JIRSS (2025)
Vol. 24, No. 01, pp 1-10
DOI: 10.22034/jirss.2025.2021358.1050

Functional Principal Component Analysis of Intracranial Pressure Data

Arezoo Orooji[1,2], Mohammad Hosseini-Nasab[3], Hassan Doosti[4], Human Baharvahdat[5], Habibollah Esmaily[1,2]

[1] Social Determinants of Health Research Center, Mashhad University of Medical Sciences, Mashhad, Iran.
[2] Department of Biostatistics, School of Health, Mashhad University of Medical Sciences, Mashhad, Iran.
[3] Department of Statistics, Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran.
[4] Department of Mathematics and Statistics, Macquarie University, 2109 Sydney, Australia.
[5] Department of Neurosurgery, Ghaem Hospital, Mashhad University of Medical Sciences, Mashhad 99199-91766, Iran.

Received: 11/02/2024, Accepted: 06/06/2025, Published online: 26/06/2025

Abstract. The term "functional data" refers to data where the units of observation are functions defined over a time interval. The fundamental philosophy behind functional data is that the repeated measurements for each individual are considered as a stochastic process over time. One of the commonly used analyses for such data is functional principal component analysis. In this study, since the intracranial pressure was measured over time in patients with aneurysmal subarachnoid hemorrhage, functional principal component analysis was employed to identify the main factors contributing to increased intracranial pressure. The first four functional principal components account for 87.8 percent of the total variation in the intracranial pressure curve. The first, second, third, and fourth principal components explain approximately 52.3, 21.9, 8, and 5.6 percent of the overall variation, respectively. These four components are linked to the total Glasgow Coma Scale score, diastolic blood pressure, age, and systolic blood pressure, respectively.

Arezoo Orooji (arezoo.orooji23@gmail.com).
Mohammad Hosseini-Nasab (M_hosseininasab@sbu.ac.ir).
Hassan Doosti (hassan.doosti@mq.edu.au).
Human Baharvahdat (humainbv@yahoo.fr).
Corresponding Author: Habibollah Esmaily (esmailyh@mums.ac.ir)

Keywords. Aneurysmal Subarachnoid Hemorrhage, Functional Data Analysis, Functional Principal Component, Intracranial Pressure Data.

MSC: 62-XX, 62Rxx, 62R10.

1 Introduction

Nowadays, with the passage of time and advancements in science and technology, data collection over time or space has become easier. This allows for denser sampling of observations over time and space Ullah and Finch (2013). Various methods have been designed to analyze such data, including different types of time series models and various longitudinal models. One of the analyses that can be used for such data collected over time is functional data analysis (FDA). With functional data, the units of observation are functions defined over a time interval. The fundamental philosophy of functional data is to consider the repeated measurements for each individual as a stochastic process over time, resulting in the observation of a function for each individual Kim (2010). In this approach, derivatives of the functions can also be used for data analysis. It is usually assumed that the observed functions belong to the separable Hilbert space $L^2[a, b]$, which indicates that they are square-integrable. In FDA, it is acknowledged that the data may be collected discretely at only a few time points. However, this approach analyzes the data by considering its underlying continuous nature rather than discrete measurements. The analysis takes into account the effect of time over the entire time interval, rather than focusing solely on discrete time points or equal intervals. This allows for a more comprehensive exploration of temporal changes, such as estimating and examining first and second derivatives Ramsay et al. (2009).

Due to the unique characteristics of functional data, such as high correlations between measurements and infinite dimensionality, classical statistical methods are not suitable for their analysis. Instead, specialized tools from functional statistics are employed. These tools are often nonparametric and rely on operator theory to handle the complexities of functional data Kim (2010).

One of the methods used for analyzing such data is Functional Principal Component Analysis (FPCA). Functional principal components (FPCs) are defined as linear combinations of the observed sample curves. Similar to classical principal components (PCs), FPCs generate a reduced number of variables using the original data that are uncorrelated and explain a substantial amount of variability, thus leading to dimension reduction. Analyzing the FPCs provides a way to investigate the covariance structure, which can offer more insight compared to examining variance-covariance matrices and correlations Ramsay (1997).

Besse and Ramsay (1986) introduced FPCA for a sample of n functions observed at p argument values. Silverman (1995) extended functional principal component analysis

(FPCA) for functional data to incorporate both functional and parametric effects. Yao et al. (2005) proposed a method for FPCA when data is sparsely collected over time. Hall et al. (2006) discussed PCA methods for analyzing functional and longitudinal data. Gervini (2008) developed robust estimators for FPCs.

