Dimensionality Reduction of Multivariate Functional Data Using FPCA

Alaa H. Jalob[1] and Lekaa Ali Mohamed[2]
[1]*Department of Studies and Planning, Al- Karkh University of Science, 10001 Baghdad, Iraq*
[2]*University of Baghdad, 10001 Baghdad, Iraq*
alaahasaan@kus.edu.iq, lekaa, a@coadec.uobaghdad.edu.iq

Keywords: Functional Data Analysis, Multivariate Functional Data, Orthonormal Basis Expansion, Functional Principal Component Analysis, Covariance Function, Dimensionality Reduction.

Abstract: In many scientific and engineering applications, data are collected over discrete, often equidistant, time intervals. While such data can be analyzed using traditional statistical methods, these methods are often very limited in their ability to capture the underlying continuous nature of dynamic processes of the phenomenon under study. This study aims to present and develop a statistical methodology for analysing multivariate functional data characterised by structural complexity, nonlinear properties, and continuous nature of its observations and variables. By using the Functional Principal Component Analysis (FPCA) method to analyse high-dimensional data through dimensionality reduction, and accurately discovering structural patterns in the data without relying on fixed distributional assumptions. Also, improving model selection using the Bayesian Information Criterion (BIC), which determines the optimal number of orthogonal basis functions and enhances the model's fit to the complexity of the data, thereby contributing to the accuracy of the analysis and understanding of relationships within the functional data.

1 INTRODUCTION

The rapid technological advancement in our current era has made the field of advanced data analysis a fundamental pillar for making strategic decisions in various scientific and practical fields, especially with the increase in the volume and complexity of available data. While much of the current research focuses on traditional data analysis methods, these approaches often struggle to handle the multidimensional and complex nature of modern datasets. Previous studies have not fully addressed the challenges associated with analyzing continuous functional data, which are characterized by their temporal or spatial nature. These data are typically represented as curves or functions instead of discrete points, requiring advanced analytical methods capable of capturing their complexities.

For example, functional data is represented in applications across various fields such as radio signal processing, where signals are continuously measured over time, monitoring brain electrical activity (EEG), analyzing growth rates over time, as well as climate changes like continuous temperature readings. Therefore, functional data analysis is of utmost importance. These examples highlight the necessity of using advanced techniques that can handle the continuous and overlapping nature of this data.

In particular, Functional Data Analysis (FDA) has emerged as a fundamental tool for handling these continuous data [1]-[3], yet its application remains limited in some fields, especially in the context of multivariate analysis. It has become increasingly necessary to use analytical techniques that can effectively manage the nonlinear, intertwined, and high-dimensional structure of functional data.

Despite the development of some methods, such as Functional Principal Component Analysis (FPCA), to reduce dimensions and discover principal components, there remains a gap in the ability to ideally determine the number of basic functions and fit the model to the complexity of the data.

This study aims to address these gaps by presenting a statistical approach for Functional Principal Component Analysis (FPCA), utilizing the Bayesian Information Criterion (BIC) for optimal selection of the number of orthogonal basis functions. By improving the methodology of functional data analysis, this study contributes to the advancement of multivariate analysis of functional data, providing

more accurate and efficient tools for handling complex datasets. The value of this research lies in its ability to provide a more flexible and statistically robust approach to functional data analysis, thereby expanding the potential applications of this technique in various scientific and practical fields.

Functional data analysis (FDA) is considered the fundamental basis for analyzing random processes of phenomena that change continuously and over short time intervals within a specific time frame, where each observation is represented as a continuous function composed of a set of connected points. From here, the method of multivariate functional principal component analysis (FPCA) emerged as an analytical tool to handle this continuous nature, unlike the traditional PCA method which deals with discrete standard data [4].

The traditional covariance matrices used in conventional analysis methods differ from the variance function, which plays an important role in functional principal component analysis due to its ability to handle data with intertwined or irregular nature [1], [5]. There are three basic strategies for estimating functional principal components identified et al. [6], which are:

1) Using the sample variance function. Using the sample variance function.
2) The estimation directly from the curves fitted for each function individually. Estimation directly from the curves suitable for each function individually.
3) Using the Qatari version of the variance function.

