedure for permutation test to obtain the t-statistic for observed values, the pointwise 0.05 critical value, and the maximum 0.05 critical value is as follows:

1. Let n and d denote the number of samples per curve and number of permutation, respectively;

2. rearrange the labels of the curves $d = 1000$ times and calculate the $T(t)$ at $n = 101$ time points every time, the result is a $n \times d$ matrix and is called $T_{nullvalues}(t)$;

3. record the maximum of $T_{nullvalues}(t)$ over 101 time points for each permutation and the result is a column vector of length 1000; This forms the null distribution and is denoted by $T_{null}(t)$.

4. calculate $T(t)$ in formula (3.3.1) for the data in hand (the original data without shuffling) at 101 time points; this forms the observed curve; call it $T_{obs}(t)$ (shown as a blue solid line in Figure 3.4.1, bottom right), then call the maximum of these values $T_{\max obs}(t)$;

5. the pointwise critical value is the curve formed by the quantile $(T_{nullvalues}(t), 0.05)$ over all permutations at each time point and this is a column vector of length 101; it is shown as a blue dashed line in Figure 3.4.1, bottom right.

6. the maximum critical value over time is a constant and it is simply quantile $(T_{null}(t), 0.05)$; it is shown in Figure 3.4.1, bottom right as a red dashed line.

7. compare the observed $T_{obs}(t)$ curve with the pointwise and maximum critical values at each time point;

8. calculate the average time that $T_{\max obs}(t)$ appears to be smaller than $T_{null}(t)$ and this is the p-value at the time point where the observed curve reaches its

maximum, p-value= $mean[T_{\max obs}(t) < T_{null}(t)]$; if p-value$< 0.05$, we reject the null hypothesis of $H_0 : \mu_1(t) - \mu_2(t) = 0$ at the significance level of 5% for the two-sided hypothesis of $H_a : \mu_1(t) - \mu_2(t) \neq 0$ for at least one t;

9. to calculate the p-value at any given time point we need to calculate the average time that $T_{obs}(t)$ at that given time t, is smaller than $T_{null}(t)$; the rest is similar to what was explained in step 8.

3.4 Analysis of activity scores

The top left panel in Figure 3.4.1 illustrates the smoothed curves for activity scores. The smoothed curves are color coded as blue for 27 EAE mice and red for 13 CFA mice. The top right plot in this figure illustrates the mean curve of each group.

From the mean plot we see that on average, the control group has about equal scores as the EAE mice up to around day 4 or 5, but the order is reversed after that; we need to test whether this finding is statistically verifiable. The functional T-test plays an important role to answer this question. The null and alternative hypotheses are denoted as $H_0 : \mu_{EAE}(t) - \mu_{CFA}(t) = 0$ vs $H_a : \mu_{EAE}(t) - \mu_{CFA}(t) \neq 0$ for at least one t. From Figure 3.4.1 lower right panel, we see that there is no evidence of any difference between the means of two groups up to around day 9. After day 9 there is a statistically significant difference between the healthy and the sick mice; particularly, the mean activity scores of CFA are larger than EAE group and this behaviour is aligned with what was expected by the researchers. The p-value at this point is reported as 0.05.

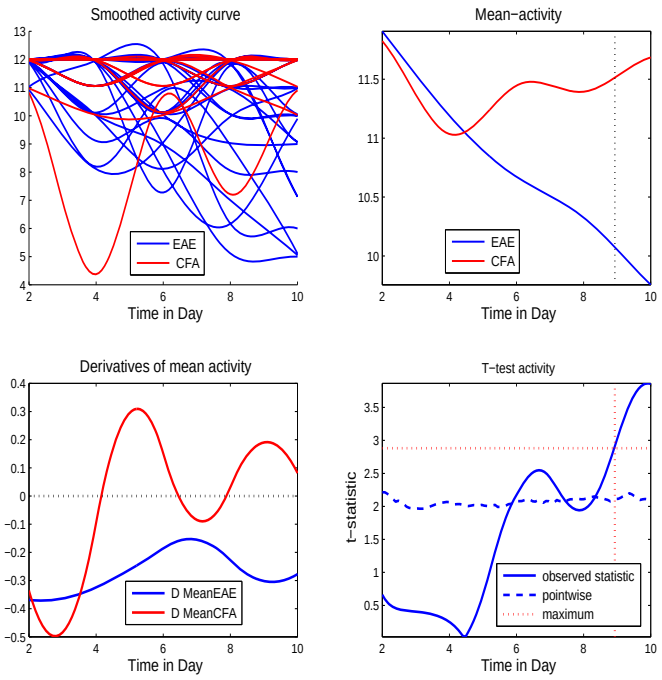

Figure 3.4.1: Plot of the smoothed activity curves by group (top left), means of each group (top right), derivatives of the mean of groups (bottom left), and permutation test for equality of the EAE and CFA groups (bottom right).

The derivative of the means of functions EAE and CFA over time has been illustrated in the lower left panel of Figure 3.4.1. By looking at the zero crossings it seems that on average, the control group has more fluctuations than the EAE group. There is a slow change in the EAE group throughout the study period with a slight jump in the trend at around day 7.

Next, we analyze phase-plane plots which are informative in the sense that we will be looking at the rate of change in the function as function itself changes over time. The means of EAE and CFA curves are plotted along with the traceable time points in Figure 3.4.2.

The cyclic behaviour of the phase-plane plot of the CFA group implies a relatively stable process. The sick animals in the EAE group tend to have a more unstable

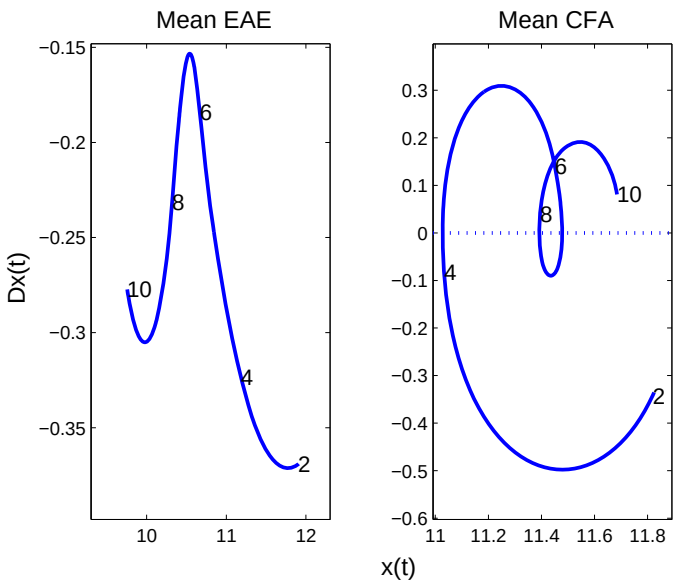

Figure 3.4.2: Phase-plane plot, the activity scores vs first derivative of the activity scores.

process. For the CFA group, there are 4 cusps or changing points that happen in between the sampling points approximately at days 3, 5, 7, and 9. Also, the process for this group starts with a negative rate of change and it becomes positive after day 4 and remains almost the same, circling around zero. For the EAE group however, the rate of change is always negative. Before day 7, the higher the value of function EAE, the smaller the rate of change in the process. The process behaves in the opposite manner after around day 7.

3.5 Analysis of crossings scores

Figure 3.5.1 top left shows the smoothed curves by groups. We investigate any differences in the pain behaviour of the EAE and the control mice and detect the points at which the differences between two groups occur. On average, the CFA mice obtain higher scores than the sick mice, as expected, but do the data support this claim?

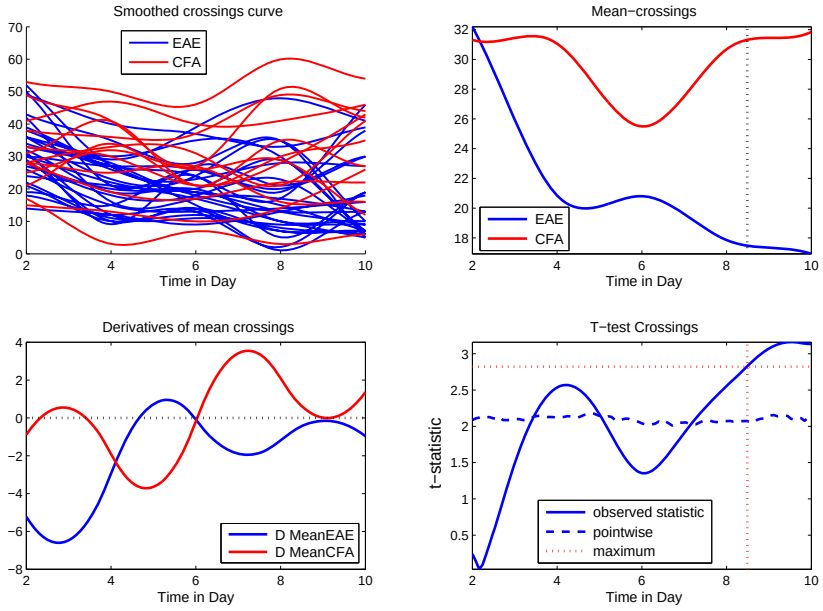

Figure 3.5.1: Plot of the smoothed crossings curves by group (top left), means of each group (top right), derivatives of the mean of groups (bottom left), and permutation test for equality of the EAE and CFA groups (bottom right).

The result of the functional T-test is used and it is shown in Figure 3.5.1 bottom right panel with the null and alternative hypotheses as defined in Section 3.4. The difference between two groups, if considering the pointwise critical value as a reference, appears somewhere at the beginning of the study interval between the days 3-5 but if we consider the more conservative reference line, the maximum 0.05 critical value, more strong evidence appears between the days 8 and 9. Consider two plots, top right and bottom right together; we see that the mean crossings scores of the CFA group are larger than the EAE group.

In Figure 3.5.1 bottom left, both EAE and CFA groups oscillate during the entire study period. The pattern in this plot clearly shows that the two groups vary in the opposite order- whenever the mean rate of change in the control group is higher, it appears to be lower in the EAE group and vice-versa. Detecting any changes in the

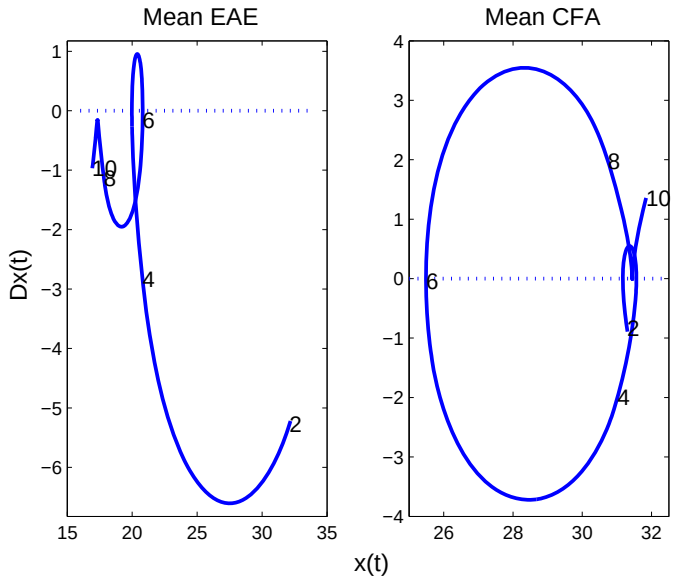

Figure 3.5.2: Phase-plane plot, the crossings scores vs first derivative of the crossings scores.

process while looking at the rate of these changes can be done through the invaluable phase-plane plots in Figure 3.5.2. The stability of the control mice is confirmed by the almost perfect cyclic shape of the plot in this figure. The beginning and the end of the process are broadly consistent with the rest of it. In EAE group we see three cusps happening at approximately day 3, day 5, and around day 7, with more changes happening in the middle of the study period. Behaviour of the change in the process, $Dx(t)$, over time t (Figure 3.5.1 bottom left) is consistent with its behaviour captured in Figure 3.5.2, $Dx(t)$ over $x(t)$ for the two groups.