Berrendero et al. (2011) proposed a PC method for analyzing multivariate functional data. Yao and Lee (2006) addressed computational aspects of FPCA and suggested an iterative procedure for estimating FPCs using penalized spline regression. Other contributions to FPCA include kernel-based smoothed FPCs discussed by Boente and Fraiman (2000).

2 Patients and Methods

The environmental pressure for blood perfusion in most parts of the human body is either low or corresponds to atmospheric pressure. However, the environmental pressure for the brain is different because the brain is encompassed and safeguarded by a rigid skull. Intracranial pressure (ICP) is derived from the circulation dynamics of cerebral blood and cerebrospinal fluid (CSF) and can be influenced during the course of various central nervous system disorders Czosnyka and Pickard (2004). Furthermore, ICP plays a critical role in the pathophysiology of aneurysmal subarachnoid hemorrhage (aSAH) (see Zoerle et al. (2020)).

The aSAH is a serious condition that carries a high risk of mortality and neurological complications Svedung Wettervik et al. (2021). Following the onset of aSAH, elevated ICP and decreased cerebral perfusion pressure (CPP) can result in early and severe brain damage, which can be fatal in some cases Imberti et al. (2021). Normally, ICP ranges between 5 and 15 mmHg, but in cases where the brain is damaged, this value can exceed 20 mmHg. Monitoring ICP is crucial after subarachnoid hemorrhage (SAH) to prevent secondary brain injuries and to customize individualized treatments Zoerle et al. (2020). Currently, ICP is measured invasively by creating a hole in the skull and inserting a catheter into the brain's ventricle. This procedure is performed selectively for patients at risk of increased ICP. The ventricles of the brain contain CSF, and when a patient's brain pressure reaches a dangerous level, doctors can drain the CSF using a catheter inserted into the brain to reduce the severity of the pressure.

In this research, individuals affected by aSAH were investigated due to its life-threatening nature. Since this condition is often accompanied by increased ICP, and measuring ICP is an invasive process, the study focused on examining factors associated with increased ICP in these patients. As ICP is measured over time, FPCA was used to analyze the data.

A total of 12 aSAH patients were included in our research. They were patients who referred to Ghaem Hospital for the period of March to October 2020. Their information was collected through health information system (HIS) and file reading at Neurosurgery Intensive Care Unit. Inclusion criteria comprised patients with definitive diagnosis of

spontaneous subarachnoid hemorrhage, placement of an External Ventricular Drain (EVD) in brain and low level of consciousness ($3 \leq$ Glasgow Coma Scale (GCS) ≤ 14). Patients were treated and their ICP was measured every 6 hours. This process lasted for 72 hours (i.e., 13 repetitions). If the number of times the ICP was measured less than 13 times, the patient was excluded.

2.1 Basis Function Systems for Constructing Functions

We need to work with functions that exhibit both unpredictable and complex features. Therefore, we require a method for constructing functions that works with parameters that are easy to estimate and can accommodate a wide range of curve characteristics, regardless of how localized they may be. Two of the most recognized and frequently used basis functions in practice are Fourier basis functions and Splines. Fourier basis functions are particularly suited for periodic data, while Spline bases are more flexible and, consequently, more complex than finite Fourier series. Splines are defined by their range of validity, the placement of knots, and their order. The total number of basis functions in a Spline basis system is determined by adding the number of interior knots to the spline order. We utilize a set of functional building blocks, denoted as $\phi_1, \phi_2, \ldots, \phi_k$ for $k = 1, \ldots, K$, which are combined linearly. Thus, a function $x(t)$ defined in this manner can be expressed mathematically as:

$$x(t) = \sum_{k=1}^{K} C_k \phi_k(t) = \mathbf{C}^T \boldsymbol{\phi}(t), \tag{2.1}$$

where $\mathbf{C} = (C_1, C_2, \ldots, C_K)$ and $\boldsymbol{\phi}(t) = (\phi_1(t), \phi_2(t), \ldots, \phi_K(t))$ are the coefficients and basis functions of the expansion respectively.

