MSI Journal of
Multidisciplinary Research (MSIJMR)

MULTIVARIATE FUNCTIONAL DATA ANALYSIS AND ITS APPLICATION TO GROWTH CURVES

Ortese Collins Aondona[1*], Nwaosu Chigozie Sylvester[2]

[1*,2]Joseph Sarwuan Tarkaa University, Makurdi – Nigeria.

*** Correspondence:** Ortese Collins Aondona

The authors declare that no funding was received for this work.

OPEN ACCESS

Received: 02-February-2026
Accepted: 06-April-2026
Published: 13-April-2026

This article is published in the **MSI Journal of Multidisciplinary Research (MSIJMR) ISSN 3049-0669 (Online)**

The journal is managed and published by MSI Publishers.

Volume: 3, Issue: 4 (April-2026)

ABSTRACT: Functional data are high dimensional data recorded continuously during a time interval at discrete time points. Statistical methods for analyzing functional data have been widely studied; however, dependent functional data that exhibit additional complex features such as time or contemporaneous dependence, stochastic volatility, jumps or rapid-changing smoothness, sparsely or irregular with non-negligible error is a challenge in existing methods. In this article, a proposed nonparametric approach to analyzing functional data that includes a variety of such complexities is used to analyze growth trajectories. The monthly weight of children registered and receiving immunization at Bishop Murray Medical Centre, Makurdi for a period of nine months was retrieved and used as training dataset to estimate change over time within each individual and then compare change across individuals. Exploratory analysis of the curve, mean, standard deviation and bivariate correlation functions were conducted. The principal component analysis reduced the multivariate data to a finite-dimensional vector of basis and visualizes the variation in the functional data. The least square regression model for fitting of basis expansions procedure was used for smoothing of the curve. The principal component curves and coefficient for the functional linear regression model and the functional parameters was estimated for the selected basis functions. Model adequacy was checked using root mean square error of approximation. The fitted model is significant as it provides the estimation of the principal components using functional regression.

Keywords: *Model, Functional data, functional observations, functional time series, model, inference.*

1.0 INTRODUCTION

Functional data analysis deals with the analysis and theory of data that are in form of functions, images and shapes or more general objects. Functional data are intrinsically infinite dimensional which serve as a rich source of information for research and analysis but poses computational challenges. This is due to the fact that statistical analysis of complex data measured over time, across space or in multiple dimensions presents a variety of methodological and computational challenges.

Functional methods for independently and identically distributed data analytics are widely applicable in diverse fields such as economics and finance, brain imaging, Chemometric analysis, speech recognition and electricity consumption, environmental monitoring. However, there have been limited methodology on the analysis of dependence features which are inherent in functional data analysis. Such dependence is very common and can arise via multiple responses, temporal and spatial effects, repeated measurements, missing covariates or simply because of natural groupings in the data. Such data allows researchers to ask questions about their interdependence, rapid-changing smoothness and sparsely observed or irregular features.

While it is also possible to model functional data with parametric approaches usually mixed effects nonlinear models, the massive information contained in infinite dimensional data and need for a large degree of flexibility, combined with natural ordering (in time) within a curve datum facilitate non- and semi-parametric approaches which are the prevailing methods in literature [19]. Functional data analysis serves as a viable alternative approach to multi-level mixed model approach.

First generation functional data typically consist of a random sample of independent real-valued functions, $X_1(t), \ldots, X_n(t)$, on a compact interval $I = [0, T]$ on a real line. Such data is called curve data [7]. These real valued functions can be viewed as a realization of a one-dimensional stochastic process often assumed to be in Hilbert space such as $l^2(I)$.

Growth curve model typically refers to statistical methods that allow for the estimation of inter-individual variability in intra-individual patterns of change over time.

Growth models are fit to data by multilevel modeling framework or structural equation modeling framework. The multilevel model was originally developed to allow for the nesting of multiple individuals within a group. However, the model can equivalently be applied to multiple repeated measures nested within each individual that allows for the direct estimation of a variety of powerful and flexible growth models. The Structural Equation Modeling (SEM) incorporates the observed repeated measures as multiple indicators on one or more latent factors to characterize the unobserved growth trajectories.

Methodologies for independent and identically distributed functional data analysis have been well developed, but in the case of dependent functional data, the independent and identically distributed methods are not appropriate. Dependent Functional data exhibit additional complex features such as time or contemporaneous dependence, stochastic volatility, jumps or rapid-changing smoothness, sparsely or irregular with non-negligible errors. These expose the data to a variety of methodological and computational challenges. It is therefore expedient to develop a unifying method for modeling functional data to address the insufficiency of existing methods. Current approaches to growth modeling are highly flexible in terms of the inclusion of a variety of complexities including partially missing data, unequally spaced time points, non-normally distributed or discretely scaled repeated measures, complex nonlinear or compound-shaped trajectories, time-varying covariates (TVCs), and multivariate growth processes. In this article, we shall present a methodology for modeling functional data in a new context for which existing methods are inadequate to address. We explore Multivariate Functional Data Model framework for dependent functional data and apply the model to capture continuous trajectory of change in growth curve data. We shall also perform exploratory analysis to review characteristics, patterns and variations in the functional data using functional principal component and canonical correlations and finally we perform confirmatory analysis using a functional linear model.

Functional data are the type of data that is in the form of functions, images, shapes or more general objects usually recorded continuously during a time interval or intermittently at several discrete time points. Functional data presents data recorded on surfaces over a region over some dormain.

Hormann [10] defined functional data as multivariate data with an ordering on the dimension. It means that this type of data consists of curves varying over a continuum such as time, frequency or wavelength. Thus, functional data are often present when measurements at various time points are analyzed. The curves of functional data are usually interdependent, which means that the measurement at a point, example t_{i+1}depends on measurement at some other points example $t_1, \ldots, t_i \in \mathbb{N}$

Functional data are intrinsically infinite dimensional. This high dimensionality presents functional data as a reservoir source of information which brings many opportunities for research and data analysis. On the other hand, however, it poses great computational challenges.

Functional data analysis (FDA) is a new branch of statistics but the mathematical foundations dates back to Grenander [9] and Rao [20]. Functional data analysis was coined by Ramsay [18] and Ramsay [16]

Functional data consists of a random sample of independent real valued functions $X_1(t), \ldots, X_n(t)$ on a compact interval I=[0, T] on a real line. These real valued functions can be viewed as the realizations of a one-dimensional stochastic process often assumed to be in Hilbert space. Although scientific interest is in underlying stochastic process and its properties, in reality, this process is often latent and cannot be observed directly as data can be collected discretely over time, either in a fixed or continuous time grid. The time grid can be dense, sparse or neither; nor may vary from subject to subject. Functional data are therefore regarded as samples of fully observed trajectories.

Functional data analysis (FDA) has been widely applied in functional time series analysis. Aue [2] proposed a prediction algorithm which combines the idea of Functional data analysis and functional time series analysis.

While it is possible to model functional data with parametric approaches, usually with mixed effects nonlinear model models. The need for a large degree of flexibility combined with a natural ordering (in time) within a curve and prior available data facilitate Bayesian approach to modeling which are the prevailing methods in the literature as well as the focus of this research.

A functional time series is a sequence of (random) functions. These functions can arise from successive measurements made over a time interval, which is divided into several consecutive time intervals with equal length. Examples of such include daily curves of financial transaction data and daily patterns of geophysical and environmental data.

When the high-dimensional data are repeatedly measured on the same object over a period of time, a time series of continuous functions is observed within a common bounded interval. Functional data consists of infinitely many values$X_n(t), \ t \in \tau$. Such data has also been termed curve data [7]. A distinctive feature of functional data is that the curves are assumed to be smooth. In classical statistics, the data consists of samples of scalars or vectors; however, data analysis for functional data can be done using complex mathematical objects such as graphs. These graphs have more complicated structures than scalars or vectors in the fact that they are characterized not only by magnitude or direction but also by shape [15].

The shape of a random curve plays a role analogous to the dependence between the coordinates of a random vector. Some data can be naturally viewed as curves which are often very easy to visualize but difficult to compute derivatives of such curves. In many situations, the points (t_j) are extremely dense while at the other extreme end are sparse longitudinal data. Both arise from long records of observations showing a very rough periodic pattern. Modeling the periodicity of this pattern is often difficult. It is hence natural to treat the long continuous record as consisting of consecutive curves. This gives additional support to treat these data as a time series of curves evolving shape called functional time series [15].

Functional time series does not need to arise from cutting a continuous time record into adjacent pieces of natural length but focus on temporal or sequential dependence between curves. These curves are available at points in space. Such spatially indexed functional data denoted by $X_{(s_k,t)}$, where $S_1, S_2, \ldots, S_N$ are locations in some regions and$X_{(s_k,t)}$ is the value of the function $X_{(s_k)}$ at time t. the main feature of such data is that there is dependence of the curves which depend on the distance.

Most existing functional time series modeling methods, including those in the references cited above, rely on FPCA to project the intrinsically infinite-dimensional functional objects onto directions of a small number of leading FPCs. FPCA extracts the dominant modes of variation of a functional object over its entire domain, with captured information referred to as the "main features" of the considered process. However, the "minor" components neglected by FPCA often have highly localized features possessing information on functional variations over particular short intervals within the function domain.

A relatively recent dynamic FPCA introduced by Hormann [12]. Their work employed longrun covariance to include serial dependence of the data but suffered the same problem of loss of local features in dimension reduction.

At an early stage of development, FDA focused mainly on independent and identically distributed (i.i.d) functional data. In recent years, FDA has also been widely applied in functional time series analysis [11].