3.6 Analysis of rearing scores

The upper left panel in Figure 3.6.1 indicates the smoothed curves by EAE and CFA groups. The mean scores of 27 EAE mice and 13 CFA mice are shown in the upper right panel. As it is clear from this plot, the average scores of CFA group are always

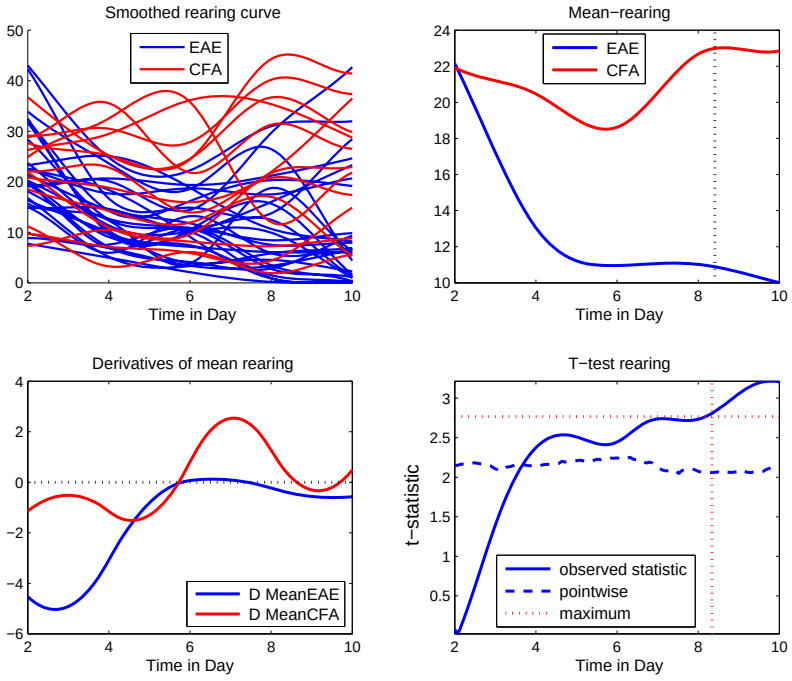

Figure 3.6.1: Plot of the smoothed rearing curves by group (top left), means of each group (top right), derivatives of the mean of groups (bottom left), and permutation test for equality of the EAE and CFA groups (bottom right).

higher than EAE group, as expected by the investigators.

To test if the higher scores for CFA group compared to EAE group is supported by the data, we use the functional T-test. Figure 3.6.1 lower right plot bears out the result of this test over time. The differences between the average of two functions happens after day 8. Whether the scores for EAE group decrease after this time can be seen by considering both top and bottom right plots in Figure 3.6.1. We see that the mean rearing scores are higher for CFA group. The p-value at the time when this difference starts is reported to be 0.05, which confirms the above mentioned result.

From Figure 3.6.1 lower left plot, it seems that the mean of the control group has more variability compared to the EAE group. Comparing this plot with the plot

of the mean of EAE and CFA groups, in Figure 3.6.1 top right, we can see how well we estimated the rearing curves. Ramsay and Silverman (2005) point out that an efficient way of checking the goodness of estimated curves is to visually test if the first or second order of its derivative acts reasonably well. Here, by testing the behaviour of derivatives of a curve we mean checking whether the same behaviour of the mean of the curve can be observed in its first derivative. For the EAE group in Figure 3.6.1 top right, for example, we have a slow decrease in the rearing scores up to day 6; afterward, the process seems to achieve scores which are slightly below 12 and stays there till the end of the study. The derivative of this mean curve indicates the same behaviour but in terms of the rate of change in the process. This means that the initial decrease up to time 6 and the subsequent steady process are all captured by the velocity curve. Similarly, for CFA group there are three phases in the mean CFA curve: a) a negative rate of change or a decrease in the mean curve up to time 6; b) an increase in the mean curve until slightly after day 8; c) a steady mean curve till the end of the study. The derivative of CFA curve reflects all the above mentioned patterns, which confirms a good approximation of the mean curves.

The phase-plane plot in Figure 3.6.2 right panel reflects the consistent behaviour of the start and the end with the rest of the process for the control group. Two cusps at days 4-6 and days 6-8 reflect the change in the process at these time points, also readable from the plot of velocity over time. The left panel in Figure 3.6.2 shows a somewhat linear relationship between the curve itself and its rate of change; the smaller the value of the mean EAE curve is, the larger the rate of change in this curve

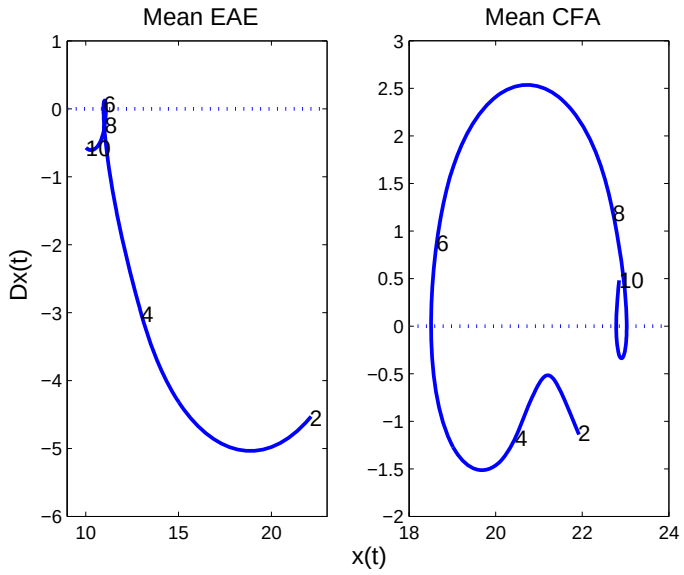

Figure 3.6.2: Phase-plane plot, the rearing scores vs first derivative of the rearing scores.

will be. In other words, sicker mice have higher rates of change in the process.

3.7 Analysis of grooming scores

Plotting the smoothed grooming curve in Figure 3.7.1 reveals some hidden features of this variable. Some extreme curves in the EAE group can be observed that might act as outliers. More investigation needs to be done to explore the characteristics of these specific subjects 4, 5, 7, and 19 among the sick mice. Investigators expect to have a higher scores for the CFA mice; however, this happens only between the days 4-6. This might be evidence showing that variable grooming is not a good indicator to detect the difference between the EAE and control subjects.

Also, no evidence of any difference between the pain behaviour of EAE mice and the control mice can be detected from the T-test plot in Figure 3.7.1, since the

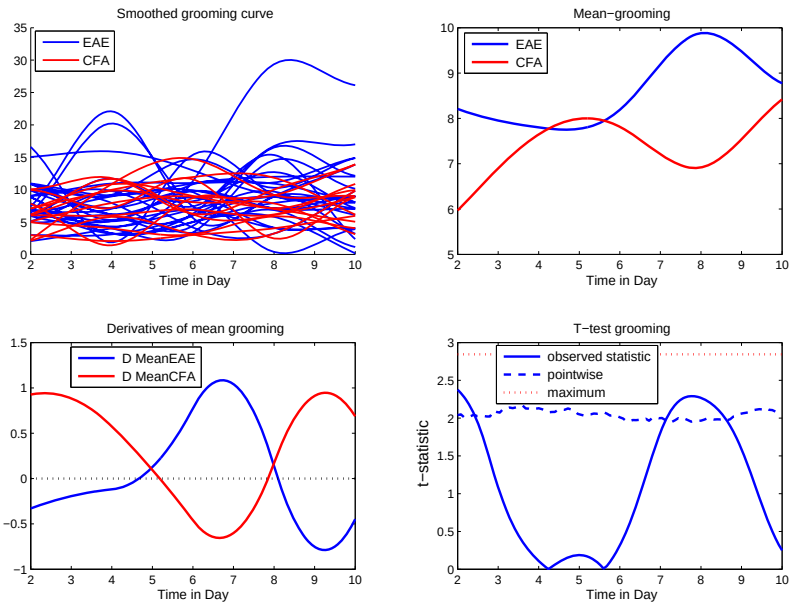

Figure 3.7.1: Plot of the smoothed grooming curves by group (top left), means of each group (top right), derivatives of the mean of groups (bottom left), and permutation test for equality of the EAE and CFA groups (bottom right).

observed line does not cross the maximum 5% point-wise reference line at any point. The velocity curve over time in Figure 3.7.1 illustrates the fluctuations around zero line, which happen in opposite order for two groups. That is to say, the higher rate of change in EAE group is associated with the lower rate of change in CFA group at any given time point, and this is quite conspicuous throughout the study period.

The reverse order of the rate of change in two groups is also reflected in the phase-plane plot shown in Figure 3.7.2. The plots show the interplay between $x(t)$ and $Dx(t)$ which can be traced by the times indicated along the curves. The existence of the same stability for two processes is implied but in a different arrangement; in the mean of the control group, the process starts with a positive rate of change and it decreases slowly up to the days 6-8, and an increase in the process is evident until the end of the process. The EAE group has a negative rate of change and it increases

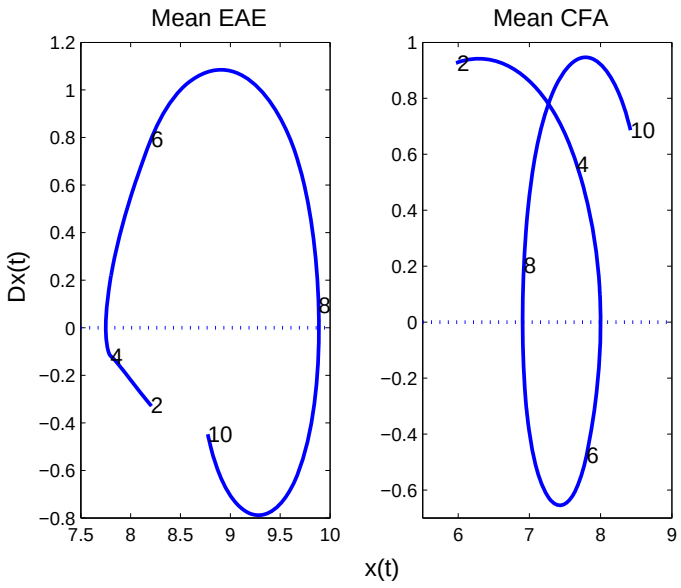

Figure 3.7.2: Phase-plane plot, the grooming scores vs first derivative of the grooming scores.

up to days 6-8, followed by a decrease till the end of the study period.

3.8 Conclusions and future studies

To sum up, we put all the results extracted from Sections 3.4-3.7, for 4 variables, activity, crossings, rearing, and grooming. First, comparisons between the healthy and sick mice for activity, crossings, rearing, and grooming scores was done. As a result, the time at which the two groups start to differ was found to be between the days 8 and 9 except for the grooming score where no difference has been observed. This answers the MS researchers' question about the day when the difference between the two groups emerges. Second, in all cases, the stability of the control group and the non-stable process of the EAE group have been confirmed through the phase-plane plots. Finally, more investigation may be needed to confirm the validity of the

grooming scores as indication of difference between the EAE and CFA groups since the behaviour of the grooming scores is not aligned with what is expected by the researchers.

We will explore the pain behaviour among the EAE mice including mild EAE, moderate EAE, and strong EAE, which is one of the MS researchers' interests. Functional Analysis of Variance (FANOVA) will be employed to serve this purpose. FANOVA is particularly useful when there is a comparison among more than two groups. This method uses the same underlying rationale as in the F-test in conventional statistics.