These techniques flexibly deal with continuous stochastic processes that change over time and within a specific time frame, such as electroencephalography, electrocardiography, stock prices, securities, and environmental changes. Wei et al. [7] also proposed a new methodology to improve the selection of an appropriate model that fits the complexity of continuous data and made adjustments to the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) for dual modelling of functional data. The importance of this study lies in providing efficient computational procedures for selecting the number of principal components of functions and tuning parameters using partial least squares regression. This helps in determining the model's accuracy and computational procedures compared to traditional methods such as cross-validation. The numerical results show that the use of modified information criteria yields results similar to or close to traditional methods with a significant reduction in time, effort, and cost, making them suitable for real-world scientific applications such as environmental, biological, and medical data analysis [8].

2 RESEARCH METHODOLOGY

The methodology transforms discrete data (observations) into continuous data (functions) using orthogonal basis function representations. We assume a stochastic process for a continuous phenomenon that varies over time and within a specified time domain in a Hilbert space [2]. The coefficients are estimated using the least squares method. FPCA is used to reduce the dimensionality while preserving the maximum amount of explained variance in the data [9]-[10]. The Bayesian information criterion (BIC) is then used to determine the model's fit and accuracy to the inherent complexity of the data. The final model captures the hidden relationships and dynamics of the original data and the essential components for interpreting the data.

3 MATERIALS AND METHODS

3.1 Transformation of Discrete Data to Multivariate Functional Data

In many scientific applications and fields, data are collected as discrete points for the phenomenon under study. In order to analyze the phenomenon accurately and clearly, it is necessary to represent the data as continuous functions to study and analyze the function in its entirety. This occurs especially when dealing with continuous random processes that change over time and within a specific time range. The FDA allows the analysis of functions using orthogonal basis functions and estimating their parameters. This involves using a series of orthogonal basis functions, then estimating the coefficients using the least-squares method, followed by smoothing the data and determining the appropriate model using the BIC.

3.1.1 Mathematical Representation of the Stochastic Process

Suppose we have a random process $X(t)$, where $t \in I$ and I represents the time domain. $X(t)$ belongs to the quadratic integrable Hilbert space $L^2(I)$,, in the domain of the function I. The random process is

represented as a linear combination of orthogonal basis functions $\varphi_b(t)$ of the domain I:

The process can be expressed as a sum of orthonormal basis functions [1] $\varphi_b(t)$:

$$X(t) = \sum_{b=0}^{B} c_b \, \varphi_b(t), \quad t \in I. \tag{1}$$

Where:

- $\varphi_b(t)$ are orthonormal basis functions;
- c_b are the random;
- B is the total number of basis functions.

3.1.2 Discrete Observations

The process $X(t)$ is observed at discrete time points $t_1, t_2, \ldots, t_J$, where $t_j \in I$ for $j = 1,2,\ldots,J$. Let x_j represent the observed value of $X(t)$ at time t_j. The observed data is represented by the vector:

$$\mathrm{x} = \left(x_1, x_2, \ldots, x_J\right)^T.$$

These are the discrete observations, which we will transform into a continuous function.

3.1.3 Coefficient Estimation Using Least Squares

To convert the discrete data into a continuous function, we use the least squares method to estimate the coefficients c_b. The coefficients c is estimated using the least squares method using the following formula:

$$S(\mathrm{c}) = (\mathrm{x} - \Phi(t)\mathrm{c})^T(\mathrm{x} - \Phi(t)\mathrm{c}),$$

Where:

- $\Phi(t)$ represents the matrix of basic functions at the discrete time points t_j.
- $\mathrm{c} = (c_0, c_1, \ldots, c_B)^T$ represents the vector of coefficients to be estimated.