Wang [22] presented a rainbow plot for visualizing functional time series, where the distant past data are shown in red and most recent data are shown in purple. Aguilera [1] proposed functional principal component regression (FPCR) to model and forecast functional time series.

From a univariate functional time series modeling perspective, Bosq [3] proposed the functional autoregressive of order one (FAR (1)) and derived one-step-ahead forecasts that are based on a regularized form of the Yule–Walker equations. Later, FAR(1) was extended to FAR(p), under which the order p can be determined via Kokoszka [10] hypothesis testing procedure. To overcome the difficulty of infinite dimensional parameter estimation, Aue [2] showed the connection between a functional autoregressive and a vector autoregressive (VAR) model via functional principal component analysis (FPCA) and proposed a forecasting method based on the VAR forecasts of principal component scores. The approach of Aue [2] can also be considered an extension of Hormann [10], who forecast principal component scores via univariate time series forecasting methods, for example, the autoregressive integrated moving average (ARIMA) method. Klepsch [13] proposed the functional moving average process and introduced an innovative algorithm to obtain the best linear predictor. Klepsch [14] extended the VAR model to the vector autoregressive moving average model, which can be considered a simpler estimation approach of the functional autoregressive moving average model.

Many of the various modeling and forecasting methods for functional time series adopt FPCA as a dimension reduction tool. FPCA can decompose a time series of functions into a set of functional principal components (FPCs) and their corresponding principal component scores [2]. Correlation among each set of principal component scores obtained in FPCA decomposition possesses temporal dependence information on the original functional time series. To forecast principal component scores, kokoszka [15] considered univariate time series forecasting methods (e.g., autoregressive integrated moving average), while Aue [2] applied a multivariate time series forecasting method. Conditioning on the past curves and estimated FPCs, point forecasts can be obtained by multiplying the forecast principal component scores by the estimated FPCs.

In practice, FPCA extracts the dominant modes of variation of a functional object over its entire domain, with captured information referred to as main features of the process considered. However, the “minor” components neglected by FPCA often have highly localized features that possess information about functional variations over particular short intervals within the function domain. A relatively recent dynamic FPCA introduced by Hormann [12] employed long-run covariance to include serial dependence of the data but suffered the same problem of loss of local features in dimension reduction. This problem of FPCA inadequately extracting local features often leads to inferior estimation and forecasts. Moreover, the presence of outliers can diminish the performance of FPCA. Modeling and forecasting volatile functional data, such as high-frequency financial time series, must control the destabilizing effect of outliers.

Growth curve modeling is a broad term that has been used in different contexts during the past century to refer to a wide array of statistical models for repeated measures data. However, within the past decade, this term has primarily come to define a discrete set of analytical approaches.

More specifically, the contemporary use of the term growth curve model typically refers to statistical methods that allow for the estimation of inter-individual variability in intra-individual patterns of change over time. In other words, growth models attempt to estimate between-person differences in within-person change. Often these within-person patterns of change are referred to as time trends, time paths, growth curves, or latent trajectories. These trajectories might take on a variety of different characteristics that vary from person to person: They might be flat (i.e., showing no change over time), they might be systematically increasing or decreasing over time, and they might be linear or curvilinear in form.

The most basic growth model is composed of the fixed and random effects that best capture the collection of individual trajectories over time. Loosely speaking, a fixed effect represents a single value that exists in the population (e.g., the population mean weight for children), and a random effect represents the random probability distribution around that fixed effect (e.g., the population variance in weight of children). Consistent with these definitions, in the growth model, the fixed effects represent the meaning of the trajectory pooling of all the individuals within the sample, and the random effects represent the variance of the individual trajectories around these group means. For instance, for a linear trajectory, the fixed effects are estimates of the mean intercept (i.e., starting point) and mean slope (i.e., rate of change) that jointly define the underlying trajectory pooling of the entire sample; in contrast, the random effects are estimates of the between-person variability in the individual intercepts and slopes. Smaller random effects (i.e. smaller variances of intercepts and slopes) imply that the parameters that define the trajectory are more similar across the sample of individuals; at the extreme situation where the random effects equal 0, all individuals are governed by precisely the same trajectory parameters (i.e., there is a single trajectory shared by all individuals). In contrast, larger random effects (i.e., larger variances of intercepts and slopes) imply that there are greater individual differences in the magnitude of the trajectory parameters around the mean values; that is, some individuals are reporting higher or lower intercepts, or steeper or less-steep slopes relative to others. Taken together, the fixed and random effects capture the general characteristics of growth for both the group as a whole and for the individuals within the group.

There are two general approaches used to fit growth models to observe data that share certain similarities but are also characterized by certain distinct differences.

The first approach is to fit the growth model within the multilevel modeling framework [4]. The multilevel model was originally developed to allow for the nesting of multiple individuals within a group, such as children nested within classroom or siblings nested within family. However, the model can equivalently be applied to multiple repeated measures nested within each individual that allows for the direct estimation of a variety of powerful and flexible growth models.

The second approach is to fit the growth model within the structural equation modeling (SEM) framework [5]. The SEM incorporates the observed repeated measures as multiple indicators on one or more latent factors to characterize the unobserved growth trajectories. In many situations, the multilevel and SEM approaches to growth modeling are numerically identical, yet in others, there are important differences. For example, the multilevel model naturally expands to estimate higher levels of nesting (e.g., repeated measures nested within child, and child nested within classroom); the SEM approach is currently more limited in these situations. In contrast, the SEM is well suited to the estimation of latent variables that estimate and remove the effects of measurement error that might exist in the predictors or the outcomes; the multilevel model is currently more limited with respect to the estimation of comprehensive measurement models.

However, the similarities between the multilevel and SEM approaches often outweigh the differences, and the optimal approach should be selected as a function of the particular research application at hand.

It is as essential to establish the adequate fit of the hypothesized model within the growth modeling framework as it is in any other statistical model. How this is best done directly depends upon the specific analytic strategy used to estimate growth models. Within the SEM, it is possible to judge the fit of a hypothesized model relative to a saturated baseline model allowing for the estimation of standalone indices of overall fit for a given model. Examples include the model chi-square test statistic and fit indices such as the RMSEA (root mean squared error of approximation), CFI (comparative fit index), and TLI (Tucker-Lewis's index), among many others. Within the multilevel framework, it is not possible to estimate a saturated baseline model to which to compare the hypothesized model. As such, there are no standalone measures of overall fit for a hypothesized model (although other indices of appropriate fit can be used such as residuals and Wald tests). Instead, comparisons of competing alternative models are required (which we believe is a strategy that could be used to a much greater extent within the SEM framework). If two comparison models are nested (i.e., if the parameters of one model are a direct subset of the parameters of the second model), then formal likelihood ratio tests can be calculated based on the differences between model deviance. For models that are not nested, informal comparisons can be made using indices such as the Bayesian Information Criterion or the Akaike Information Criterion to rank order models. Regardless of approach, it is extremely important that clear evidence be presented that supports the adequacy of fit of the hypothesized model to the observed data prior to drawing theoretical inferences from the results.

Shang [21] attempted to simultaneously model and forecast state-specific mortality in Australia by determining a common trend across populations and the sex-specific trends of each state. Recently, Gao [6] considered consecutively applying a dynamic version of FPCA and a factor model to reduce multivariate functional time series to a manageable dimension. Apart from the inadequate feature extraction of FPCA, multivariate functional time series methods often fail to explicitly consider significant correlations among multiple functional processes under consideration. In addition, large collections of time series often have aggregation constraints. For example, male and female sub-national populations in different cities and states should aggregate consistently to the country total. Thus, multivariate functional time series methods should follow appropriate grouping structures to ensure coherence in the forecasts.

In Hormann [10], they proposed the notion of L^p-m-approximability, which can be used to quantify the temporal dependence of functional time series. It is an extension of m-dependence in scalar and multivariate time series analysis. Based on the work of Hormann [10], the ideas of time series analysis and FDA have been merged and many results in FDA under the assumption of i.i.d have been extended to L^p -m-approximable functional time series.

Providing reliable predictions is one of the most important goals of functional time series analysis. Bosq [3] has studied the functional best linear predictor for (stationary) functional linear process. But the problem is, we do not know the exact math formula for the best functional linear predictor, so in fact it is difficult to implement. Since we still lack advanced functional time series methodology, we often assume the functional time series to follow the first order functional autoregressive model (FAR (1)). And the prediction is also based on the assumption of FAR (1) structure.

Aue [2] proposed a prediction algorithm which combines the idea of FDA and functional time series analysis. And the prediction algorithm is not restricted to FAR structure. The basic idea is to use FPCA to reduce the finite-dimensional data to infinite-dimensional data. Then the issue of predicting functional time series is transformed to the prediction of multivariate time series.

METHODOLOGY

3.1 Functional Data Analysis

Functional variable is one whose values depend on a continuous magnitude such as time. They are functional in the sense that they are evaluated at any time in the domain, instead of the discrete way, in which they were originally measured or observed. Thus, a functional dataset is a set of curves $\{x_1(t), \ldots, x_n(t)\}$, with $t \in T$. Each curve can be observed at different time points of his argument t as $x_i = (x_i(t_0), \ldots, x_i(t_{mi}))'$ for the set of times $t_0, \ldots, t_m$, i=1, …, n.