Chapter 4: Application of FDA to Cervical Vertebrae data

4.1 Background of the data

This study was conducted by orthodontists in the department of dentistry at the University of Alberta with the approval of Human Ethics Research Board. In this experiment, 62 adolescents between the ages of 11-17 have been followed up for one and a half years, among which 19 were randomly assigned to the control group and 43 were assigned to the treatment group. Two measurements have been taken for each patient, at times prior to the treatment ($T1$) and after the treatment ($T2$). Time $T2$ was different for each patient and the time interval between $T1$ and $T2$ ranged from 9.8 to 16.5 months. Variables gender, Hand Wrist (HW) skeletal maturation scores, and lateral Cephalogram (Ceph) skeletal maturation scores for each patient at two time points $T1$ and $T2$ are available.

The Cervical Vertebral Maturation Staging (CVMS) method is used to score the lateral cephalometric radiographs of each patient and these Cervical Vertebrae (CV)

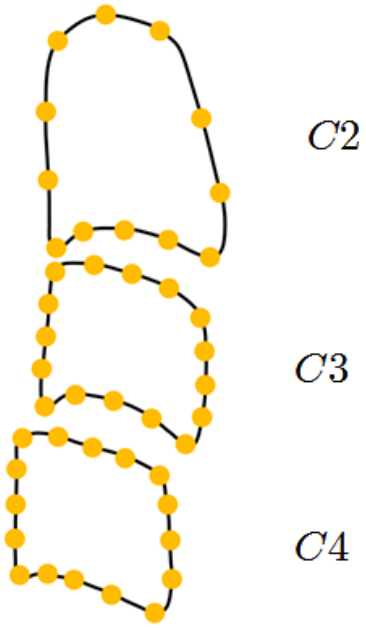

Figure 4.1.1: Outlines of Cervical Vertebrae.

scores indicate the Skeletal Maturation Levels (SML) which are used to determine the pattern of the development and growth of the adolescents. This information is in turn used by orthodontists to determine the optimal treatment time (Shim, Bogowicz, Heo, and Lagravère, 2011).

It is of interest to find out if FDA methods can distinguish the features of the CV corresponding to the SML. The aim is to explore any changes in the shape of CV curves at times $T1$ and $T2$ using techniques in FDA.

The outlines of the shapes of three cervical vertebrae, which we call C2, C3, and C4 curves, are sketched by an orthodontist. Then, 12, 16, and 16 landmarks are positioned along the outline of C2, C3, and C4, respectively. Figure 4.1.1 shows the result. The coordinates of each landmark are extracted using software called Avizo (Visualization Science Group, United States). We import these coordinates to

a Matlab program to reconstruct the curves in a two dimensional space. Only the shapes of C3 curves at times $T1$ and $T2$ are analysed.

4.2 Converting raw data to smoothed curves

A plot of the raw data is illustrated in the Figure 4.2.1. The curves are misaligned and they obscure the view. There are both vertical and horizontal mismatches which are corresponding to the amplitude and phase variation, respectively. Before carrying out any analysis, we need to use a curve alignment technique or registration method to match the curves. There are several methods for curve registration including Generalized Procrustes Analysis (GPA), Landmark Registration (LMR), and Continuous Registration (CR). We used GPA method to algin the loops. This method works based on the translating, scaling, and rotating of the raw data.

To introduce this method, let $X = (x_1, x_2, ..., x_n)^T$ and $Y = (y_1, y_2, ..., y_n)^T$ be the coordinates of the n landmarks in a 2-dimensional space and form a $n \times 2$ matrix $Z = (X, Y)$ which includes these coordinates. Further, assume that the mean of coordinates are calculated as $\overline{x} = \frac{x_1+x_2+...+x_n}{n}$ and $\overline{y} = \frac{y_1+y_2+...+y_n}{n}$. The procedure of superimposing the curves is as follows:

1. translating: calculate $(x_i - \overline{x},\ y_i - \overline{y})$ for $i = 1, 2, ..., n$; this forms the centered coordinates of landmarks denoted by Z_c.

2. scaling: calculate $\frac{Z_c}{||Z_c||} = Z_n$ where $||Z_c|| = \sqrt{\sum_{i=1}^{n}[(x_i - \overline{x})^2 + (y_i - \overline{y})^2]}$;

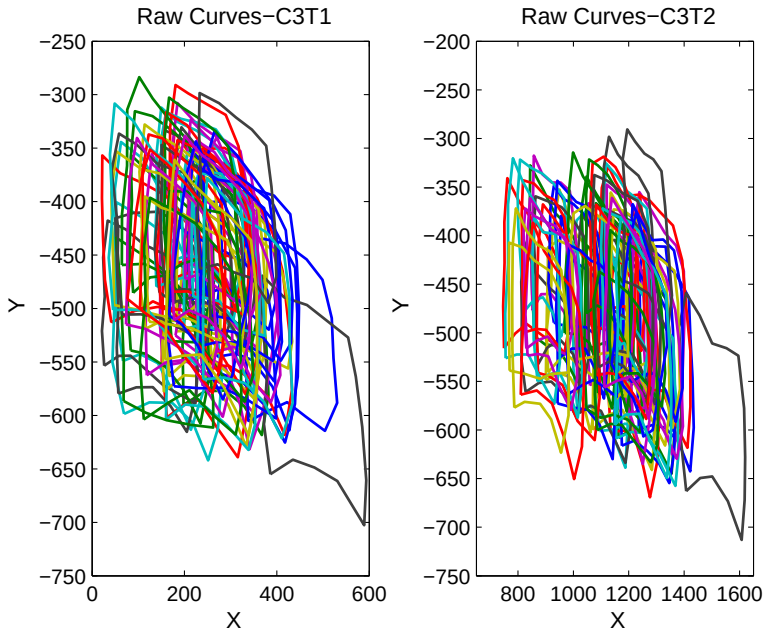

Figure 4.2.1: The Landmarks of C3 at times T1 and T2. Landmarks are connected with the solid lines to give a better visualizations.

3. rotating: rotation is done by minimizing the criterion $||Z_n Q - \overline{Z}_n||$ where Q is the orthogonal rotation matrix that aligns matrix Z_n to the average matrix $\overline{Z}_n$.

For more details on GPA see Dryden and Mardia (1998). The results of applying the GPA method to C3 landmark data for all 62 curves at times $T1$ and $T2$ are shown in Figure 4.2.2.

In the study of the CV curves the data have a non-periodic nature, thus the B-spline basis system is considered. We used a 15 B-spline basis functions of order 5 with 10 equally spaced knots and penalized second derivative. We model the data as in formula (1.2.4) where $n = 17$ is the number of landmarks per curve and $N = 62$ is the number of curves. It is worth mentioning that the time points in CV data correspond to the spatial locations of the X-Y coordinates. The results of smoothing the 2-D curves at times $T1$ and $T2$ are shown in Figure 4.2.3.

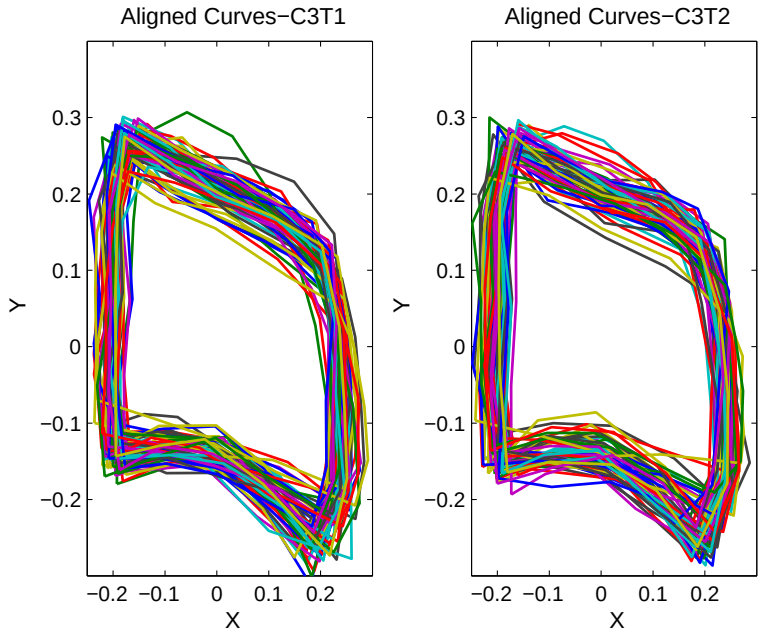

Figure 4.2.2: Aligned C3 curves at times T1 and T2.

Figure 4.2.4 shows the value of λ which minimizes the GCV criterion. The smoothed parameters are reported as 0.0562 and 0.0398 at times $T1$ and $T2$, respectively. Choosing the appropriate level of smoothness plays an important role in smoothing C3 curves. Figure 4.2.5 is an illustration of C3 curves at time $T1$ which are oversmoothed (left) and undersmoothed (right) using the extreme values of λ.

4.3 Descriptive analysis

We are interested in knowing whether there are any changes in the shape of the body of CV curves over time. To answer this question we use the information from Figure 4.3.1 and compare the means of two smoothed curves at times $T1$ and $T2$. On average, there is a slight change in the shape of C3 bone mostly at the corner of the concavity of lower border and the top border of the bone. As expected by the orthodontists,

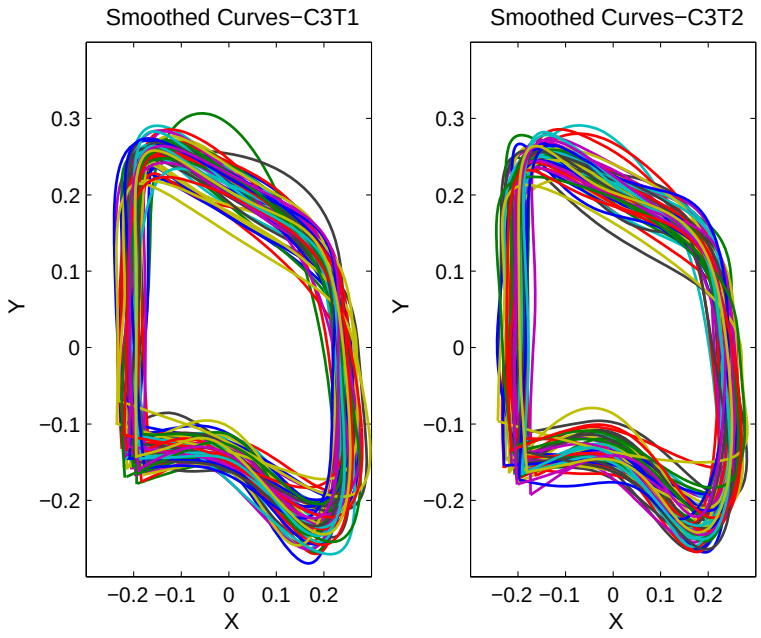

Figure 4.2.3: Smoothed curves of C3 at times T1 and T2.