3.1.4 Inner Product for Functional Data

The inner product is used to compute the projections of the data onto the basis functions using the least squares method. The inner product between two functions $u(t)$ and $v(t)$ is given by:

$$\langle u, v \rangle = \int_I u'(t) v(t) \, dt. \tag{2}$$

Where:

- $\langle u, v \rangle$ represents the inner product of the two functions $u(t)$
- $v(t)$, $u'(t)$ is the derivative of $u(t)$ (if needed for the particular function space)
- The integral is taken over the interval I.

The inner product is used to calculate the projection of the observed data x_j onto the basis functions $\varphi_b(t)$. to ensure the smoothness of the function.

3.1.5 Reconstructing the Continuous Function

After estimating the coefficients c, the random process $x(t)$ is characterized by the following formula:

$$x(t) = \textstyle\sum_{b=0}^{B} \hat{c}_b \, \varphi_b(t), \quad t \in I. \tag{3}$$

The continuity equation for the observed data **x**.

3.1.6 Multivariate Functional Data

Multivariate functional data are handled by expanding the underlying process. Let $X(t) = \left(X_1(t), X_2(t), \ldots, X_p(t)\right)^T$ where p is a vector of random processes that are summed to the same discrete time points.

The representation of each component $X_d(t)$ for $d = 1,2,\ldots,p$ is given by:

$$X_d(t) = \textstyle\sum_{b=0}^{B_d} c_{db} \, \varphi_b(t), \quad t \in I, \quad d = 1,2,\ldots,p.$$

Where $\varphi_b(t)$ represents the basis functions of each component, and c_{db} are the coefficients of the d-th component. The multivariate functional data is represented as:

$$\mathrm{X}(t) = \sum_{d=1}^{p} \Phi_d(t) \mathrm{c}_d, \quad t \in I. \tag{4}$$

Where:

- $\Phi_d(t)$ is the matrix of basis functions for the d-th component,
- c_d is the vector of coefficients for the d-th component.

3.1.7 Matrix Representation of Coefficients

The coefficient vector $c$and the matrix $\Phi(t)$are obtained using the following:

- Coefficient Vector c:

$$\mathrm{c} = \left[c_{1B1}, c_{1B2}, \ldots, c_{pBp}\right]^T$$

- Matrix $\Phi(t)$:

$$\Phi(t) = \begin{bmatrix} \varphi'_B 1(t) & 0 & \cdots & 0 \\ 0 & \varphi'_B 2(t) & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & \varphi'_B p(t) \end{bmatrix}.$$

Multiplying $\Phi(t)$ by the coefficient vector c reconstructs the multivariate functional data:

$$\mathrm{X}(t) = \Phi(t)\mathrm{c}. \tag{5}$$

3.1.8 Model Selection Using Bayesian Information Criterion (BIC)

To determine the optimal number of basic functions B, we use the Bayesian Information Criterion (BIC) to improve and fit the model.
The formula for BIC is:

$$BIC = \ln\left(\sum_{j=1}^{J}\left(x_j - \sum_{b=0}^{B}\hat{c}_b\,\varphi_b(t_j)\right)^2\right) + (B+1)\left(\frac{\ln J}{J}\right).$$

3.1.9 Final Transformation to Functional Data

After estimating the coefficients $\hat{\mathrm{c}}$, the continuous functional data is obtained:

$$\mathrm{X}(t) = \Phi(t)\,\hat{\mathrm{c}}. \tag{6}$$

It represents the reconstructed functional data from the discrete data.

3.1.10 Principal Component Analysis for Functional Data

MFPCA is an advanced analytical statistical method that reduces dimensions and extracts main patterns from complex data containing many functional variables. These variables are described as mathematical functions that change continuously over space or over closely spaced time intervals within a specific time frame, allowing for the analysis of the variance present within the complex data.

The mathematical steps of the MFPCA method begin by identifying the principal components that explain the greatest amount of variance in the data. This is done by finding weight functions or (weight vectors) that represent these components. Where the first principal component, which represents the largest variance in the data, is extracted, followed by finding the other components with the constraint that these components are uncorrelated with the previous components. This allows for dimensionality reduction without losing the important information contained in the data.