Different approaches have been taken to the study of functional data, including nonparametric methods proposed by Ramsay [19]. The later method is adopted in this application, in which we seek to reconstruct the functional form of curves in order to evaluate them at any time point t. this method assumes that the curves belong to a finite dimensional space generating a basis of functions $\{\phi_1(t), \ldots, \phi_p(t)\}$ and so they can be expressed as

$$x_i(t) = \sum_{j=1}^{p} a_{ij}\phi_j(t), \text{ i=1, …, n} \tag{1}$$

The functional form of the curves is determined when the basis coefficients $a_i = (a_{i1}, \ldots, a_{ip})'$ are known. These can be obtained from the discrete observations either by least squares or by interpolation. In our application, the least square is considered for functional representations.

Depending on the characteristics of the curves and the observations, various classes of basis can be used. In practice those most commonly used are on one hand, the basis of trigonometric functions for regular, periodic, continuous and differentiable curves and on the other, the basis of B-spline functions which provides a better local behavior. This research uses the B-spline functions as the class of basis representation. Basis expansions can provide good approximations to functional data provided that the basis functions have the same essential characteristics as the process generating the data.

3.2 Exploratory Analysis

3.2.1 Curve representation

Let $x_1(t), x_2(t), \ldots, x_n(t)$ be a set of curves all of them observed at the same time points $t_1, t_2, \ldots, t_m$. Then the available information in this situation is the matrix

$$X = \begin{bmatrix} x_1(t_1) & x_1(t_2) & \ldots & x_1(t_m) \\ x_2(t_1) & x_2(t_2) & \ldots & x_2(t_m) \\ x_n(t_1) & x_n(t_2) & \ldots & x_n(t_m) \end{bmatrix}$$

The basis coefficient of all the curves is given as

$$A^T = (\phi^T \phi)^{-1} \phi^T X^T \quad (2)$$

Where

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ a_{n1} & a_{n2} & \dots & a_{np} \end{bmatrix} \text{ and } \phi = \begin{bmatrix} \phi_1(t_1) & \phi_2(t_1) & \dots & \phi_p(t_1) \\ \phi_1(t_1) & \phi_2(t_2) & \dots & \phi_p(t_2) \\ \phi_1(t_1) & \phi_2(t_m) & \dots & \phi_p(t_m) \end{bmatrix}$$

3.2.2 Mean and standard deviation function

From the set of curves $x_1(t)$, $x_2(t)$, … , $x_n(t)$, the mean curve is defined as

$$\bar{x}(t) = \frac{1}{n} \sum_{i=1}^{n} x_i(t) \quad (3)$$

The mean curve can be expressed in terms of basic functions through basis coefficient

$$\bar{x}(t) = \frac{1}{n} \sum_{i=n}^{n} x_i(t) = \sum_{i=1}^{n} \sum_{j=1}^{p} a_{ij} \phi_j(t) \quad (4)$$

$$= \sum_{j=1}^{p} \bar{a}_j \phi_j(t)$$

$$\bar{a}_j = \frac{1}{n} \sum_{i=1}^{n} a_{ij}$$

3.2.3 Bivariate correlation function

From the set of curves $x_1(t)$, $x_2(t)$, …, $x_n(t)$ with mean curve $\bar{x}(t)$, the covariance surface is defined as

$$C(s,t) = \frac{1}{n-1} \sum_{i=1}^{n} \left(x_i(s) - \bar{x}(s)\right) \left(x_i(t) - \bar{x}(t)\right) \quad (5)$$

And from the correlation surface

$$r(s,t) = \frac{C(s,t)1}{\sqrt{C(s,s)C(t,t)}} \quad (6)$$

3.3 Functional Principal Component Analysis

Let $x_1(t)$, $x_2(t)$, …, $x_n(t)$ be a set of curves with mean curve $\bar{x}(t) = \frac{1}{n}\sum_{i=1}^{n} x_i(t)$ *and* covariance surface $C(s,t) = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i(s) - \bar{x}(s)\right)\left(x_i(t) - \bar{x}(t)\right)$.

Functional Principal components are defined as the vectors whose elements are obtained with the linear combinations of the sample curves.

$$\xi_i = \int_T (x_i(t) - \bar{x}(t) f(t) dt, \qquad i = 1, \dots, n \tag{7}$$

that maximizes the variance of $\xi_1, \dots, \xi_n$ and uncorrelated.

By imposing this condition, functional principal components are the solutions of this equation

$$\int_T C(s,t) f(s) ds = \lambda f(t) \tag{8}$$

Where λ is the variance of functional principal components

When curves are expressed in term of basic functions as $x_i(t) = \sum_{j=1}^{p} a_{ij}\phi_j(t)$, i=1, …, n previous equation has p solutions for p values λ that verify that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$. Each one of the λ_j is associated to a $f_j(t)$ function that define the functional principal component and is known as eigenfunctions or principal function curves

$$\xi_{ij} = \int_T (x_i(t) - \bar{x}(t) f_j(t) dt \qquad j = 1, \dots, p, \; i = 1, \dots, n \tag{9}$$

So, each principal component ξ_j is a vector of dimension n. Each principal component commutes a proportion of the total variability given by

$$\frac{\lambda_j}{\sum_{j=1}^{p} \lambda_j} \tag{10}$$

3.3.1 Explained variances and principal component curves

When the curves are expressed in terms of basic functions as $x_i(t) = \sum_{j=1}^{p} a_{ij}\phi_j(t)$, i=1…, n

Eigenfunctions are also expressed in terms of the same basic functions

$$f_i(t) = \sum_{k=1}^{p} F_{jk}\phi_k(t) \qquad j = 1, \dots, p \tag{11}$$

3.3.2 Functional principal component expansion

The original curves can be approximated by using a reduced set of functional principal components $x_i(t) = \sum_{j=1}^{q<p} \xi_{ij} f_j(t) \qquad i = 1, \dots, n$ (12)

By expressing the principal component curves $f_j(t)$ in terms of basic functions, we have an approximation of the original curves in terms of basis functions, that is, by knowing their basis coefficients

$$x_i(t) = \sum_{j=1}^{q<p} \xi_{ij} \sum_{k=1}^{p} F_{jk}\, \phi_j(t) \qquad i = 1, \dots, n \tag{13}$$

3.4 Confirmatory analysis

3.4.1 Functional principal component linear regression

The functional linear model is a functional method to explain a scalar response variable y in terms of a functional predictor variable x(t). A set of curves of the functional predictor $x_1(t), x_2(t), \dots, x_n(t)$ and an associated to a set of observations of the scalar response $y_1, y_2, \dots, y_n$, then the functional regression is formulated as

$$y_i = \alpha + \int_T^{\cdot} (x_i(t)\beta(t)dt + \varepsilon_i \qquad i = 1, \dots, n \tag{14}$$

To fit the functional linear model, it is usual to consider that the curves of the predictor variable and the functional parameter are expressed in terms of the same basis of functions.

$$x_i = \sum_{j=1}^{p} a_{ij}\phi_j(t), \quad i=1, \dots, n \quad \text{and } \beta(t) = \sum_{k=1}^{p} \beta_k \phi_k(t)$$

Under these conditions, the functional linear regression model turns to a classical linear regression model.

$$y_i = \alpha + \int_T^{\cdot} \left(\sum_{j=1}^{p} a_{ij}\phi_j(t)\right)\left(\sum_{k=1}^{p} \beta_k \phi_k(t)\right)dt + \varepsilon_i \tag{15}$$

$$y_i = \alpha + \sum_{j=1}^{p} \sum_{k=1}^{p} a_{ij} \left(\int_T^{\cdot} \phi_j(t)\phi_k(t)dt\right)\beta_k$$

$$= \alpha + \sum_{j=1}^{p} \sum_{k=1}^{p} a_{ij} \psi_{jk} \beta_k, \quad i = 1, \dots, n$$ where

$$\psi_{jk} = \int_T \phi_j(t) \phi_k(t) dt$$

3.5 Data Analysis Tool: A statfda (www.statfda.com) is the statistical software used for analysis of functional data and obtaining relevant results.

4.0 RESULTS

4.1 Data Used

A sample of 50 children who enrolled for post-natal immunization at the centre were sourced, extracted and used as training data at Bishop Murray Hospital, Makurdi for a period of 9 months (January 2025 to September 2025). The sample represents the population of children under immunization in a prospective study.