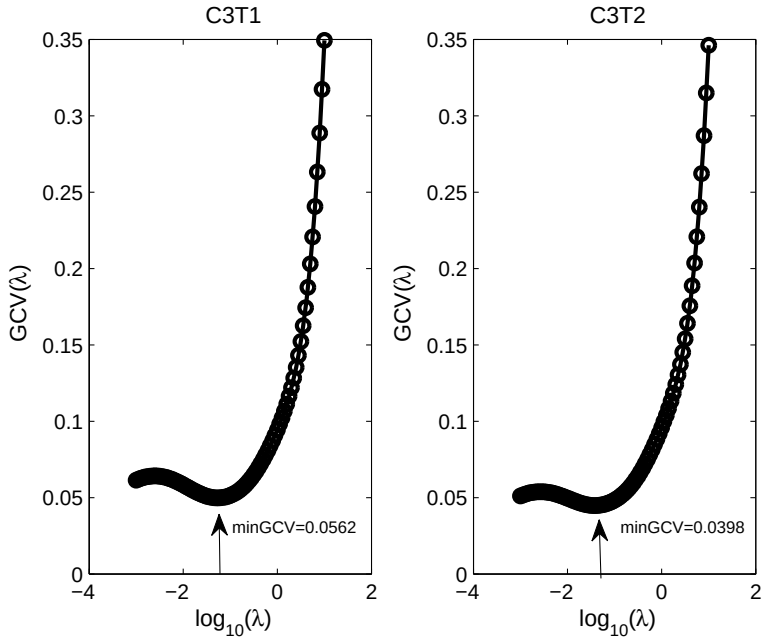

Figure 4.2.4: Smoothing parameter that minimizes GCV at times T1 and T2.

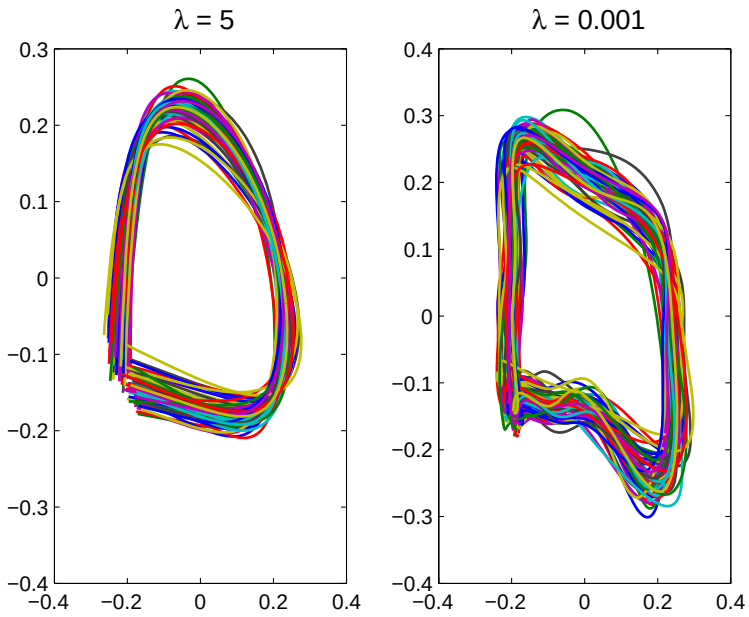

Figure 4.2.5: Oversmoothed C3 curves (left) and undersmoothed curves (right) at time T1.

the rectangular shape of the bone becomes slightly taller vertically than its original shape. Note that this change cannot be solely attributed to the treatment effect since this is the overall mean of both control and treatment groups.

The first row in Figure 4.3.2 is a clear illustration of the rate of change (velocity) in the mean of X and Y coordinates which can be associated with the width and height of C3, respectively. On average, there is not much change in the width of this bone over time, whereas the rate of change in the height of C3 is more pronounced, particularly at the right side of the upper convexity and at the lower corner of the concavity of C3. Figure (4.3.1) aids to more easily spot these areas.

Similarly, we can detect the rate of change in the velocity curves by looking at the plots of the accelerations over time at different spatial locations (the second row in

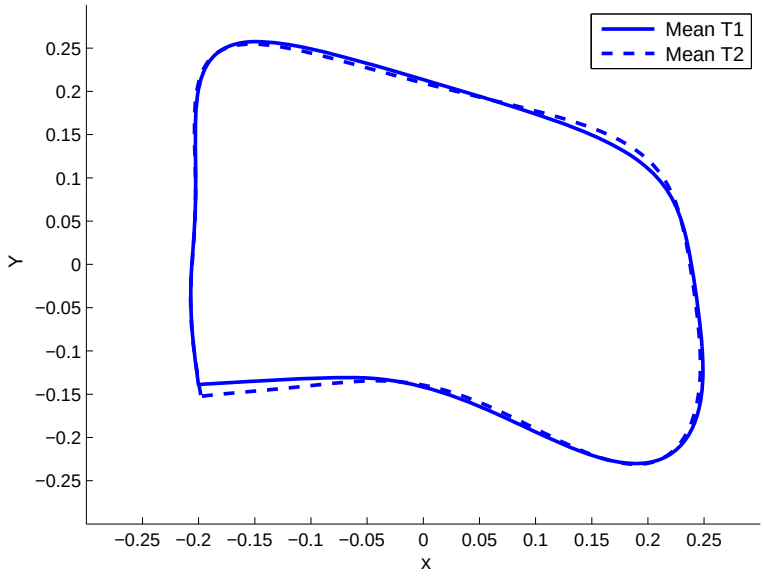

Figure 4.3.1: Mean of C3 at times T1 (solid) and T2 (dashed).

Figure 4.3.2). These plots show the magnitude at which the velocity curves change along with the direction of these changes. Several zero crossings imply the spatial locations at which there are changes in the velocity curves. We can also see where these changes are extreme from the maximum and minimum of the curves. The highest change in the height of C3 happens around the location of the 5^{th} landmark at the concavity of C3, while the maximum change in the width of C3 occurs at the upper border of the bone. Overall, the rate of change in the height of C3 bone seems higher than that in its width.

The phase-plane plot may be more revealing in detecting the changes of the shape of the mean C3 curve. The results of the first and second derivatives, at times $T1$ and $T2$ are shown for each mean curve X and Y in Figure 4.3.3. The cyclic behaviour of each plot illustrates a stable process. There is a small change in the pattern of Y at time $T2$ compared to time $T1$, which happens nearly at the beginning of the

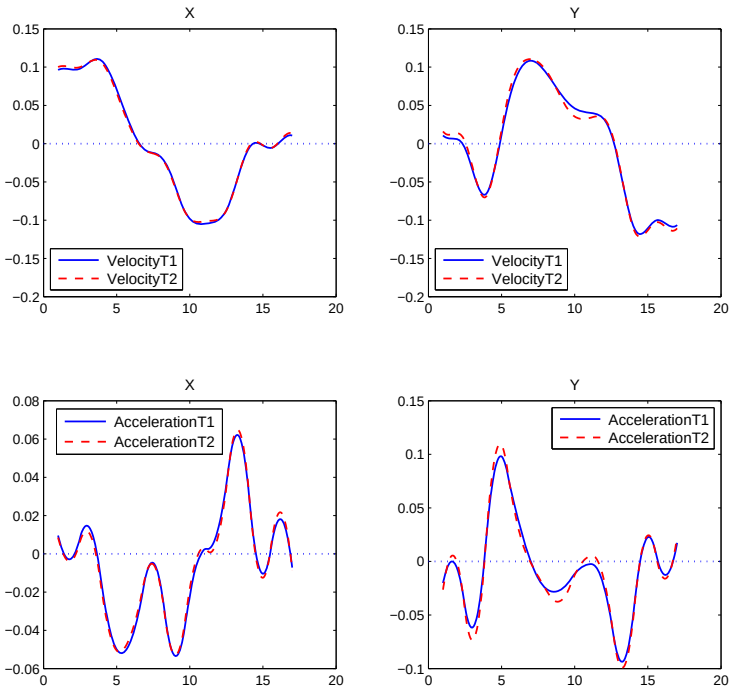

Figure 4.3.2: Derivatives of the mean of X and Y coordinates functions over time at T1 and T2.

process and is associated with the change in the left corner of the lower concavity of C3 in Figure 4.3.3. There are approximately 7 to 8 changing points in the phase-plane plots. The phase-plane plots confirm the result of having more changes in the height of the bone than in its width.

4.4 Functional Principal Component Analysis (FPCA)

A useful data reduction method, called Functional Principal Component Analysis (FPCA), is used to explore and distinguish any components of variations in the data. The idea is similar to what we have in the multivariate Principal Component Analysis (PCA) except that the weight vector and the data vector are now functions. The

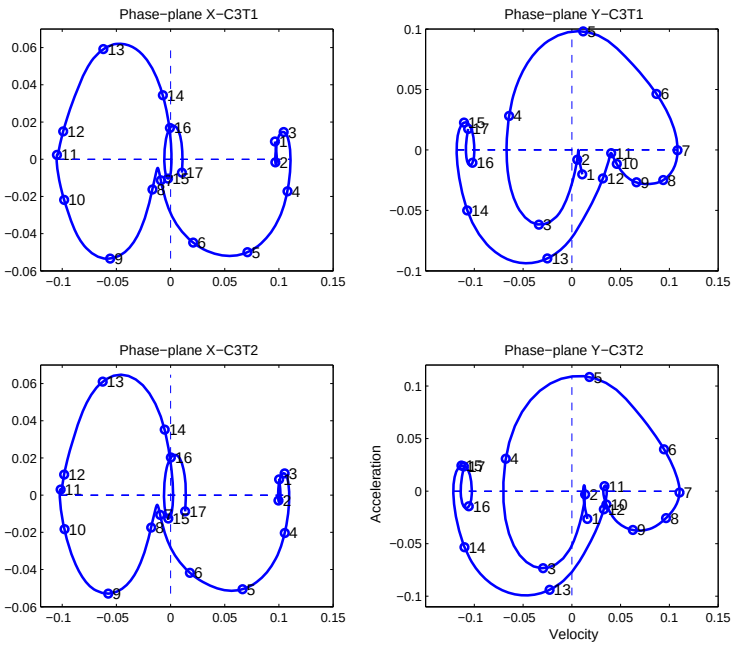

Figure 4.3.3: Phase-plane plots of the mean of X and Y coordinate functions at times T1 and T2.

method rotates and projects the function values through taking a linear combination of them such that the maximum variability in the function values is visible. FPCA method is a more informative way of looking at the variability structure in the variance-covariance function, which is hard to interpret. To define FPCA we need to introduce some notations.

Let $x_i(t)$ be the i^{th} smoothed function estimated by the discrete data with mean functions removed. Further, assume that $\boldsymbol{\xi}(t)$ is the m-vector of the weight functions where m is the number of the first most important principal components. We take a weighted linear combination of the estimated function. This linear combination of the function $x_i(t)$ is defined to be the inner product of the smoothed function and the weight function. This forms the principal component (PC) scores indicated by f_i. Thus, the principal component scores corresponding to the weight function $\xi(t)$

for the i^{th} curve is

$$f_i = \int \xi(t)\, x_i(t)\, dt. \tag{4.4.1}$$

The formula (4.4.1) is a general formula for the principal component scores. The procedure to find the first m principal component scores, that are the most informative ones, is as follows:

1. the first PC scores are defined as $f_{i1} = \int \xi_1(t)\, x_i(t)\, dt$;

2. the first weight function $\xi_1(t)$ is found such that $\frac{1}{N}\sum_i f_{i1}^2(t)$ is maximized subject to the constraint $||\xi_1||^2 = \int \xi_1^2(t)dt = 1$;

3. the weight function $\xi_1(t)$ defines the largest component of variation;

4. the second PC scores are defined as $f_{i2} = \int \xi_2(t)\, x_i(t)\, dt$;

5. the second weight function $\xi_2(t)$ is found such that $\frac{1}{N}\sum_i f_{i2}^2(t)$ is maximized subject to the constraints $||\xi_2||^2 = \int \xi_2^2(t)dt = 1$ and ξ_1 and ξ_2 are orthogonal. That is, $\int \xi_1(t)\xi_2(t)dt = 0$;

6. the weight function $\xi_2(t)$ defines the second largest component of variation;

7. we repeat this procedure until we find all m important PC; the maximum number of important PC is $m = min\ (N-1,\ K,\ n)$ where K is the number of basis functions in estimating $x(t)$.