In applications that involve variables changing continuously over time or any other continuous data, we use the MFPCA method to analyse those variables. MFPCA allows for understanding the complex relationships between variables and discovering hidden patterns that may not be visible when analysing the variables separately at discrete data points. This is achieved by transforming the data into a new set of uncorrelated principal components, ordered from greatest variance to least, while preserving the variance of the original data. If we have a random vector $X = \left(X_1, X_2, \ldots, X_p\right)' \in \mathbb{R}^p$We look for a linear combination $U_1 = u_{11}X_1 + u_{12}X_2 + \cdots + u_{1p}X_p = u_1'X$, where X is the vector of components that has the maximum variance. Where U_1 represents the first principal component, we then find a linear combination $U_2 = u_2'X$ which is unrelated to the first component U_1 and so on until we reach the k-th stage, followed by a linear combination $U_k = u_k'X$ for the k-th principal component, which has the maximum variance and is also unrelated to the previous principal components.

PCA helps understand the structural composition of the variance in the covariance matrix of one-dimensional functional data [4], [12], [13].

3.2 Functional Principal Component Estimation

In multivariate functional PCA, we aim to find the inner product:

$$U = \langle u, X\rangle = \int_I u'(t)X(t)\,dt. \tag{7}$$

Equation (5) represents the projection of the functional data onto the vector function u, which has the largest variance for all $u \in L_2^p(I)$, where $\langle u, u\rangle = 1$, and it is assumed that:

$$\lambda_1 = \sup_{u \in L_2^p(I)} \mathrm{var}(\langle u, X\rangle) = \mathrm{var}(\langle u_1, X\rangle). \tag{8}$$

Where:

- u_1 represents the first eigenfunction;
- λ_1 represents the first eigenvalue.

The first principal component is determined as the direction that maximizes the variance, defined as $U_1 = \langle u_1, X\rangle$, where $u_1(t)$ represents the eigenfunction corresponding to the first principal component.

Finally, then find the second principal component $U_2 = \langle u_2, X\rangle$ to maximize the variance $\mathrm{var}(\langle u, X\rangle)$, provided the condition $\langle u_1, u_2\rangle = 0$ (i.e., orthogonality of the first component).

3.2.1 General Principal Components

In general, the k-th principal component $U_k = \langle u_k, X \rangle$ satisfies the following:

$$\lambda_k = \sup_{u \in L_2^p(I)} \text{var}(\langle u, X \rangle) = \text{var}(\langle u_k, X \rangle).$$

Where, $\langle u_{k_1}, u_{k_2} \rangle = \delta_{k_1 k_2}, k_1, k_2 = 1, \dots, k$

$\delta_{k_1 k_2}$ is the Kronecker delta, which is:

$$\delta_{k_1 k_2} = \begin{cases} 1 & if\ k_1 = k_2 \\ 0 & if\ k_1 \neq k_2 \end{cases}.$$

Thus, the functional principal component system $\left(\lambda_k, u_k(t)\right)$ for the process $X(t)$ is defined, and we represent the process $X(t)$ as:

$$X(t) = \Phi(t)c, \quad t \in I. \tag{9}$$

Where $E(X) = 0$, $E(c) = 0$, and $\text{Var}(c) = \Sigma$. The k-th principal component $U_k^* = \langle \omega_k, c \rangle$ for this vector satisfies:

$$\gamma_k = \sup_{\omega \in \mathbb{R}^{k+p}} \text{Var}(\langle \omega, c \rangle) = \omega' \Sigma \omega. \tag{10}$$

Where:

- ω_k is the eigenvector corresponding to the k-th principal component,
- Σ is the covariance matrix of c, The k-th principal component system $\left(\lambda_k, u_k(t)\right)$ is related to the principal component (γ_k, ω_k) of the random vector c via the equation:

$$\lambda_k = \gamma_k, \quad u_k(t) = \Phi(t)\omega_k, \quad t \in I.$$

Where, $k = 1, \dots, s$, and $s = \text{rank}(\Sigma)$.