4.2 Exploratory Data Analysis

The figure shows the plot of the raw data

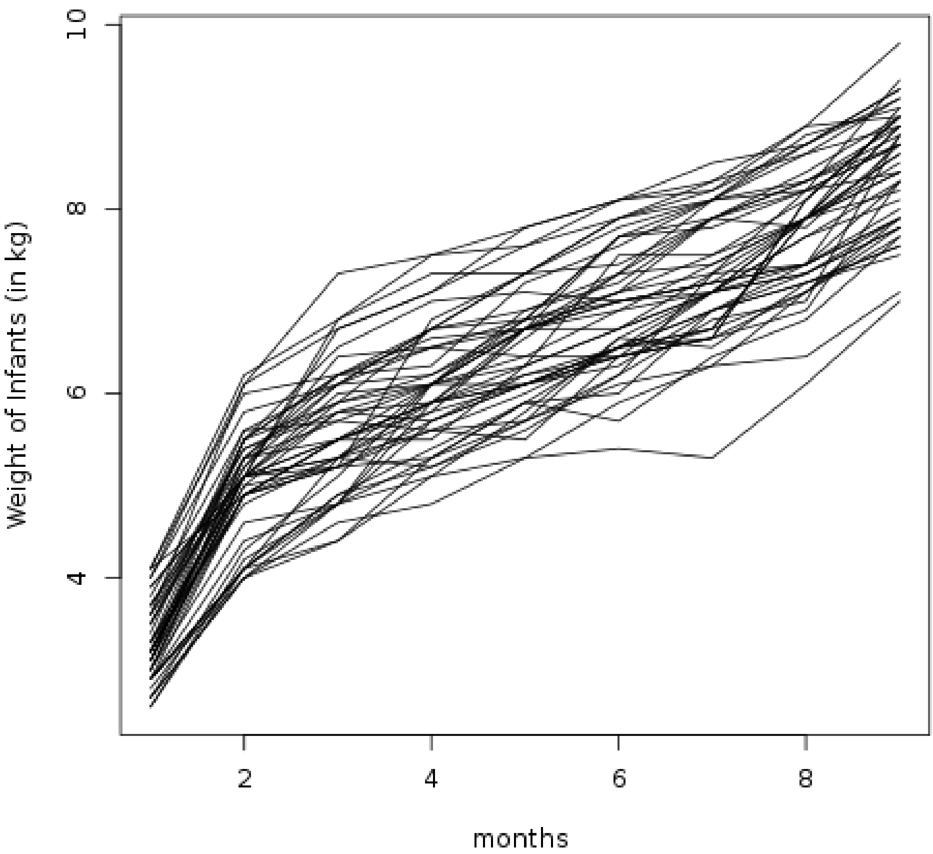

Figure1: Raw data representation

Each curve is a functional observation which represents the evolution of a child's weight in the range from 0 to 9 months. This is a continuous curve because the weight is defined as a function of time.

Each functional variable can be represented as a linear combination of known basis functions. A functional variable X belongs to an infinite dimensional space L^2 and, as a consequence, an infinite number of basis functions are needed for representing the functional variable with zero error. Hence, all the realizations $\{X_n\}_n \in Z$ of a functional random variable X can be expressed in relation to the same functional basis

Using a basis dimension of 6, the Smoothed curve for the data is as given below

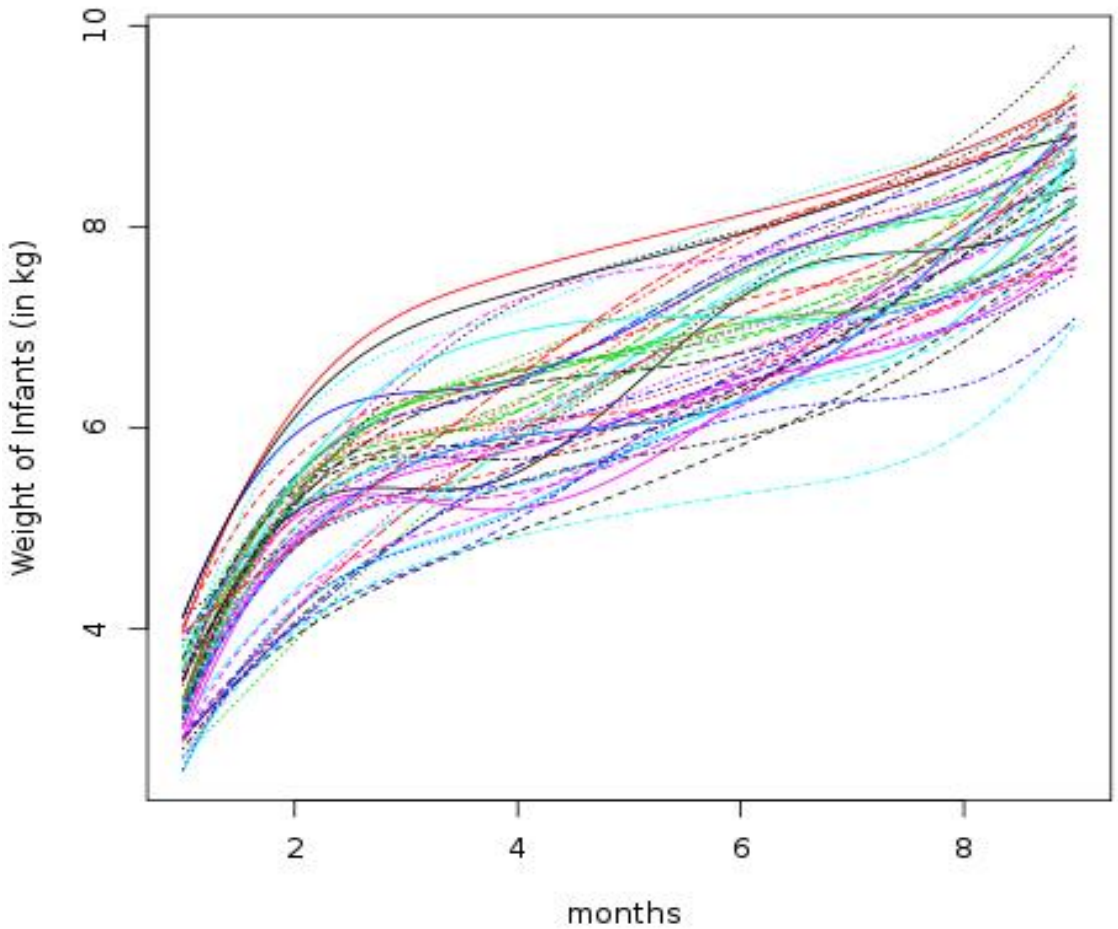

Figure 2: Plot of functional basis representation Functional Representation

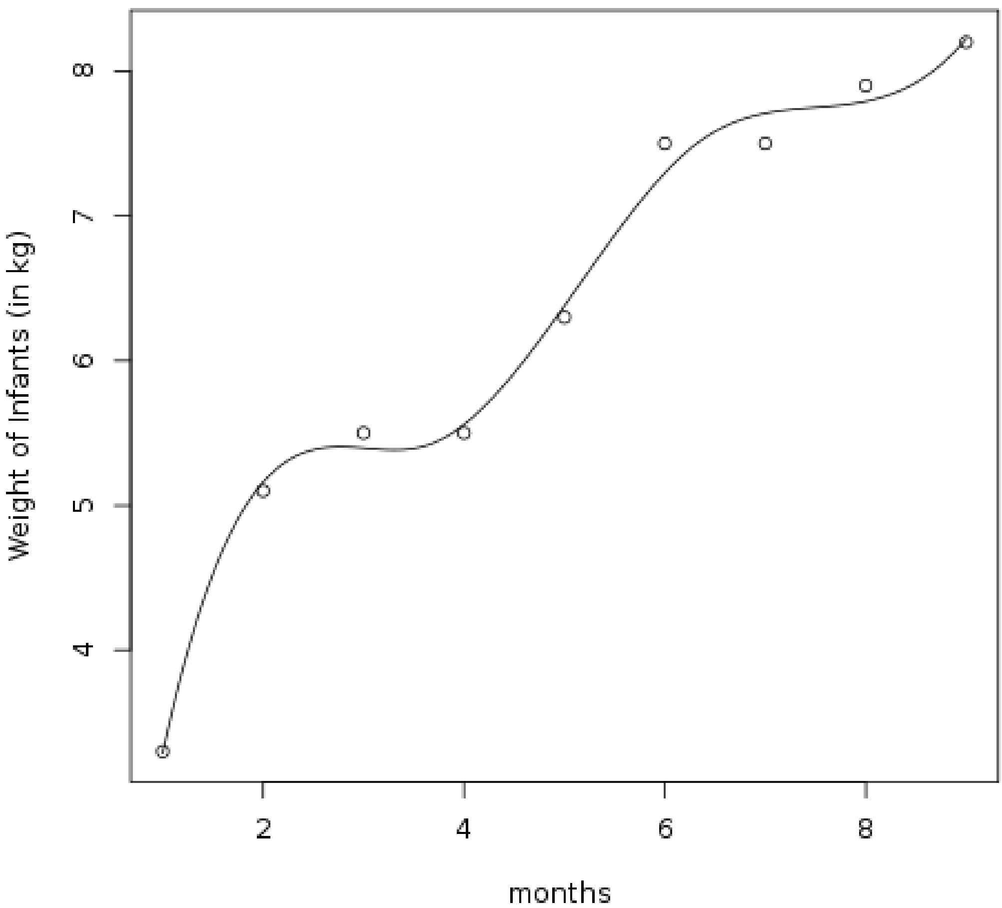

Figure 3: showing the functional representation at ϕ=6

BASIS REPRESENTATION OF THE SMOOTHED CURVES

Table 2: Basis Coefficients of the smoothed curves using B-spline with basis dimension of 6

	bspl4.1	**bspl4.2**	**bspl4.3**	**bspl4.4**	**bspl4.5**	**bspl4.6**
Curve 1	3.286941429	6.146796577	4.248109174	8.451081539	7.480429673	8.223303581
Curve 2	3.121746113	5.360825994	5.196249901	6.886295782	7.192010189	7.805858029
Curve 3	2.72888015	3.711653988	5.972048599	7.246398957	7.644802287	8.526663207
Curve 4	3.11714151	5.638400188	4.955900377	6.566194539	6.126388865	7.094754743
Curve 5	3.047850422	4.608809621	5.363401224	8.019067488	8.077668231	8.7013164
Curve 6	3.199944493	6.056017785	4.44157193	6.484318248	8.256999149	8.385599573
Curve 7	2.916442154	4.048935749	4.882064892	5.852796917	7.072124339	7.90887847
Curve 8	3.476207907	6.164064096	5.430303443	8.343393856	8.186062085	8.791791681
Curve 9	3.684832826	5.524368835	6.01075442	7.468026827	8.478759939	8.996113134
Curve 10	3.497784528	6.226070218	5.831228387	8.174968228	8.431361819	9.209847279
Curve 11	2.582856603	4.656269377	4.702495534	6.83536068	6.568736953	8.770685148
Curve 12	3.020800591	5.652309016	5.570182836	6.674905373	6.902104264	7.806020974
Curve 13	4.10764059	4.988952403	5.404882846	7.691087368	8.835126441	9.198490275