8. note that the m^{th} weight function is chosen such that it is orthogonal to the previous weight functions mutually.

The above procedure is useful when there is only one function. In the case of having more than a function we need to incorporate the effect of the second function in defining the PC scores. Therefore, the bivariate PCA in the formula (4.4.1), adapted for our 2-D data, the coordinate functions X and Y, is

$$\begin{aligned} f_i &= < \xi(t),\ Coord_i(t) > \\ &= \int \xi^X(t)\, X_i(t)dt + \int \xi^Y(t)\, Y_i(t)\, dt, \end{aligned} \tag{4.4.2}$$

where $Coord_i(t) = (X_i(t), Y_i(t))$ is the vector of functions and $\xi(t) = (\xi^X(t), \xi^Y(t))^T$ is the 2-vector of the weight functions corresponding to the functions X and Y coordinates, respectively. The constraints are now:

$$\begin{aligned} ||\xi||^2 &= ||\xi^X||^2 + ||\xi^Y||^2 = 1 \\ < \xi_1, \xi_2 > &= \int \xi_1^X(t)\xi_2^X(t)dt + \int \xi_1^Y(t)\xi_2^Y(t)dt = 0, \end{aligned}$$

where the inner product shows the orthogonality of functions ξ_1 and ξ_2. The weight function $\xi(t)$ is now the results of solving the eigenequation system $V\xi = \rho\xi$. The system of eigenfunction equations can be written as

$$\begin{aligned} \int v_{XX}(s,t)\xi^X(t)dt + \int v_{XY}(s,t)\xi^Y(t)dt &= \rho\xi^X(s) \\ \int v_{YY}(s,t)\xi^Y(t)dt + \int v_{YX}(t,s)\xi^X(t)dt &= \rho\xi^Y(s), \end{aligned} \tag{4.4.3}$$

where $v_{XY}(s,t) = v_{YX}(t,s)$ are cross-covariance functions, $v_{XX}(s,t)$ and $v_{YY}(s,t)$ are covariance operators, and ρ is the eigenvalue.

The term *harmonics* is used to indicate "principal component of variation in curves in general" (Ramsay and Silverman 2005, p.151). The results of the first four rotated harmonics of C3 at time $T1$ are shown in Figure 4.4.1. The purpose of using the rotated harmonics is for a better interpretability of the harmonic plots. Rotation of the weight functions is done using a method called Varimax strategy which is explained in Section 4.4.1. These four harmonics account for 85% of the total variability. Each plot shows the mean curve of the X-Y coordinates along with the ± 0.08 of each harmonics. The value 0.08 is chosen subjectively and it is used only for a better interpretability of the harmonic plots; we tried several values among which the chosen value seemed to give a more interpretable plot. After applying the Varimax rotation, the total proportion of variability explained by these 4 harmonics remained about the same but the proportion explained by each of these harmonics changed. Before rotating the weight functions, PC1 accounted for 57% and the rest for 11, 11, and 5% of the variability, respectively (a total of 84%). These values become 30, 18, 22, and 15% after using the rotated harmonics. The first harmonic accounts for the highest variability in the curve which can be associated with the

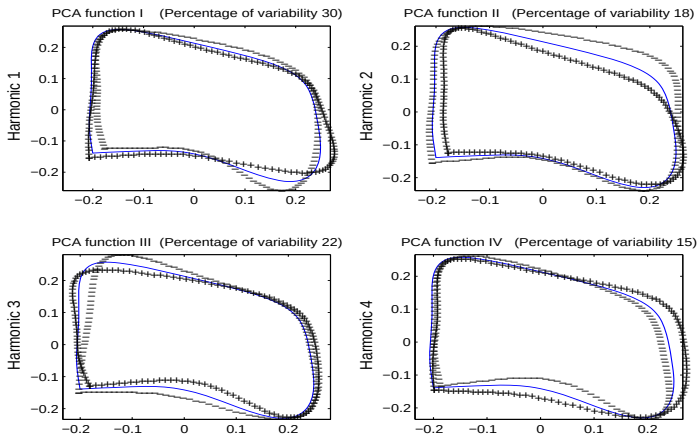

Figure 4.4.1: The first four important harmonics, each plot shows the mean function (solid blue) +/- small amount of harmonics.

change in the corners of the lower concavity. The second harmonic contributes to the vertical variability of C3 bone. The third and fourth PC account for the variability in the upper left corner, the concavity in the lower border, and the width of the bone. As mentioned, PC1 accounts for 30% of the total variability alone but we need to know which one of the curves X or Y is mostly responsible for causing this variation. From Figure 4.4.2 we see that the Y coordinate contributes slightly more than X in making the variability in the first PC; to see the rest of the harmonics of the X and Y coordinates separately see Appendix C.

The number of important principal components were obtained using Figure 4.4.3. It shows how the choice of 4 harmonics in our analysis was made. We plot the first 10 largest eigenvalues against the logarithms of these eigenvalues and as Ramsay, Hooker, and Graves (2009) pointed out, the first few logarithms of these eigenvalues

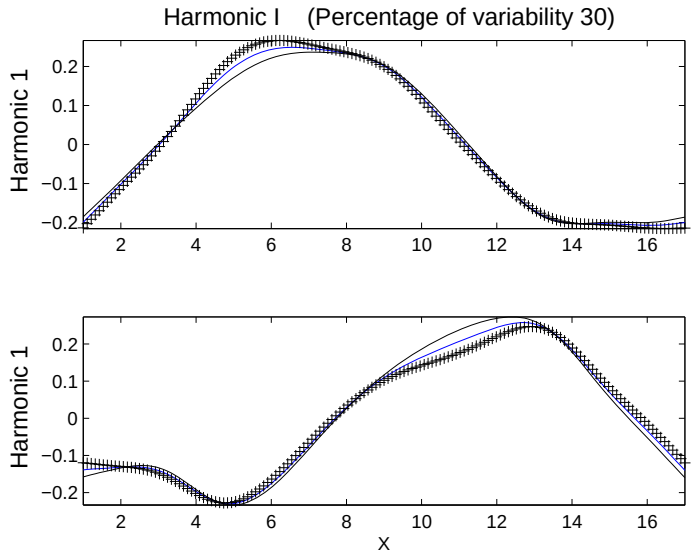

Figure 4.4.2: The first harmonic shown for X and Y coordinates separately to aid the detection of the source of variability in the first harmonic.

are large but then they decrease, as the eigenvalue does, in a linear fashion. This linear trend is shown as the blue dashed lines. From the plot we see that the linear decrease starts approximately after the first 4 eigenvalues.

To analyse the components of variability in C3 curve at time $T2$ we consider an alternative way of plotting principal components which is illustrated in Figure 4.4.4 and is called cycle plot. Each dot shows the mean of the (X, Y) coordinate across 62 patients and each plot shows the overall mean shape of C3 bone. If the direction of an arrow is parallel to the X-axis, then the width of the bone contributes more to the first PC and if it is parallel to the Y-axis, this contribution is made more to the second principal component. Whenever the arrows are along the mean, there is a small or no contribution within the specified spatial location. At $T2$ the total variability in C3 curve accounts for 83% which is almost the same as what was observed at $T1$.

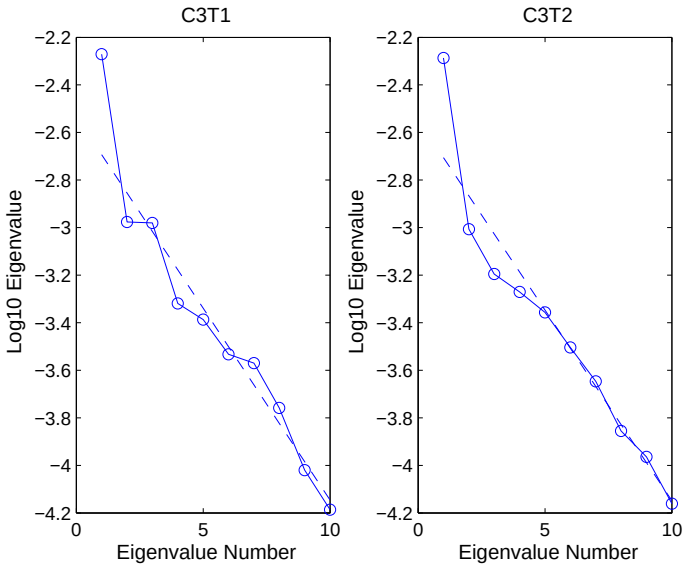

Figure 4.4.3: This plot shows how to choose the number of important harmonics. The blue dashed line shows the linear trand.

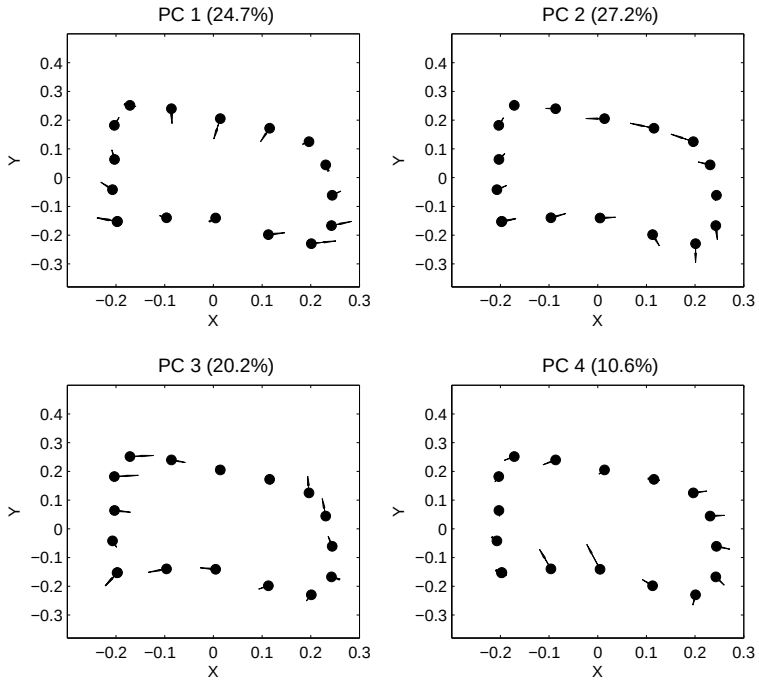

Figure 4.4.4: Cycle plots for the first four harmonics at time T2.

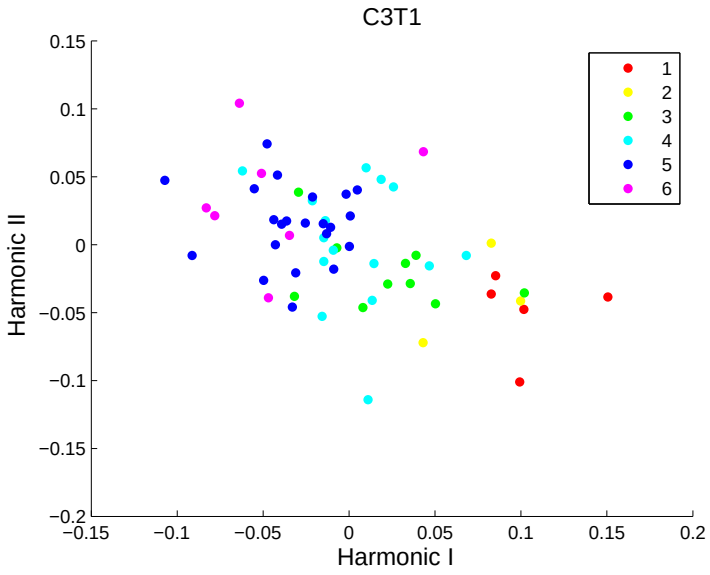

Figure 4.4.5: Plot of PC scores by the Ceph at time T1. There are 6 different Ceph scores.