2.2 Functional Principal Component Analysis

PCA is a widely used method for exploring variability in multivariate data. PCA employs an eigenvalue decomposition of the variance matrix to identify directions in the observation space where the data exhibit the greatest variability. For each principal component, the analysis produces a loading vector or weight vector that indicates the direction of variability corresponding to that component. In a multivariate context, y_{ij} represents the value of variable j for the i^{th} individual, but in functional data, the discrete index j has been replaced by the continuous index t, so that the counterparts of variable values y_{ij} are function values $y_i(t)$. In the functional context, if $y_1(t), y_2(t), \ldots, y_n(t)$ be the observed curves, then each principal component is represented by a principal component weight function $\xi(t)$, defined over the same range of t as the functional data. The principal component scores for the individuals in the sample are the values z_i, calculated as follows Ramsay and Silverman (2002):

$$z_i = \int \xi(t) y_i(t) dt. \tag{2.2}$$

The purpose of simple PCA is to identify the weight function $\xi_1(t)$ that maximizes the variance of the principal component scores z_i subject to the constraint

$$\int \xi(t)^2 dt = 1. \tag{2.3}$$

Without such a constraint, we could increase the variance by simply multiplying by a large factor. The second, third, and higher-order principal components are defined similarly, but with additional constraints. The second component function $\xi_2(t)$ is specified to maximize the variance of the principal component scores subject to the constraint (2.2) as well as an additional constraint

$$\int \xi_2(t)\xi_1(t)dt = 0. \tag{2.4}$$

In general, for the j^{th} component, we impose the additional constraints

$$\int \xi_j(t)\xi_1(t)dt = \int \xi_j(t)\xi_2(t)dt = \ldots = \int \xi_j(t)\xi_{j-1}(t)dt = 0. \tag{2.5}$$

2.3 Plotting Components as Perturbations of the Mean

Interpreting the components is often not a straightforward matter in most functional PCA problems. A useful method is to examine plots of the overall mean function alongside the functions obtained by adding and subtracting a suitable multiple of the principal component function in question. In these plots, the solid curve represents the overall mean curve, while the dotted and dashed curves illustrate the effects of adding and subtracting a multiple of each principal component curve. In this context, it is important to select an appropriate multiple of the principal component function. Ramsay and Silverman (2005) define a constant C as the root-mean-square difference between $\hat{\mu}$ and its overall time average. Namely, they add and subtract

$$C^2 = T^{-1} \left\| \hat{\mu} - \bar{\mu} \right\|^2, \tag{2.6}$$

to each principal component curve, where

$$\bar{\mu} = T^{-1} \int \hat{\mu}(t)dt. \tag{2.7}$$

3 Data Analysis

In this section, we begin by describing the participants of the study using descriptive statistics. Next, we explore the data structure based on the features present in the data. Finally, we extract the pattern of variability in the data using FPCA to identify the primary factors contributing to ICP and determine their importance.

3.1 Summary Statistics for the ICP Data

In a study involving 12 patients diagnosed with spontaneous SAH, it was found that 6 patients (50 percent) were female, while the remaining were male. The mean age of the patients was 62.25. The dependence of ICP on time led to the utilization of FDA. When analyzing functional data, if the desired phenomenon is collected discretely at different time points, the first step is to convert them into curves. In this study, the raw data from the 12 patients with aSAH were transformed into continuous curves using a second-order B-spline with 13 basis functions. The curves were smoothed at this stage, before FPCA. The measurements and the corresponding fitted curves of ICP are depicted in Figure 1.

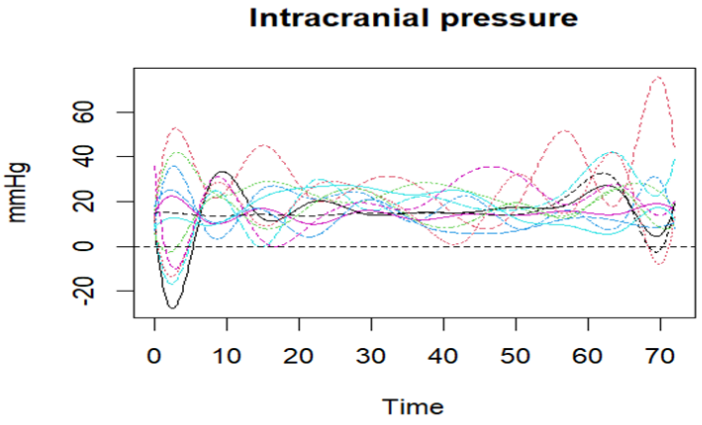

Figure 1: ICP curves of 12 patients with aSAH. Time is measured in hours.