Curve 14	3.292231227	5.595666733	6.189062565	6.081444512	7.515539786	7.580907437
Curve 15	3.59677642	6.152217136	5.334944606	8.265152367	8.012195449	8.415250912
Curve 16	4.114199365	6.695634779	5.800618188	8.125799644	8.097176278	8.904244853
Curve 17	3.285905252	5.478563211	4.899341423	6.710609409	6.54125056	8.281732171
Curve 18	3.035930599	5.541876558	5.481098473	7.155302936	7.181087489	7.797590558
Curve 19	2.80885887	4.091446184	5.743147384	5.71342556	6.844578111	7.688232627
Curve 20	2.88605854	4.169717705	5.899654915	7.369130709	7.909844132	9.006250081
Curve 21	3.304056264	6.152696231	6.237040551	7.345624632	7.087566082	8.296362161
Curve 22	2.916123448	4.318371641	4.451266147	7.257889703	7.18661592	7.889780972
Curve 23	3.623039397	6.7399848	6.763267916	8.448906446	8.769218228	9.030472655
Curve 24	3.162546862	5.439239965	7.689902566	7.514809129	8.415288174	8.671697171
Curve 25	3.494440977	5.521461996	6.685501599	6.418526003	7.982214714	8.606242919
Curve 26	3.978900306	6.716261833	7.546825689	8.162455982	8.742840445	9.284399178
Curve 27	3.201073445	5.425315897	6.039087479	7.063401119	7.41642008	7.887011079
Curve 28	2.60070086	4.722549251	4.561942507	7.01768192	6.864351077	7.536324039
Curve 29	2.706777385	4.212956545	4.929179215	5.407483276	5.670640147	7.03227219
Curve 30	3.012992401	5.578640333	4.856658761	6.433843591	7.37147411	7.58977972
Curve 31	4.120895664	6.559764463	7.33892266	7.944974376	8.706257224	8.896422372
Curve 32	4.023951953	6.202483168	6.243360738	7.6363668	7.420559095	9.109367947
Curve 33	3.825734613	5.853220522	6.527405023	7.722641628	8.266176833	8.707238367
Curve 34	3.887917644	5.473278877	5.835308528	6.700041112	7.726426062	8.303436222
Curve 35	3.721758825	5.756443826	5.264153343	8.107879131	7.577626904	8.676246328
Curve 36	2.906661671	5.482005215	5.520186747	6.675208912	6.878571037	7.713660235
Curve 37	3.707722385	5.982767564	5.600215967	6.359172623	7.744710006	8.639868011
Curve 38	3.435352716	6.191957253	5.672570181	6.730124509	7.450925632	7.895034816
Curve 39	3.102488026	5.794955353	6.664480281	6.84795488	8.001013292	9.411551387
Curve 40	2.91622621	3.991035941	5.675270707	6.597558593	7.377390602	8.002685483
Curve 41	3.105743987	5.963134917	7.179440318	7.077913378	7.022057194	8.786334874
Curve 42	2.913956034	4.768557766	4.784895289	6.860093763	7.749198483	8.107652961
Curve 43	3.568341543	5.271093909	7.525475809	7.900701699	8.756857591	9.807464017
Curve 44	3.236580825	5.947605467	6.116895134	8.326795031	8.591491723	9.131342748
Curve 45	3.177628748	5.764706699	6.752257831	6.6920315	7.954243284	8.88538803
Curve 46	3.219948784	5.07235647	6.208929724	5.776241587	8.055974074	9.041770428
Curve 47	3.219948784	5.07235647	6.208929724	5.776241587	8.055974074	9.041770428
Curve 48	2.727226047	4.184757001	6.16691396	7.144224498	7.504797032	8.926313183
Curve 49	3.108772838	6.137230445	5.293288541	6.374986934	7.822250878	8.440440298
Curve 50	3.961125237	4.730805924	6.232435898	8.451861408	8.415232594	9.317525405

Summary Statistics for Functional Data

1. Mean and Standard Deviation Functions

Functional mean and standard deviation obtained as the average and the standard deviation of the sample curves pointwise across replications, respectively. Confident bounds for the mean are computed pointwise as the mean +/- two times the standard deviation

Figure 3: The mean function

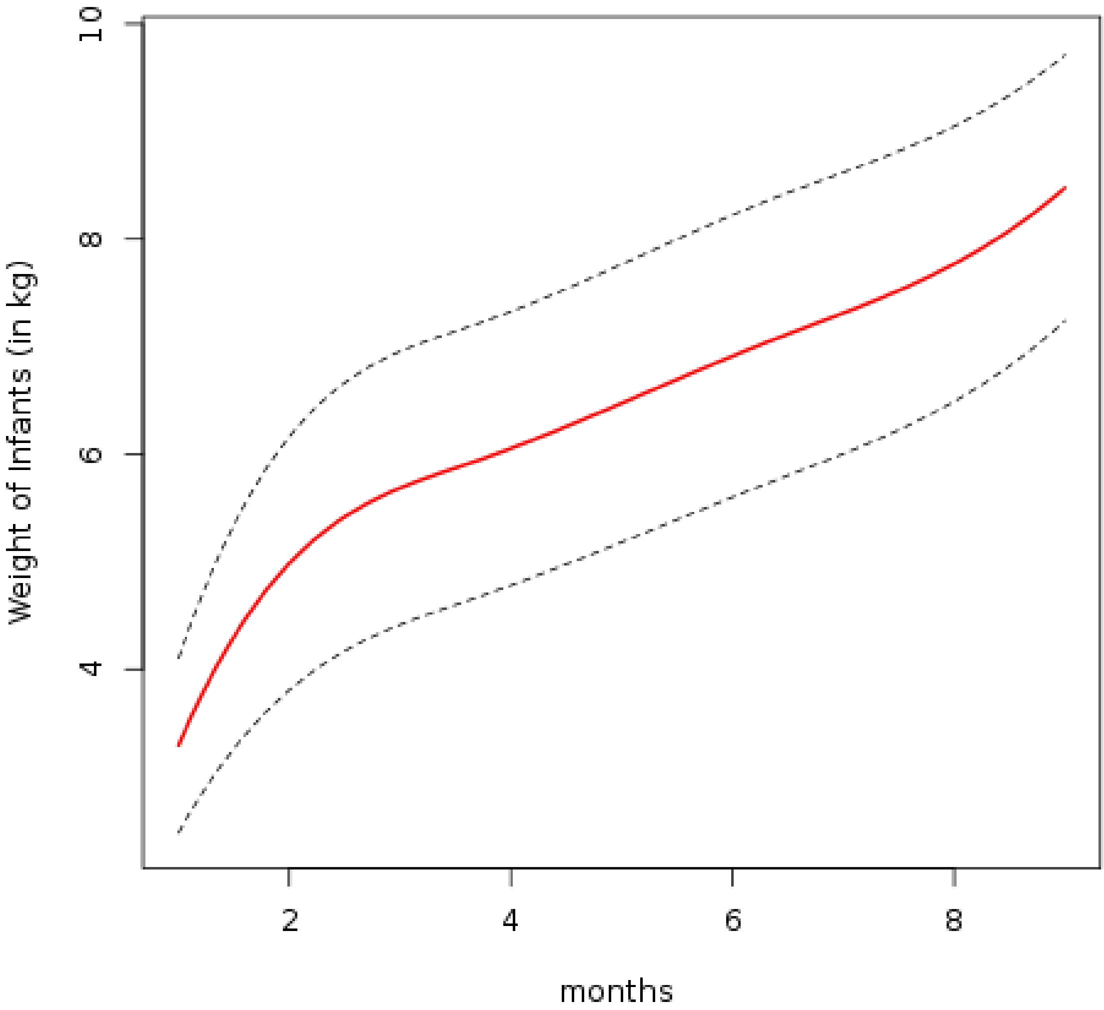

Figure 4: the meaning of the sample curves

Table 3: Mean function for the functional components

	bspl4.1	**bspl4.2**	**bspl4.3**	**bspl4.4**	**bspl4.5**	**bspl4.6**
mean	3.293953	5.430732	5.798581	7.137768	7.678772	8.475188

2. **The standard deviation**

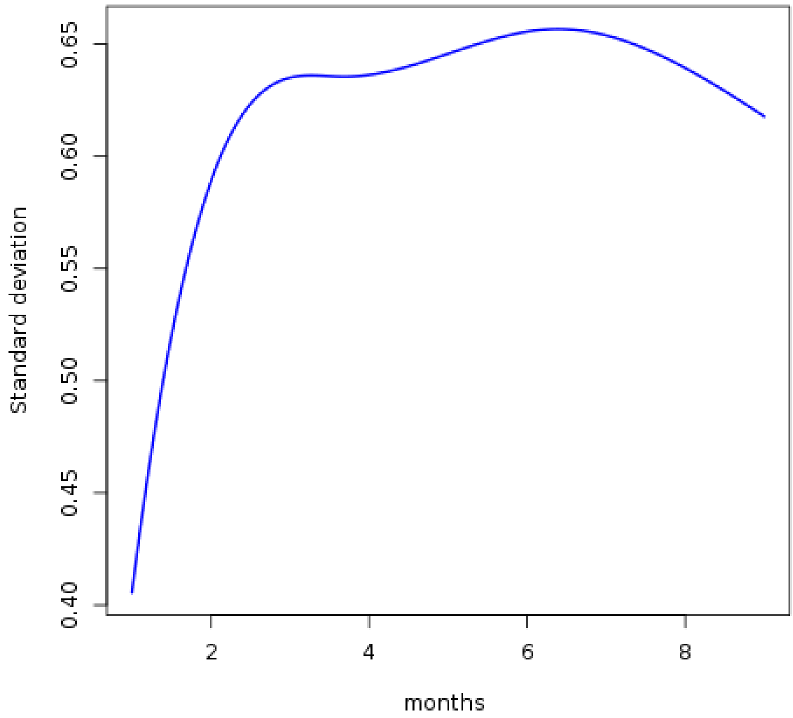

Figure 5: The standard deviation function

3. **Bivariate correlation function**

The correlation function summarizes the dependence across different arguments values and is given by a surface over the plane of possible pairs of times and also as a set of level contours