3.2.2 Principal Component Estimation from Samples

For estimating the principal components from samples, assume that we have n independent observations $x_1(t), \dots, x_n(t)$ of the random process $X(t)$. These are of the form:

$$x_i(t) = \Phi(t)\hat{c}_i. \tag{11}$$

Where $\hat{c}_i$ are the sample coefficients, and $\hat{C} = (\hat{c}_1, \hat{c}_2, \dots, \hat{c}_n)'$.

The covariance matrix $\hat{\Sigma}$ is estimated as:

$$\hat{\Sigma} = \frac{1}{n} \hat{C}' \hat{C}. \tag{12}$$

Where, $\hat{\Sigma}$ is the sample covariance matrix.

The eigenvalues $\hat{\gamma}_1 \geq \hat{\gamma}_2 \geq \cdots \geq \hat{\gamma}_s$ are the non-zero eigenvalues of $\hat{\Sigma}$, and the eigenvectors $\hat{\omega}_1, \hat{\omega}_2, \dots, \hat{\omega}_s$ are the corresponding eigenvectors.

The principal components for the random process $X(t)$ based on the sample data are:

$$\left(\hat{\lambda}_k = \hat{\gamma}_k, \hat{u}_k(t) = \Phi(t)\hat{\omega}_k\right) \qquad k = 1, \dots, s. \tag{13}$$

4 SIMULATION OVERVIEW

4.1 Simulated Functional Data and FPCA Analysis

The simulation produces artificial multivariable functional data consisting of 10 sample curves, each observed over 50 equally spaced time points in the interval [0, 1]. Three functional components are simulated to represent distinct dynamical patterns: a sinusoidal trend (variable 1), a cosine-based oscillation (variable 2), and a linear trend with noise (variable 3). These components simulate real-world scenarios such as physiological signals, environmental measurements, or sensor outputs. The discrete data are then transformed into smooth functional representations using Fourier basis extensions, with the number of basic functions chosen via the Bayesian Information Criterion (BIC) to balance model fit and complexity, as shown in Table 1. Functional principal component analysis (FPCA) is applied to extract the dominant variance patterns, as shown in Tables 2, 3, 4, and 5, and dimensionality is reduced by projecting each curve onto the leading principal components, as shown in Table 6.

Table 1: Optimal number of basis functions and BIC values.

VAR	1	2	3
Basis 5	0.5505	-0.4945	1.6726
Basis 6	0.5788	-0.4352	1.5277
Basis 7	0.657	-0.3569	1.6059
Basis 8	0.6645	-0.3233	1.5252
Basis 9	0.7427	-0.2451	1.6035
Basis 10	0.7814	-0.2426	1.5863
Optimal Basis	5	5	8

Table 2: Explained variance ratio by functional principal component.

VAR	FPC 1	FPC 2	FPC 3	Total Variance Explained
1	0.4579	0.2903	0.1521	0.9004 (90.0%)
2	0.3135	0.2649	0.204	0.7824 (78.2%)
3	0.3847	0.2156	0.1833	0.7836 (78.4%)

Table 3: FPC scores for all samples (Variable 1).

Sample	FPC1	FPC2	FPC3
1	0.01763	0.003475	-0.008379
2	-0.002226	-0.000664	0.002051
3	-0.006237	-0.000415	-0.001019
4	-0.002955	0.003093	-0.001727
5	-0.000268	0.017201	0.00576
6	0.016524	-0.006109	0.002501
7	-0.009067	0.00712	-0.004701
8	-0.000934	-0.006015	0.012101
9	-0.014809	-0.00873	-0.004964
10	0.002341	-0.008957	-0.001623

Table 4: FPC scores for all samples (Variable 2).

Sample	FPC1	FPC2	FPC3
1	0.005197	0.004723	0.007229
2	-0.003033	0.003634	-0.000094
3	0.006536	-0.007081	-0.000759
4	-0.002312	0.003863	0.004948
5	0.00017	0.00388	-0.000792
6	-0.000846	-0.003144	-0.005397
7	0.000876	-0.008372	0.004618
8	-0.013045	-0.002258	-0.000124
9	0.003094	-0.000935	-0.003169
10	0.003365	0.00569	-0.00646

Table 5: FPC scores for all samples (Variable 3).