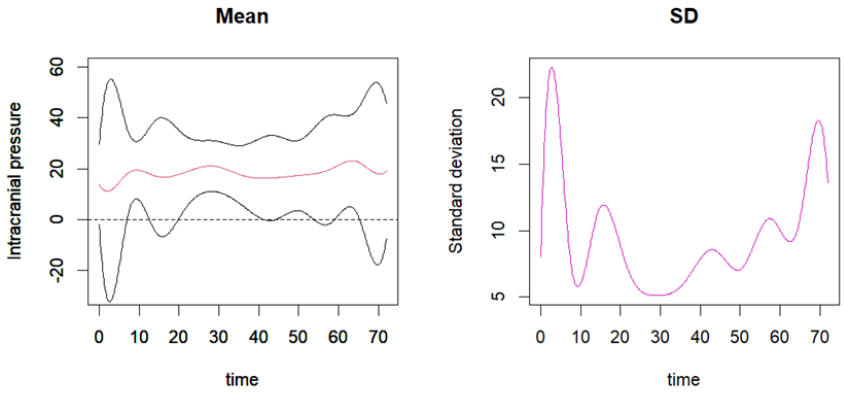

Figure 2: Mean and standard deviation functions for the ICP of 12 aSAH patients.

We show the mean and standard deviation functions for aSAH patients in Figure

2. As Figure 2 shows, the mean of ICP varies between about 13 and 22 mmHg. They are minimal at first hours of the follow-up period and are maximal at hours 60 and 66 of the study. Standard deviation values of ICP vary between 4.5 and 22 mmHg. The variability in the initial iterations of the measurement is higher than the other follow-up times, while in the fifth iteration the measurement has the least variability. It also seems that the standard deviation of ICP increases considerably from about 7 to 22 during initial hours.

3.2 FPCA Results for the ICP Data

In the given section, FPCA was employed to analyze the ICP data. The aim was to identify patterns of variation within the data and determine the main factors contributing to increased ICP. Figure 3 in the study showcases the four leading PC curves that were estimated from the ICP curves. PC curves are derived through FPCA and represent the main patterns of variation observed in the ICP data. These FPCs are plotted in Figure 3, allowing researchers and readers to visually understand the dominant modes of variation in the ICP curves. Each FPC represents a distinct pattern or trend within the data and can provide valuable insights into the factors influencing increased ICP.

These four FPCs account for 87.8 percent of the total variation of ICP curves. The first, second, third and fourth FPC explain around 52.3, 21.9, 8 and 5.6 percent of the total variation. The first and second PCs, which approximately explain the highest variability in the data, are respectively related to the patients' total GCS scores and their diastolic blood pressure. As evident from Figure 4, the patterns of variation in these two components are cyclic, and they exhibit the highest variability in the initial and final repetitions.

In contrast, the third and fourth PCs are respectively related to age and systolic blood pressure. These components explain relatively less variability compared to the first two components.

4 Discussion and Results

The total GCS score was identified as the most significant factor driving variability in ICP in patients with SAH due to an aneurysm. In this study, it explained approximately 50 percent of the variability. This primary component exhibited the highest oscillation in the initial repetitions, indicating a prominent role in the observed patterns of change.

Acknowledgments

The authors would like to express their deepest gratitude to the Ghaem and Razavi Hospitals, as well as the staff of the Neurosurgery Intensive Care Unit, who collaborated in data collection.