Figure below shows the surface plot

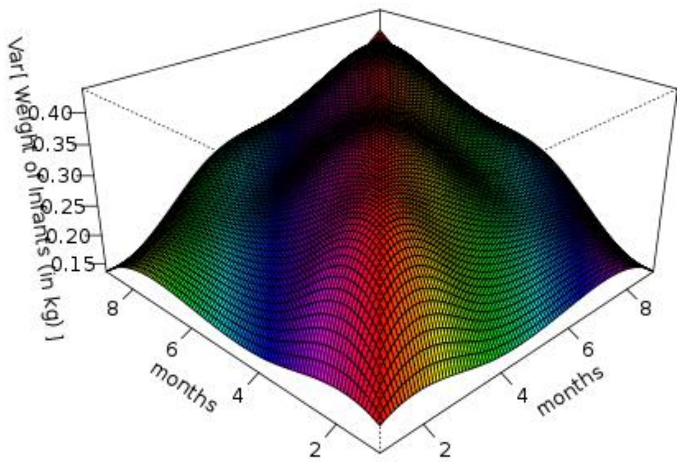

Figure 6: surface plot

Plot below show the contour plot

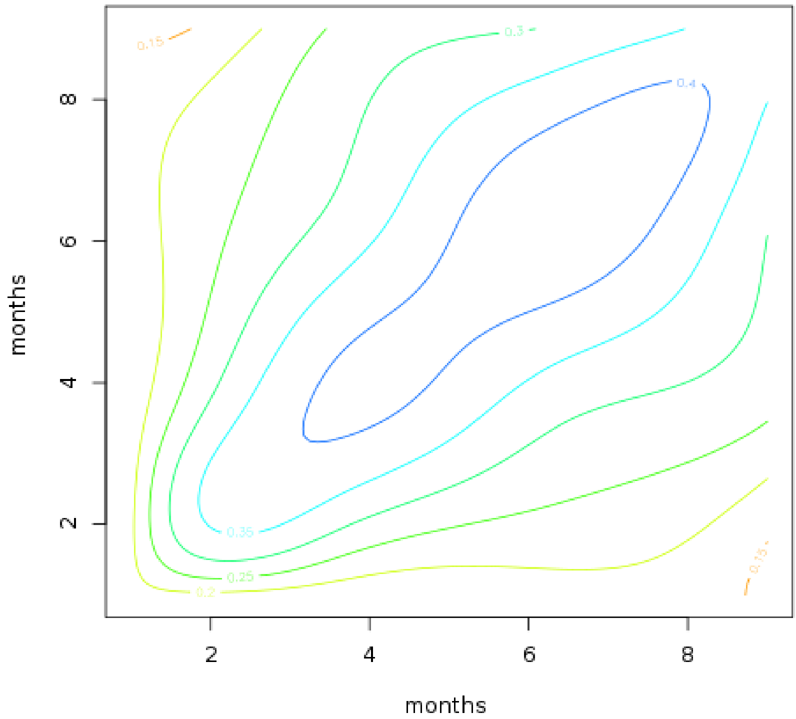

Figure 7: Contour plot

Explained variances

Variances explained by the principal components computed as the eigenvalues of the sample covariance function

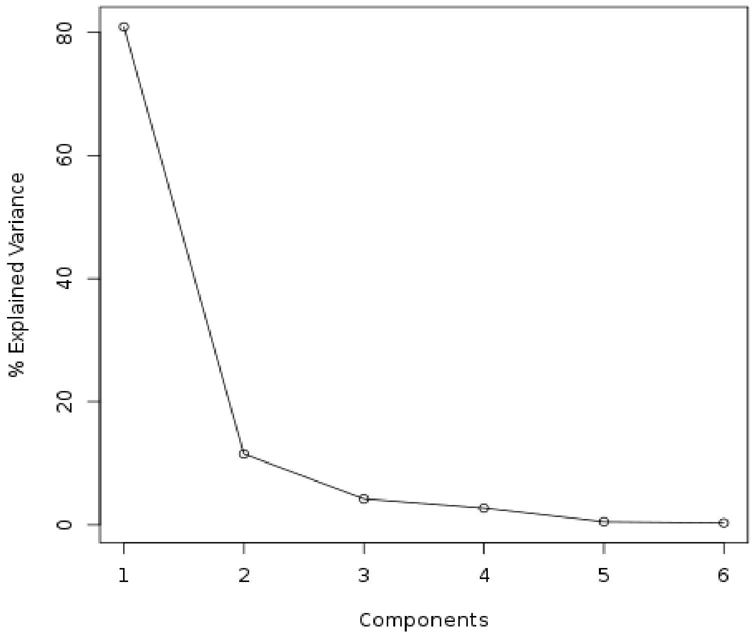

Figure 8: Graph of the explained variation

Table 4: Variance, percentage explained variance and percentage cumulative explained variance

	Variance	% Exp. Var	% Cum. Exp. Var
Comp.1	2.484277	80.9	80.9
Comp.2	0.351625	11.5	92.4
Comp.3	0.127896	4.2	96.6
Comp.4	0.082652	2.7	99.3
Comp.5	0.014073	0.5	99.8
Comp.6	0.009663	0.3	100.1

Principal Component Curves (Eigenfunctions)

Plot of principal component curves computed as the eigenfunctions of the sample covariance function and plot of the mean function and the functions obtained by adding and subtracting a suitable multiple of the principal component curve (perturbation of the mean)