Sample	FPC1	FPC2	FPC3
1	0.004874	-0.000342	0.008137
2	-0.000907	-0.003172	-0.003756
3	-0.003288	0.004258	0.001245
4	-0.000332	0.006735	-0.005194
5	-0.007451	-0.001239	0.001725
6	0.006316	-0.004657	-0.000756
7	-0.001544	-0.002391	-0.00283
8	-0.006007	-0.00526	0.000261
9	-0.002004	0.005087	0.003075
10	0.010343	0.000979	-0.001905

4.2 Analysis of Results

The results demonstrate a successful application of Functional Data Analysis (FDA) from data representation to dimensionality reduction. First, the Bayesian Information Criterion (BIC) effectively selects the optimal number of Fourier basis functions: 5 for Variables 1 and 2, and 8 for Variable 3. This is consistent with the nature of the variables: Variables 1 (sinusoidal) and 2 (cosine-based) are smooth and periodic, so they can be accurately represented with fewer basis functions, whereas Variable 3 follows a linear trend with added noise, which is less periodic and requires more basis functions (8) to capture its shape accurately. This demonstrates that BIC adapts to the complexity of each functional variable, avoiding both underfitting and overfitting.

After functional representation, Functional Principal Component Analysis (FPCA) is applied. The results show that the first three principal components explain a high percentage of total variation: 90.0% for Variable 1, 78.2% for Variable 2, and 78.4% for Variable 3 (see Figure A4 in the Appendix). Figures A1-A3 in the Appendix illustrate the functional principal components for each variable, highlighting the main patterns captured by FPCA. This indicates that most of the information in the original data is preserved in just a few components, enabling significant dimensionality reduction without major loss of insight.

The first component (FPC1) typically represents the main pattern – such as amplitude in oscillatory signals or overall slope in linear trends – while FPC2 and FPC3 capture secondary variations like phase shifts, curvature, or local fluctuations. Most samples have values close to zero when examining FPC scores, which indicates they follow the average pattern. However, some samples show varying degrees of deviation: for example, sample 6 in Variable 1 has a high positive FPC1 score, indicating a curve with above-average amplitude, while sample 9 shows the opposite. In Variable 3, sample 10 has the highest FPC1 score, reflecting a steeper upward trend, whereas sample 5 is flatter. Other deviations, such as the high FPC3 scores in samples 1 and 8, indicate unique local behaviours, including late peaks or noise patterns.

In general, the final reduced data matrix (10×9) consolidates all the scores into an interpretable and visualisable compact representation, where each sample is described by nine meaningful features, three for each variable. Table 6 presents this matrix, combining the top three FPC scores from each of the three functional variables into a single 9-dimensional feature space. This transformation reduces the original $10 \times 50 \times 3 = 1500$ discrete observations into a compact, structured, and interpretable format. The integration of scores allows for multivariate analysis – for example, detecting samples that simultaneously exhibit high amplitude in Variable 1, low offset in Variable 2, and steep slope in Variable 3. Sample 10, for instance, shows positive FPC1 scores in all three variables, indicating an above-average profile in overall magnitude, phase, and trend. In contrast, Sample 9 exhibits mostly negative scores, representing a below-average profile. Reducing high-dimensional data to a low-dimensional representation preserves the fundamental dynamics of the system and is crucial for clustering, classification, and other downstream analyses, demonstrating the practical value of FDA.

Table 6: Final reduced data matrix (10 × 9).