The plot of PC scores is shown in Figure 4.4.5, from which we can explore and identify the possible clusters formed by the patients with different Ceph scores. Each X and Y axis shows the PC scores, $f_{i1}(t) = \int \xi_1^X(t)\, X_i(t)dt + \int \xi_1^Y(t)\, Y_i(t)\, dt$ and $f_{i2}(t) = \int \xi_2^X(t)\, X_i(t)dt + \int \xi_2^Y(t)\, Y_i(t)\, dt$, respectively. From the position of the dots we see that patients with higher Ceph scores are high in the second principal components, while the ones with lower Ceph scores are higher in the first PC. See Appendix C for more plots of PC scores by groups including gender at times $T1$ and $T2$, Ceph scores at time $T2$, and HW scores at times $T1$ and $T2$.

4.4.1 The Varimax rotation

A well-known method, called the Varimax strategy, is used to orthogonally rotate the original weight functions $\xi = (\xi_1, \xi_2, ..., \xi_k)^T$ found by FPCA method. The aim is to find a set of more interpretable weight functions. The orthogonal rotation of

the weight functions, used in the Varimax criterion, means applying an orthogonal matrix of order k, denoted by Q, to the weight functions ξ where Q has the property of $QQ^T = Q^TQ = I$, that is, $\psi = Q\xi$. From here we treat $\psi = (\psi_1, \psi_2, ..., \psi_k)^T$ as the new weight functions. The vector of functions ψ will be as effective as ξ in approximating the original curves. To explain how the Varimax rotation works, let B be the $k \times n$ matrix including all the original k weight functions evaluated at n time points. Now, suppose that the m^{th} row of this matrix B includes the values $\xi_m(t_1), \xi_m(t_2), ..., \xi_m(t_n)$ where $t_1, t_2, ..., t_n$ are equally spaced time points in the interval τ. Then the values of the rotated weight functions can be stored in a $k \times n$ matrix A, thus $A = QB$. The orthogonal rotation matrix Q is then found by maximizing the variance of the a^2_{mj}, where a^2_{mj} are the diagonal elements of the matrix A^TA. Thus, very large values or approximately zero values are desirable for a^2_{mj}. Mathematically, $\sum_m \sum_j a^2_{mj} = trace(A^TA) = trace(B^TB)$.

4.5 Conclusions and future research

We conclude this chapter with some remarks. First, GPA registration method worked well in aligning C3 curves. Second, we associated each coordinates of the landmarks, X and Y, with the width and height of C3 bones. Third, our analysis shows that more changes in the height rather than width of the average shape of C3 curve have occurred over time. Finally, the result of FPCA revealed two components of variability and each was attributed to the width and height of the third cervical vertebra. Also, we plan to carry out a similar analysis on the second and fourth cervical vertebrae.

Bibliography

[**1**] Ramsay, J. O., and Silverman, B. W. (2005), *Functioinal Data Analysis*, New York: Springer.

[**2**] Ramsay, J. O., Hooker, G., and Graves, S. (2009), *Functional Data Analysis with R and MATLAB,* New York: Springer.

[**3**] Ramsay, J. O., and Silverman, B. W. (2002), *Applied Functional Data Analysis,* Springer Verlag: New York.

[**4**] Ramsay, J. O., and Dalzell, C. J. (1991), "Some Tools for Functional Data Analysis," *Journal of the Royal Statistical Society,* Ser. B, 53, 539-572.

[**5**] Ramsay, J. O. (2006), "Functional Data Analysis of Continuous Judgments in Music Cognition," Unpublished slides, ftp:-ego.psych.mcgill.ca-pub-ramsay-FDAtalks-Music.ppt.

[**6**] Ramsay, J. O. (2006), "Mouse Livers: Derivatives and Functional Linear Models," Unpublished slides, ftp:-ego.psych.mcgill.ca-pub-ramsay-FDAtalks-Mouse Livers.ppt.

[**7**] Tian, T. S. (2010), "Functional Data Analysis in Brain Imaging Studies," *Frontiers in Psychology*, 1-35, University of Houston, Dept. of Psychology.

[**8**] Ramsay, J. O. (2006), "Models for Output-Buffered Systems: An introduction to Dynamics," http://www.samsi.info/communications/jim-ramsay-models-output-buffered-systems-videos-1-2.

[**9**] Ramsay, J. O. (2006), "Models for Output-Buffered Systems: An introduction to Dynamics," http://www.samsi.info/communications/jim-ramsay-models-output-buffered-systems-videos-3-4.

[**10**] De Boor, C. (1977), "Package for Calculating with B-Splines," *Journal on Numerical Analysis*, 14, 441-472, Society for Industrial and Applied Mathematics.

[**11**] Gu, C. (1992), "Cross-Validating Non-Gaussian Data," *Journal of Computational and Graphical Statistics*, 1, 169-179.

[**12**] Hastie, T., Tibshirani, R., and Friedman, J. (2009), *The Elements of Statistical Learning, Data Mining, Inference, and Prediction*, p. 186-190, New York: Springer.

[**13**] Fan, J., and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications,* London: Chapman & Hall.

[**14**] Shim, J., Bogowicz, P., Heo, G., and Lagravère, M.O. (2011), "Interrelationship and Limitations of Conventional Radiographic Assessments of Skeletal Maturation," Unpublished manuscript, University of Alberta.

[15] Dryden, I. L., and Mardia, K.V. (1998), *Statistical Shape Analysis*, London: Wiley.

[16] Craven, P., and Wahba, G. (1979), "Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation," *Numerische Mathematik*, 31, 377–403.

Appendix A

Additional plots for chapter 2

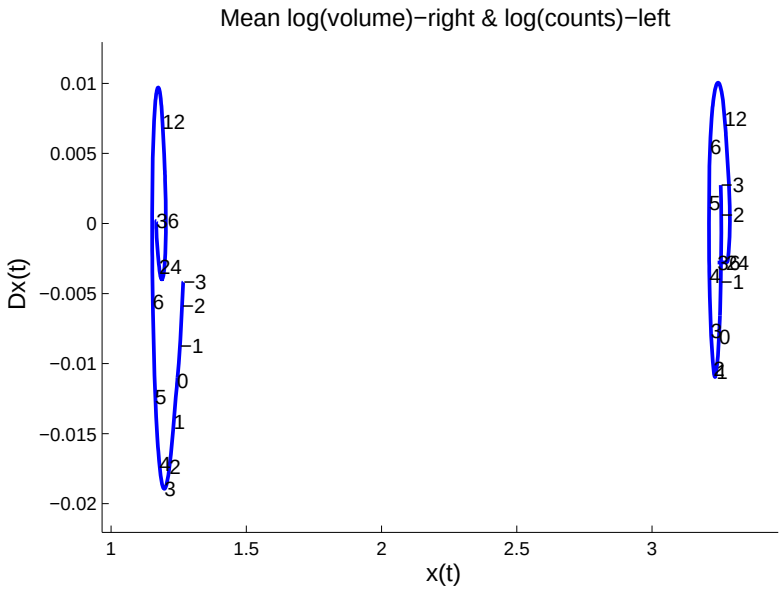

Figure A.0.1: Phase-plane plots-volume and counts.

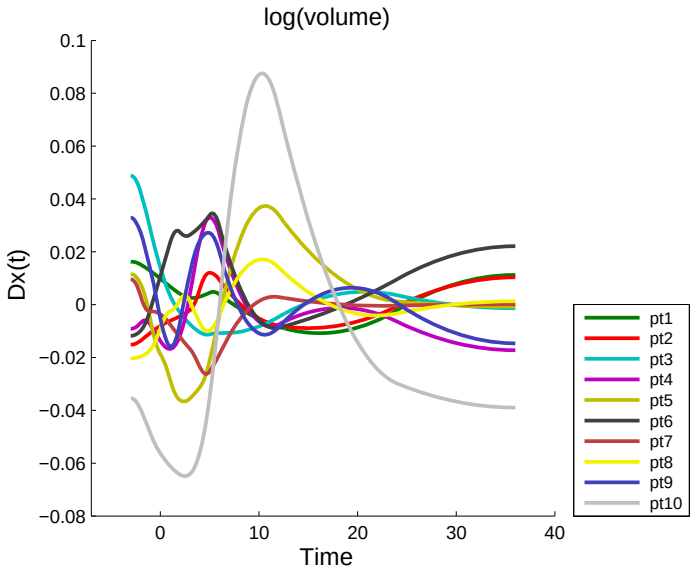

Figure A.0.2: Dlogvolume(t) over time.

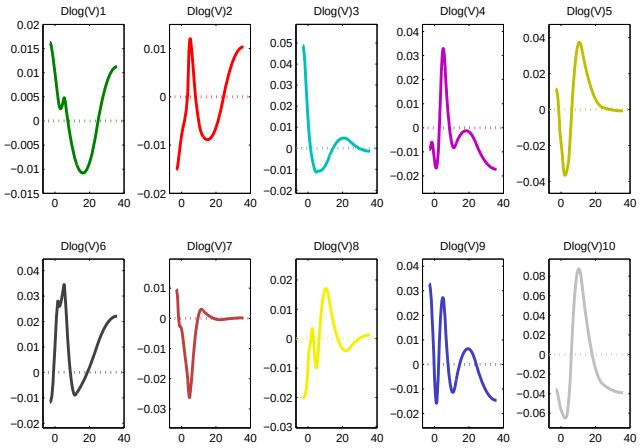

Figure A.0.3: Individual Dlogvolume(t) over time.

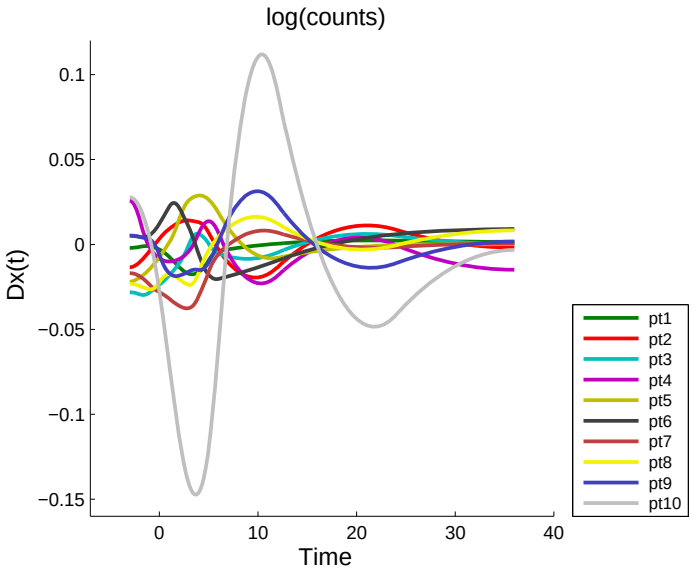

Figure A.0.4: Dlogcounts(t) over time.

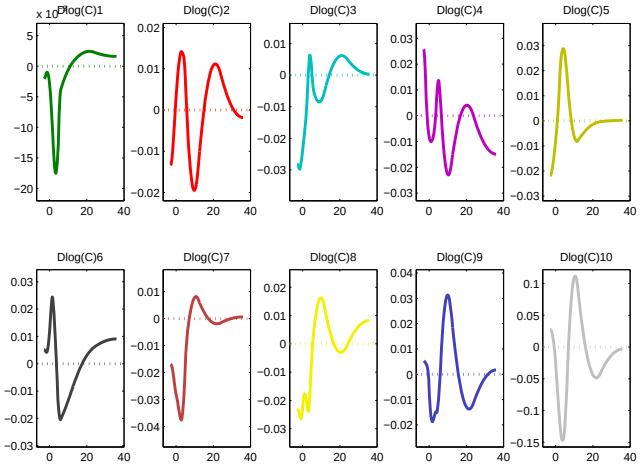

Figure A.0.5: Individual Dlogcounts(t) over time.