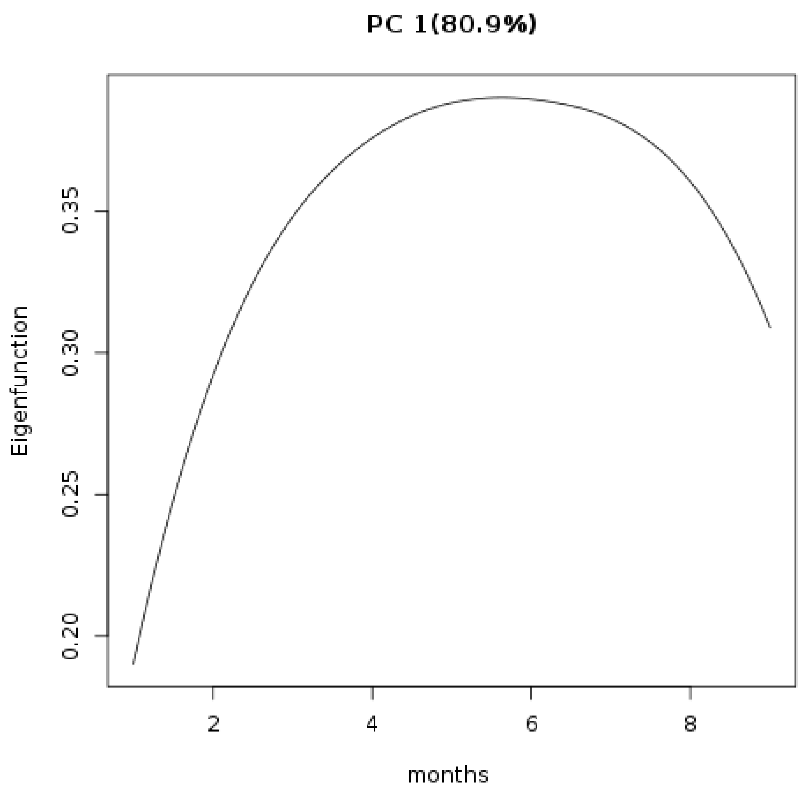

Figure 9: PC curve

Perturbation of the mean

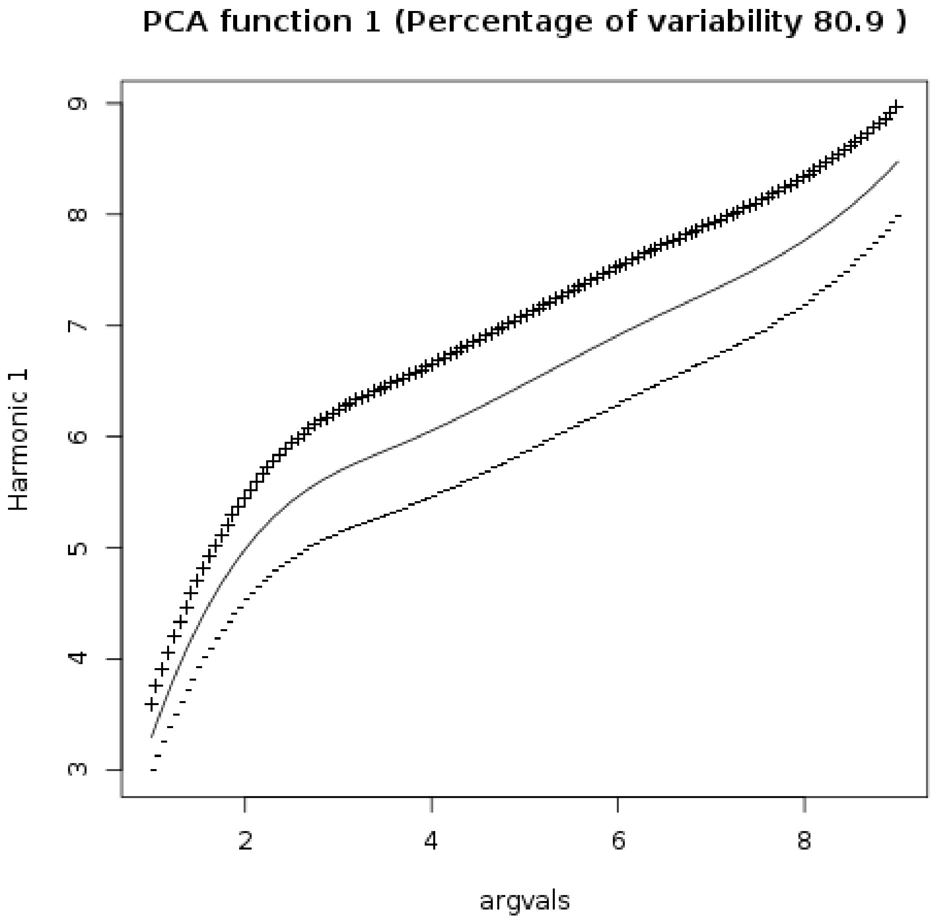

Figure 10: perturbation of mean

Basis coefficient of PC curves

Table 5: The table below shows the PC curve basis coefficient

	PC1	PC2	PC3	PC4	PC5	PC6
bspl4.1	0.190096	-0.12837	0.489334	-0.13998	1.422084	-1.29559
bspl4.2	0.306373	-0.77842	0.900814	-0.52582	-0.86658	1.581047
bspl4.3	0.38768	-0.34522	-1.06599	1.166215	0.087343	-1.65337
bspl4.4	0.395399	0.443754	0.038672	-1.42731	0.554018	1.675772
bspl4.5	0.371681	0.382592	0.516556	0.53178	-1.19625	-1.66531
bspl4.6	0.308834	0.342892	0.165641	0.964757	1.044499	1.323036

FPC expansion

Representation of a curve in terms of the first selected functional principal components. Basis coefficients of FPC expansion for all curves

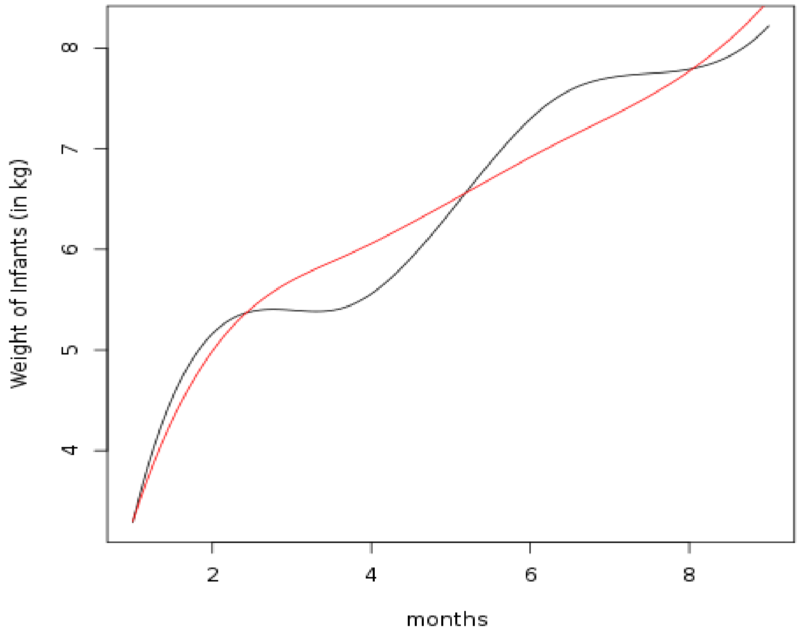

Figure 11: basis expansion

Confirmatory Analysis

Fitted model

Summary of the estimated parameters in terms of the selected principal components and their statistical significance

Table 6: Functional linear regression estimates

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	25.5	1.88	13.564	< 2e-16 ***
`Comp. 1`	1.546	1.193	1.296	0.20183
`Comp. 2`	-2.649	3.17	-0.835	0.40812
`Comp. 3`	-5.294	5.257	-1.007	0.31954
`Comp. 4`	22.894	6.539	3.501	0.00109 **
`Comp. 5`	6.696	15.848	0.422	0.67477
`Comp. 6`	-5.78	19.125	-0.302	0.76395

Functional parameter

Plot of the regression coefficient function of the functional linear mode

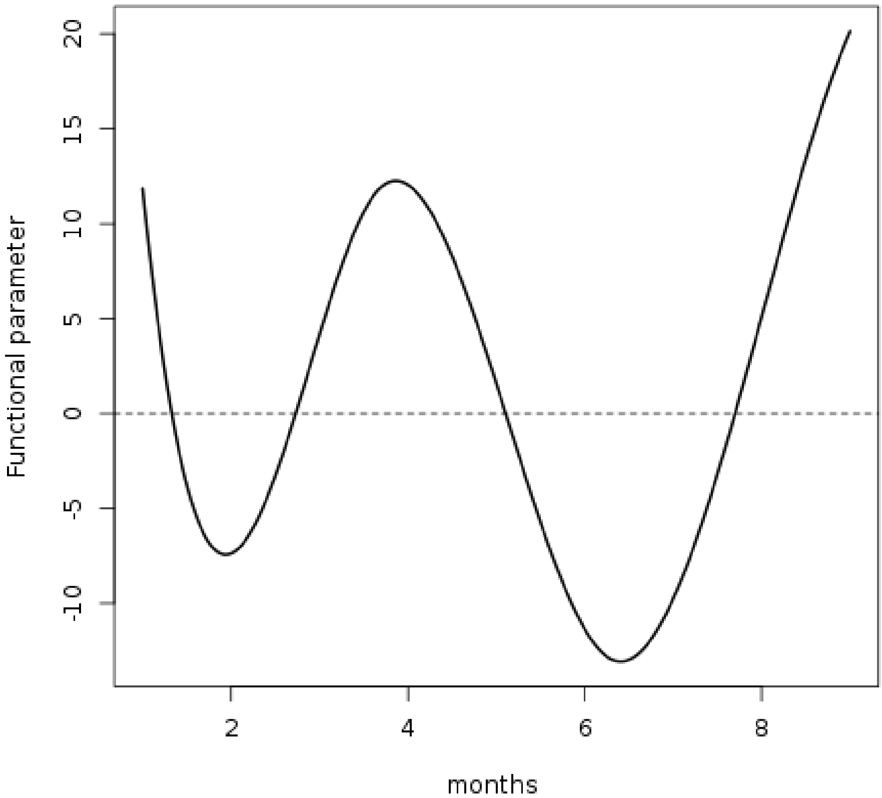

Figure 12: Regression coefficient plot

Table 7: Basis coefficients of functional parameters

	V1
bspl4.1	11.84838
bspl4.2	-29.2121
bspl4.3	43.99743
bspl4.4	-39.4218
bspl4.5	10.61668
bspl4.6	20.12643