	V1-FPC1	V1-FPC2	V1-FPC3	V2-FPC1	V2-FPC2	V2-FPC3	V3-FPC1	V3-FPC2	V3-FPC3
1	0.0176	0.0035	-0.0084	0.0052	0.0047	0.0072	0.0049	-0.0003	0.0081
2	-0.0022	-0.0007	0.0021	-0.003	0.0036	-0.0001	-0.0009	-0.0032	-0.0038
3	-0.0062	-0.0004	-0.001	0.0065	-0.0071	-0.0008	-0.0033	0.0043	0.0012
4	-0.003	0.0031	-0.0017	-0.0023	0.0039	0.0049	-0.0003	0.0067	-0.0052
5	-0.0003	0.0172	0.0058	0.0002	0.0039	-0.0008	-0.0075	-0.0012	0.0017
6	0.0165	-0.0061	0.0025	-0.0008	-0.0031	-0.0054	0.0063	-0.0047	-0.0008
7	-0.0091	0.0071	-0.0047	0.0009	-0.0084	0.0046	-0.0015	-0.0024	-0.0028
8	-0.0009	-0.006	0.0121	-0.013	-0.0023	-0.0001	-0.006	-0.0053	0.0003
9	-0.0148	-0.0087	-0.005	0.0031	-0.0009	-0.0032	-0.002	0.0051	0.0031
10	0.0023	-0.009	-0.0016	0.0034	0.0057	-0.0065	0.0103	0.001	-0.0019

5 CONCLUSIONS

The simulation results of our research show that functional data analysis (FDA) is a useful tool for converting high-dimensional, multivariate data into a simple, clear, low-dimensional representation while retaining the dynamic characteristics of the original data. That the criterion of information-theoretic virtualization turned out to optimally choose the number of base States, thereby achieving a kind of balance in the clarity of the model, balancing the accuracy of the model and complexity.

Functional principal component analysis explained the bulk of the data variance (between 78% and 90%) based on only three principal components, demonstrating the model's ability to capture primary patterns such as oscillations, linear trends and local variations.

6 RECOMMENDATIONS

In future research, researchers recommend conducting sensitivity analysis through simulated experiments or real data using various sample sizes and variable time intervals. This process will help to identify and limit the possibility of stability and accuracy of the model, in addition to increasing the researchers ' confidence in the practical application of the model across a variety of data situations. Sensitivity testing can also help in the real detection of any problems with generalization or overgeneralization of these cases, which directs to the improvement of the models and methods used.

REFERENCES

[1] J. O. Ramsay and B. W. Silverman, Functional Data Analysis, 2nd ed. Springer, 2005.

[2] T. Hsing and R. Eubank, Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators. Wiley, 2015.

[3] J. Jacques and C. Preda, “Functional data analysis: A review,” Computational Statistics & Data Analysis, vol. 71, pp. 1–15, 2014.

[4] I. T. Jolliffe, Principal Component Analysis, 2nd ed. Springer, 2002.

[5] L. Xiao, V. Zipunnikov, D. Ruppert, and C. Crainiceanu, “Fast covariance estimation for high-dimensional functional data,” Statistics and Computing, vol. 26, no. 1, pp. 409–421, 2016.

[6] K. Chen and J. Lei, “Localizing the functional principal components analysis of irregularly sampled longitudinal data,” Stat, vol. 6, no. 1, pp. 147–161, 2017.

[7] J. Wei, L. Zhou, and L. Zhang, “Model selection using modified AIC and BIC in joint modeling of paired functional data,” Statistics & Probability Letters, vol. 80, no. 23–24, pp. 1918–1924, 2010.

[8] H. Helena, R. Pardo, M. Vega, E. Barrado, J. M. Fernandez, and L. Fernandez, “Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis,” Water Research, vol. 34, no. 3, pp. 807–816, 2000.

[9] T. Górecki, M. Krzyśko, Ł. Waszak, and W. Wołyński, “Selected statistical methods of data analysis for multivariate functional data,” Statistical Papers, vol. 59, no. 1, pp. 153–182, 2018.

[10] X. Dai and H.-G. Müller, “Principal component analysis for functional data on Riemannian manifolds and spheres,” The Annals of Statistics, vol. 46, no. 6A, pp. 3334–3361, 2018.

[11] Y. Lu, X. Zhou, E. Cui, D. Rogers, C. M. Crainiceanu, J. Wrobel, and A. Leroux, “Generalized conditional functional principal component analysis,” arXiv preprint arXiv:2411.10312, Nov. 2024.