Appendix B

Additional plots for chapter 3

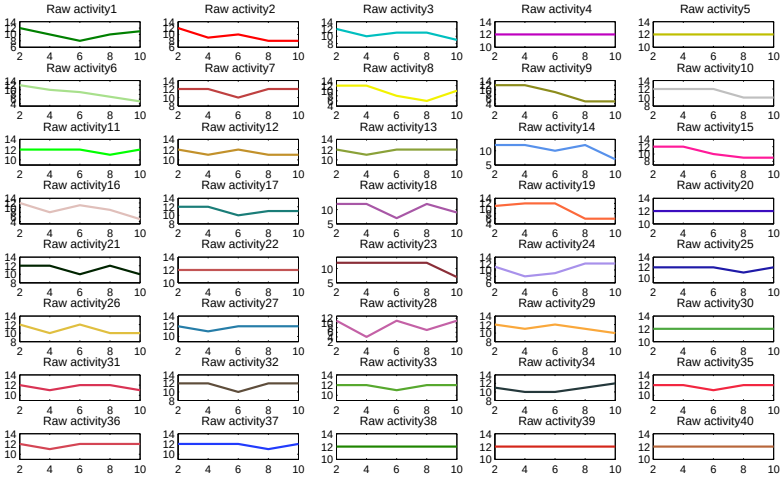

Figure B.0.1: Indivisual plots of the raw data-activity.

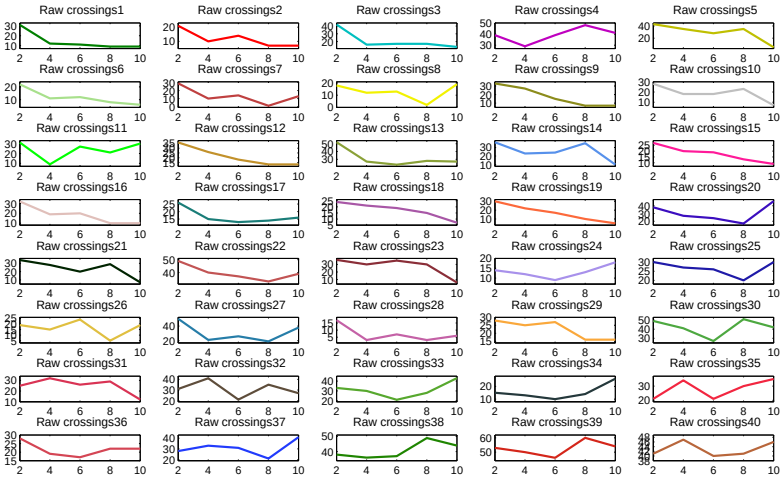

Figure B.0.2: Indivisual plots of the raw data-crossings.

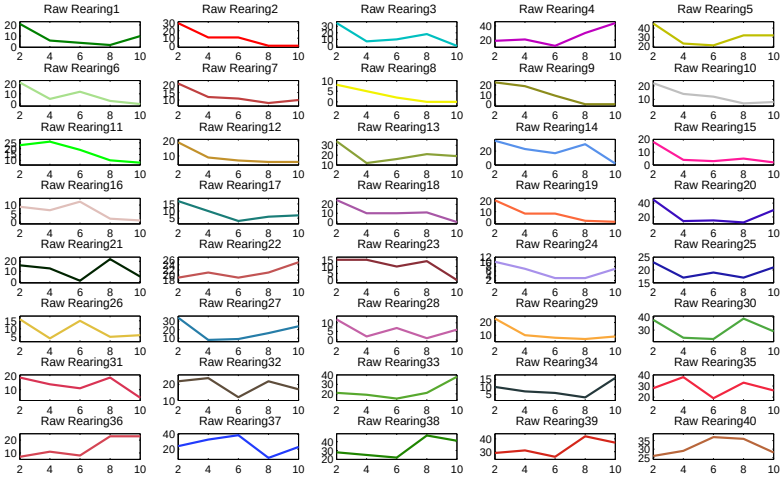

Figure B.0.3: Indivisual plots of the raw data-rearing.

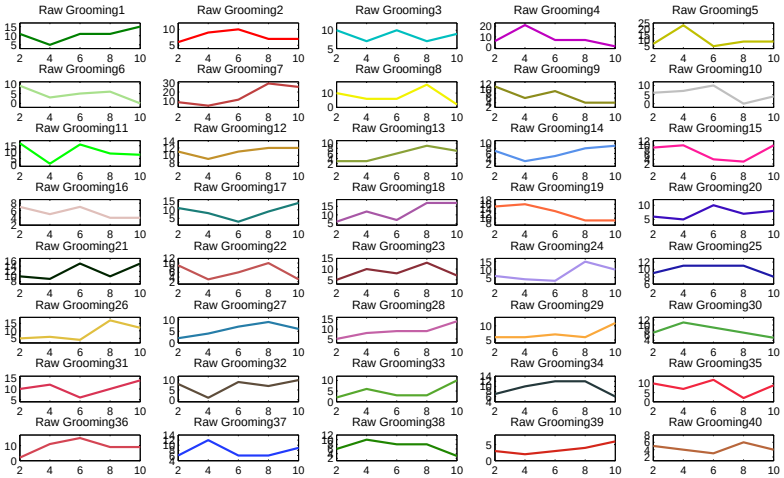

Figure B.0.4: Indivisual plots of the raw data-grooming.

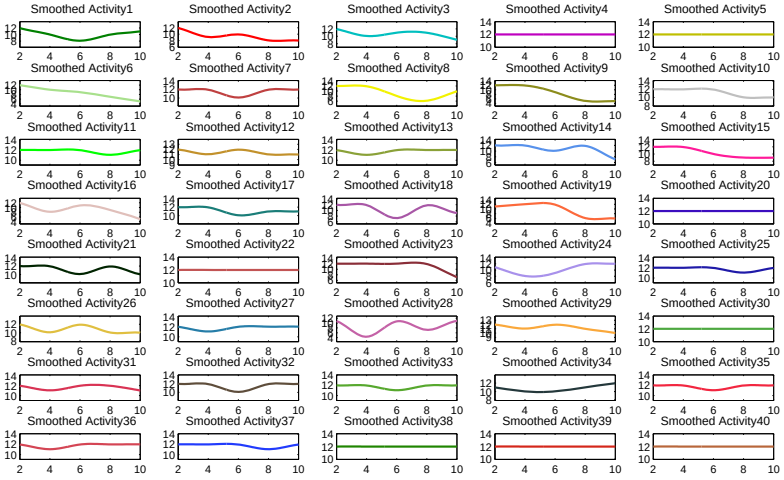

Figure B.0.5: Indivisual plots of the smoothed curves-activity.

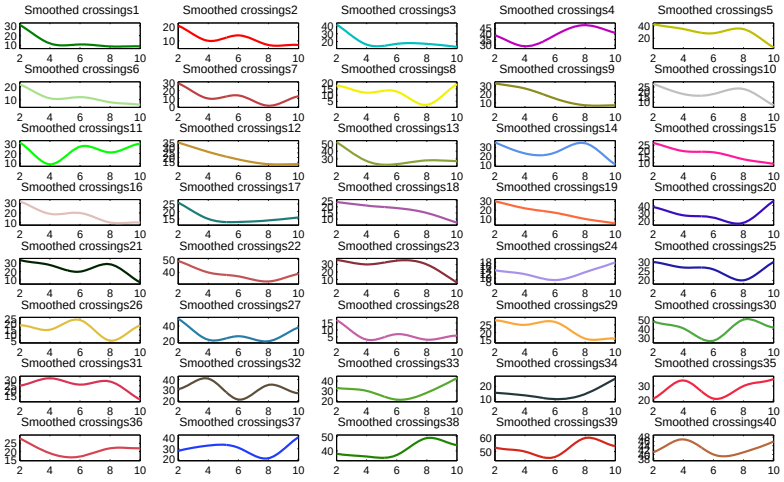

Figure B.0.6: Indivisual plots of the smoothed curves-crossings.

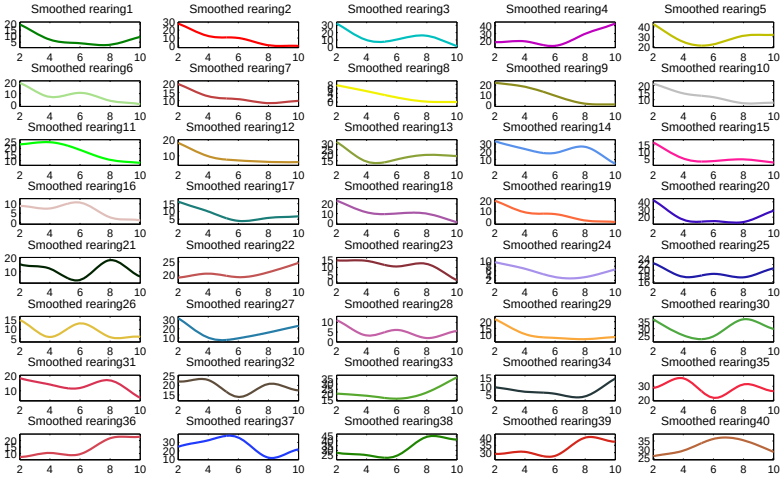

Figure B.0.7: Indivisual plots of the smoothed curves-rearing.

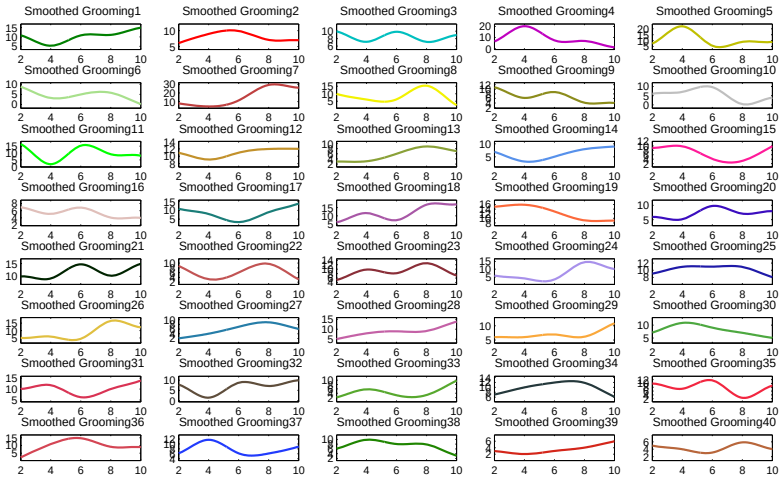

Figure B.0.8: Indivisual plots of the smoothed curves-grooming.

Appendix C

Additional plots for chapter 4

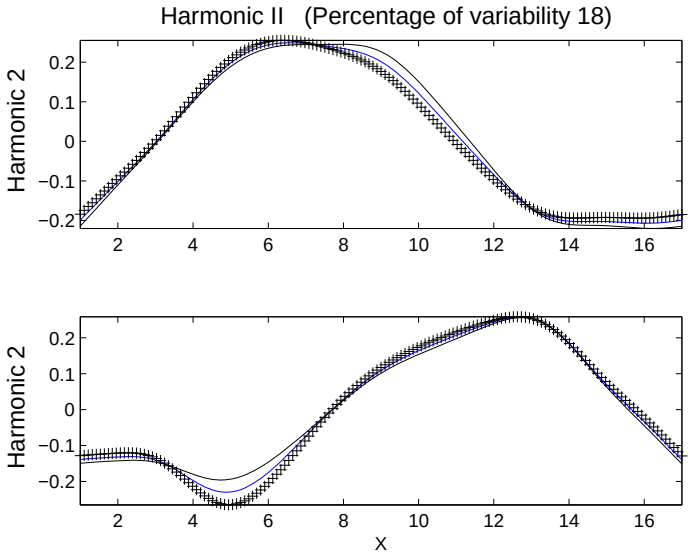

Figure C.0.1: Harmonic II of the mean C3 curve at time T1 +/- 0.08 of each PC curve.