Residual analysis

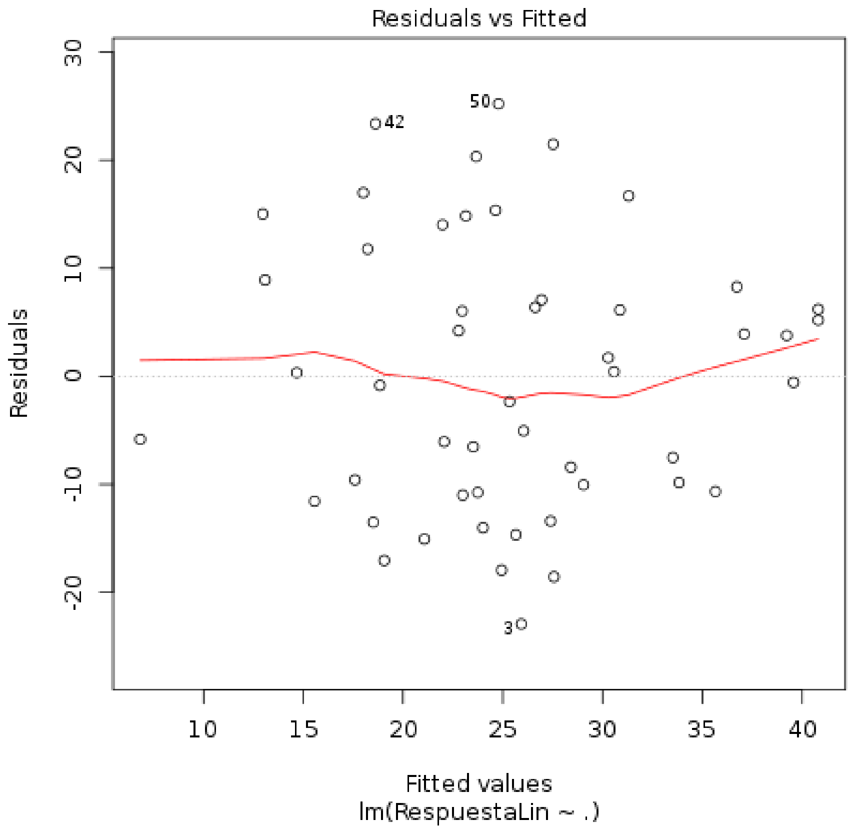

Figure 13: plot of residual vs fitted

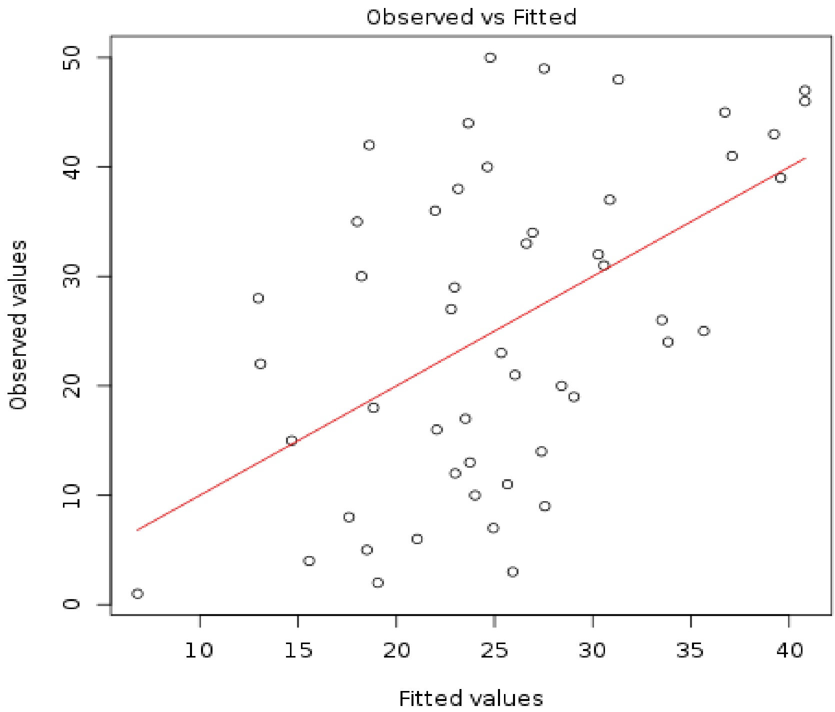

Figure 14: plot of observed vs fitted values

4.0 DISCUSSION

A multi-child weight growth curve observed on months t=1 ,…, 9 for child n= 1,…50. We are particularly interested in modeling the relationship between children's weights for the children sampled and how these relationships vary over time. Existing functional Data Analysis methods are inadequate to model the dynamic dependences among and between the curves for different children such as contemporaneous dependence, volatility clustering, covariates and change points. Our approach resolves these inadequacies and provides insights among children from childbirth to nine months old.

The figures above display the observations which are sample curves that come from a stochastic process in continuous time. Despite their continuous nature, the sample curves are unequally spaced and different among the sample units. Because of this it is necessary to reconstruct the true functional form of each sample curve from a finite set of discrete observations. The sample curves are hence represented in terms of basic functions, and the basis coefficient are fitted by least square approximation. These curves are estimated by B-spline basis. There are three basic B-spline basic least square approximation approaches namely, the Regression Spline, the smoothing spline and the P-spline.

The first step in Functional Data Analysis is to reconstruct the functional form of the sample curves from their discrete observations. The most usual way to solve this problem consists of assuming an expansion of each sample curve in terms of a basis of functions and fitting the basis coefficients using smoothing

Dynamic regression can be formulated in very general terms by using a state space representation of the observations and the hidden state of the system. With sequential definition of the processes, having conditional dependence only on the previous time step, the classical recursive Kalman filter algorithms can be used to estimate the model states given the observations

5.0 Conclusion

Functional data analysis does not interpret measured data as a sequence but as a curve. These curves vary from the traditional time series in the fact that the derivatives of a function can be evaluated at any point in time and measurements do not have taken at the sample point in time to be comparable.

Since functions are continuous, we estimate the underlying functions of the data by using the basis expansion technique where measurements are represented via a linear combination of basis functions. B-spline with a dimension of 6 was used to smoothen (reduce the noise and dimension) and estimate the underlying functions of our data.

The functional Principal Component Analysis (FPCA) was the technique employed to analyze functional data, reduce dimensionality and measure relevant information. The Functional Principal Component Analysis was used to find the principal components which maximize the variance of the data along their direction. Each principal component explains a portion of the total variance in the data and builds an orthonormal basis which means each data point as a linear combination of our principal components.

Growth models offer a plethora of exciting opportunities for testing theoretically derived hypotheses in ways not previously possible. Despite the strength and flexibility of these methods, even greater care must be taken to ensure that the estimated growth model maximally corresponds to the underlying developmental theory. Analytical results show that growth models are typically characterized by much higher levels of statistical power than comparable traditional methods applied to the same data. Any disjoint that exists between the theoretical model and the statistical model only serves to undermine our ability to draw empirically informed conclusions about our theory under study. Despite this caveat, growth models have a tremendous amount to offer to a broad array of developmental research endeavors and represent a powerful set of tools to help us continue to propel forward as a science.

Author contributions:

- Collins Aondona Ortese: Conceptualization, Methodology, Writing—original draft, Writing—review and editing;
- Nwaosu Chigozie Sylvester : Writing—review and editing;

Data Availability: The data that support the findings are available from the corresponding author upon request.

Acknowledgments: The authors sincerely thank the Editor and anonymous reviewers of the Journal of Statistical Sciences and Computational Intelligence for their timely, thorough and constructive feedback, which has significantly enhanced the clarity and overall quality of this manuscript. Thanks to God and our dear families for their support and goodwill.

Conflict of interest: The authors declare no conflict of interest.

REFERENCES

1. Aguilera, A., Ocana, F. and Valderrama, M. (1999). Forecasting Time Series by Functional PCA.Discussion of Several Weighted Approaches. *Computational Statistics*, *14*(**3**):443–467
2. Aue, A., Norinho, D. D., and Hormann, S. (2015). On The Prediction of Stationary Functional Time Series. *Journal of the American Statistical Association*, ***110***(509):378-392.
3. Bosq, D. (2000). *Linear Processes In Function Spaces: Theory and Applications,* volume 149. *Springer*
4. Bryk A.S. and Raudenbush S.W. (2019) Application of hierarchical linear models to assessing change. *Psychological Bulletin*. 101:147–158.
5. Duncan, TE and Strycker, LA.(2016) An introduction to latent variable growth curve modeling: Concepts, issues, and applications. 2. Mahwah, NJ: Lawrence Erlbaum Associates; 2006.
6. Gao, Y., Shang, H. L. and Yang, Y. (2019). 'High-Dimensional Functional Time Series Forecasting: An Application to Age-Specific Mortality Rates', *Journal of Applied Statistics*, ***41***(1):142-163.
7. Gasser T, Muller HG, Kohler W, Molinari L, and Prader A. (2017). Nonparametric regression analysis of growth curves. *Annals of Statistics* ***12***(1):210-229
8. Gasser T and Kneip A. (1995). Searching for Structure in Curve Samples. *Journal of the American Statistical Association 90*(**1**):1179-1188

9. Grenander U. (1950).Stochastic Processes and Statistical Inference. *Arkiv for Matematik **1***(1):195-277
10. Hormann S. and Kokoszka P. (2020).Weakly Dependent Functional Data. *The Annals of Statistics*, ***38***(3):1845-1884.
11. Hormann S. and Kokoszka P.(2022). Functional Time Series. Handbook of Statistics: *Time Series Analysis-Methods and Applications*, ***30***(1):157-186.
12. Hormann, S., Kidzinski, Ł. And Hallin, M. (2015), 'Dynamic Functional Principal Components', *Journal of the Royal Statistical Society*: Series B (Statistical Methodology) *77*(2), 319–348.
13. Klepsch, J. and Kluppelberg, C. (2017).An Innovations Algorithm for the Prediction of Functional Linear Processes. *Journal of Multivariate Analysis **155***(1): 252–271.
14. Klepsch, J., Kluppelberg, C. and Wei, T. (2017), 'Prediction of Functional ARMA Processes with and Application to Traffic Data', *Econometrics and Statistics*1, 128–149.
15. Kokoszka, P. (2022). Dependent Functional Data. *International Scholarly Research Network: ISRN Probability and Statistics* Volume 2012, doi:10.5402/2012/958254
16. Ramsay J.O. and Dalzell C. (2021).Some Tools for Functional Data Analysis. *Journal of the Royal Statistical Society*. Series B (Methodological) ***12***(1):539-572
17. Ramsay, J. and Silverman, B. (2005).Functional Data Analysis. *Springer*.
18. Ramsay, J. O. and Silverman, B. W. (2005). Functional Data Analysis, 2nd edition, Springer, New York.
19. Ramsay, J. O. and Silverman, B. W. (2007).Applied Functional Data Analysis*: Methods and Case studies*. Springer.MR2168993
20. Rao C. R. 1958. Some Statistical Methods for Comparison of Growth Curves. *Biometrics* 14:1-17
21. Shang, H. L. (2016), 'Mortality and life expectancy forecasting for a group of populations in developed countries: A multilevel functional data method', *The Annals of Applied Statistics **10***(3), 1639–1672.
22. Wang G., Zhou J., Wu W., and Chen M (2015) Robust functional sliced inverse regression. *Stat Papers*, Doi: 10.1007/s00362-015-0695-x