[12] G. Aneiros, R. Cao, R. Fraiman, C. Genest, and P. Vieu, "Recent advances in functional data analysis and high-dimensional statistics," Journal of Multivariate Analysis, vol. 170, pp. 3–9, 2019.

[13] K. Kim and P. Piotr, "Extremal dependence measure for functional data," Journal of Multivariate Analysis, vol. 184, p. 104746, 2022.

APPENDIX

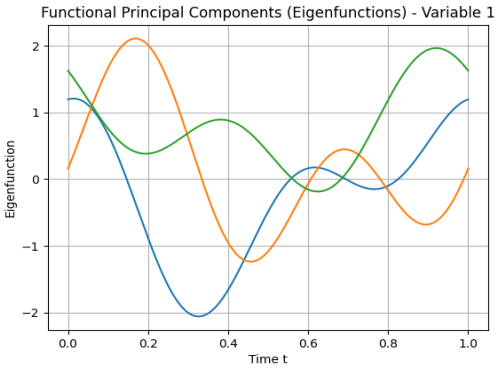

Figure A1: Functional principal components for Variable 1.

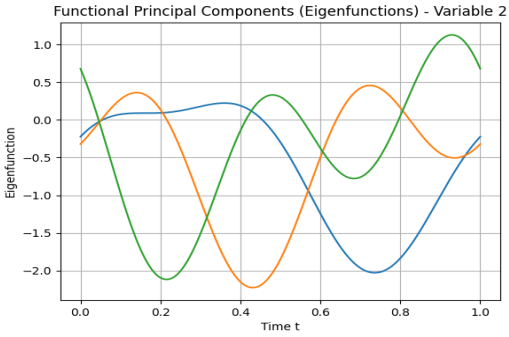

Figure A2: Functional principal components for Variable 2.

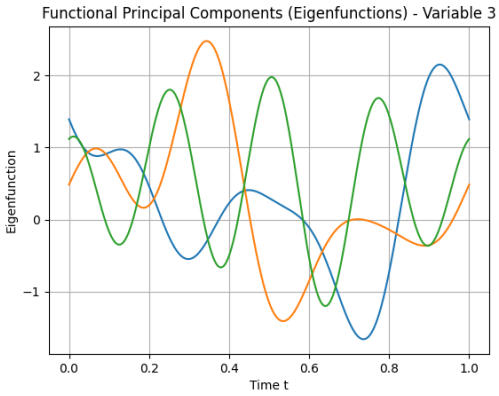

Figure A3: Functional principal components for Variable 3.

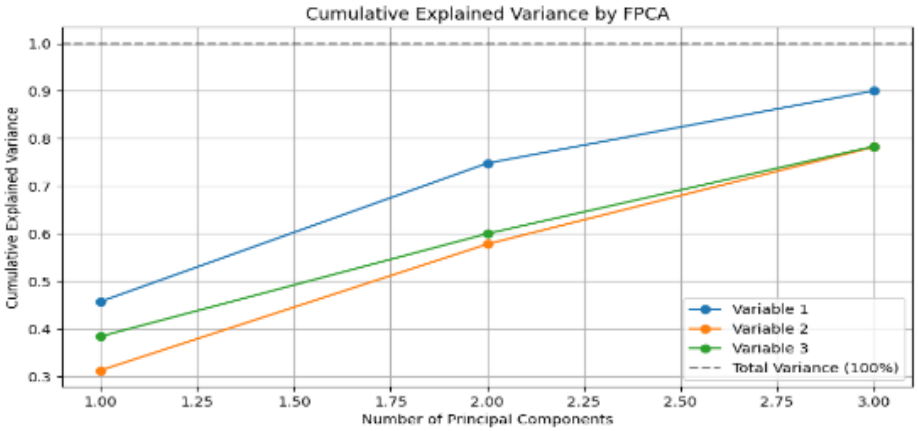

Figure A4: Cumulative explained variance by FPCA.