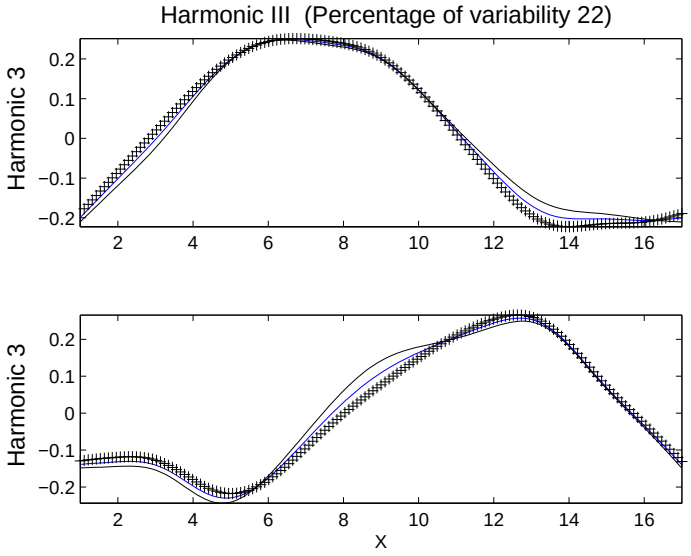

Figure C.0.2: Harmonic III of the mean C3 curve at time T1 +/- 0.08 of each PC curve.

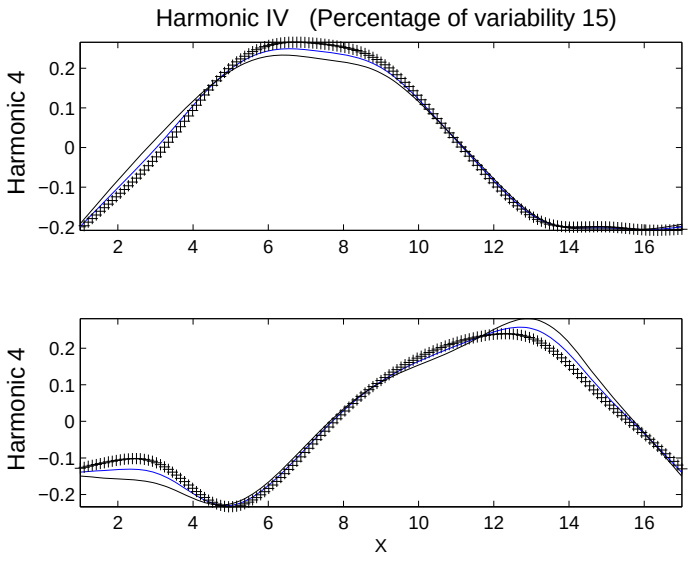

Figure C.0.3: Harmonic IV of the mean C3 curve at time T1 +/- 0.08 of each PC curve.

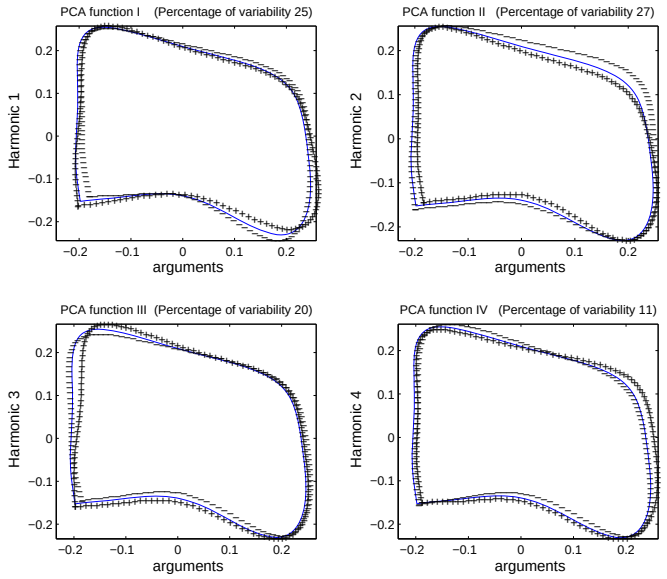

Figure C.0.4: The first four important harmonics, each plot shows the mean function (solid blue) +/- small amount of harmonics at time T2.

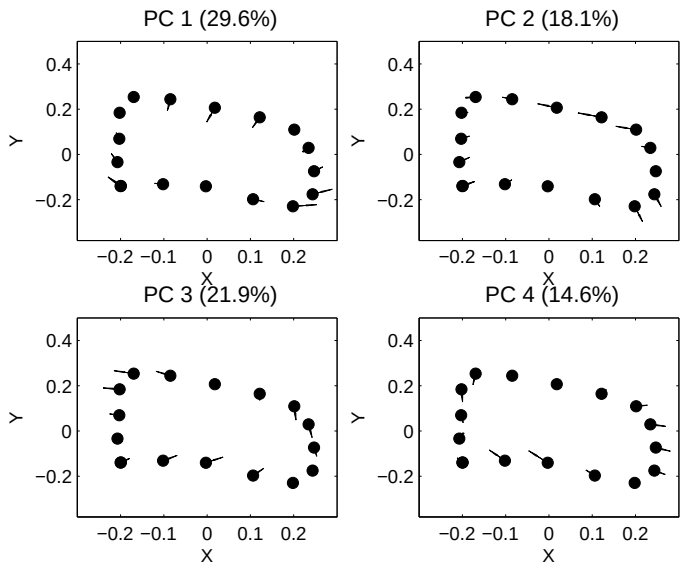

Figure C.0.5: Cycle plots for the first four harmonics at time T1.

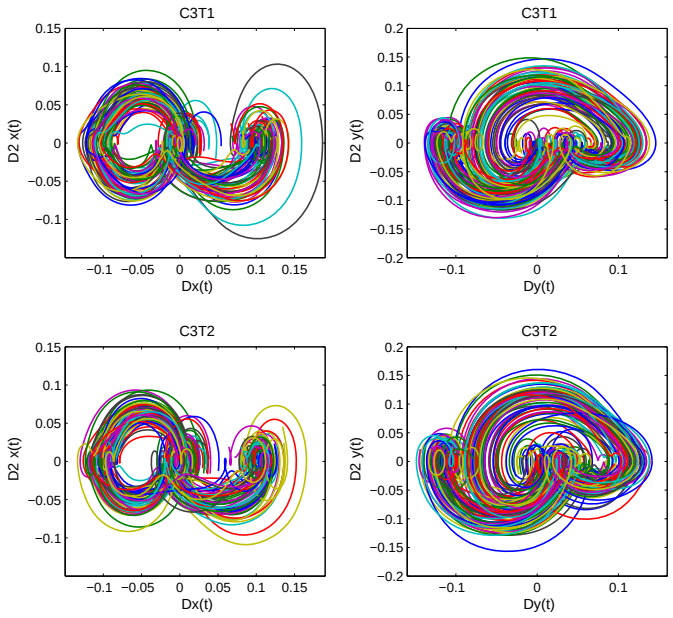

Figure C.0.6: Phase-plane plots for X and Y curves at times T1 and T2.

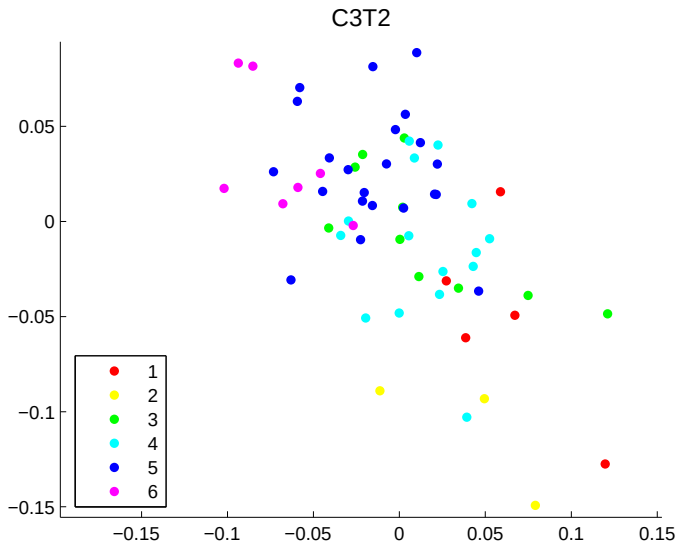

Figure C.0.7: Plot of PC scores by the Ceph at time T2. There are 6 different Ceph scores.

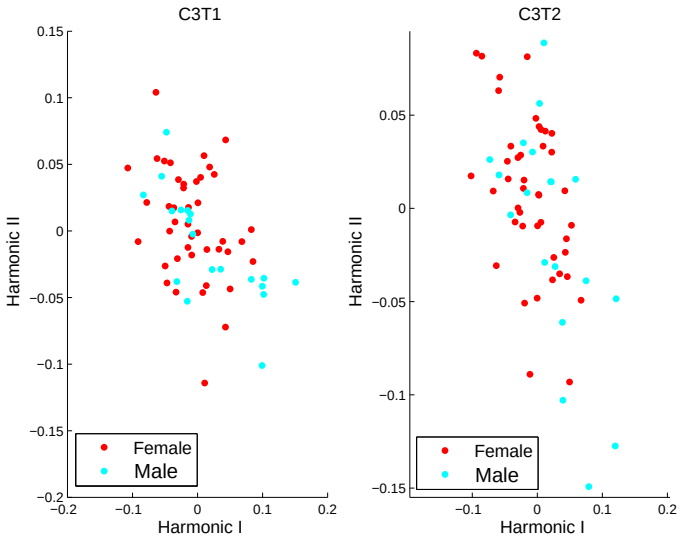

Figure C.0.8: Plots of PC scores by the gender at times T1 (left) and T2 (right).

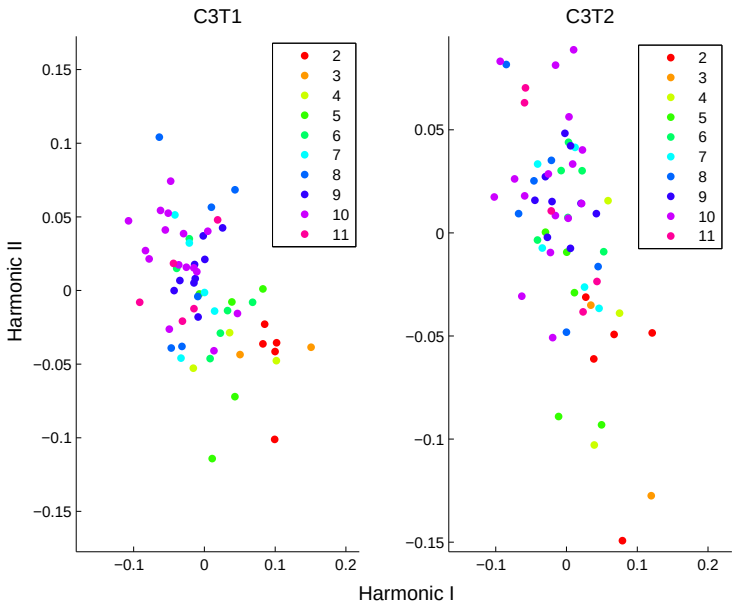

Figure C.0.9: Plots of PC scores by the HW scores at times T1 (left) and T2 (right).