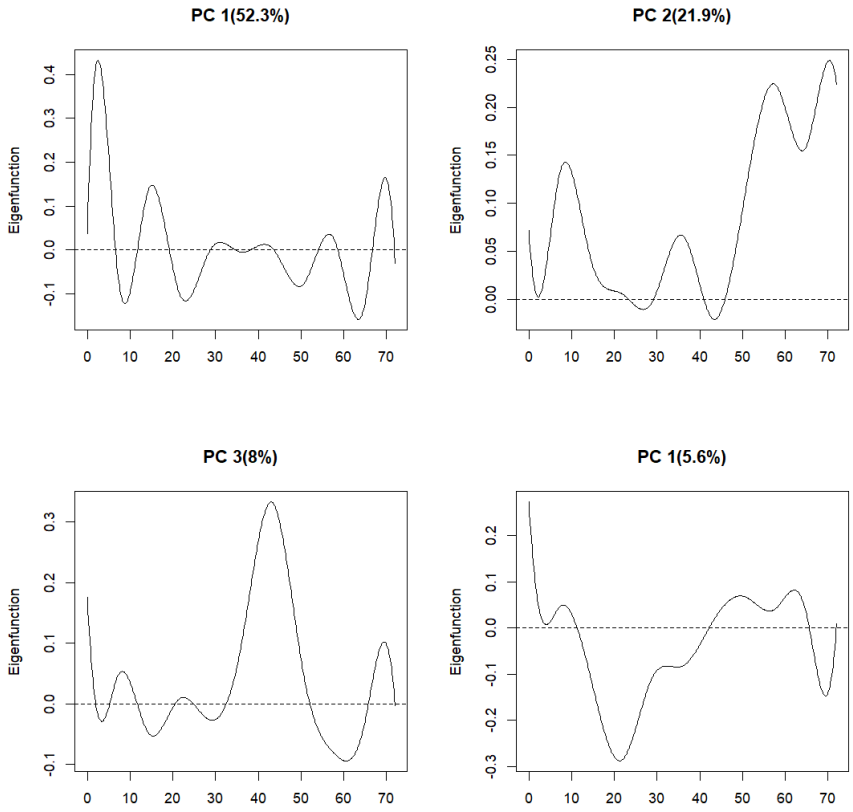

Figure 3: The first four leading PC curves estimated from the ICP data. The x-axis denotes the hours in the study period.

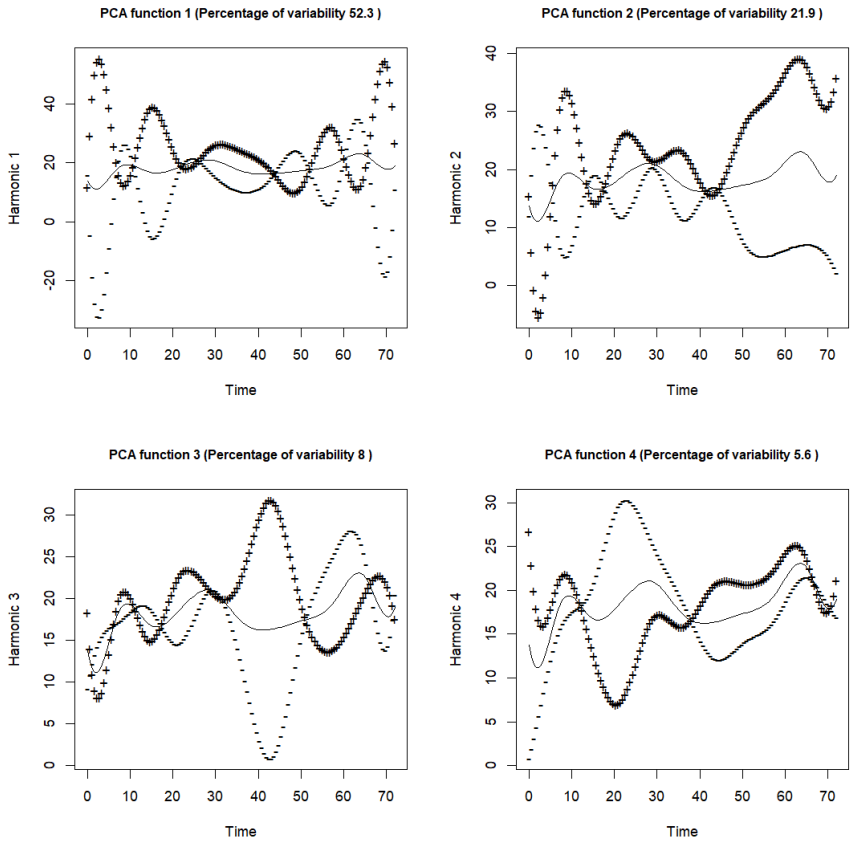

Figure 4: The mean ICP function and the effects of adding (+) and subtracting (-) a suitable multiple of each PC of curves. The x-axis denotes the hours in the study period.

References

Berrendero, J. R., Justel, A., and Svarc, M. (2011). Principal components for multivariate functional data. *Computational Statistics & Data Analysis*, **55**(9), 2619–2634.

Besse, P. and Ramsay, J. O. (1986). Principal components analysis of sampled functions. *Psychometrika*, **51**(2), 285–311.

Boente, G. and Fraiman, R. (2000). Kernel-based functional principal components. *Statistics & Probability Letters*, **48**(4), 335–345.

Czosnyka, M. and Pickard, J. D. (2004). Monitoring and interpretation of intracranial pressure. *Journal of Neurology, Neurosurgery & Psychiatry*, **75**(6), 813–821.

Gervini, D. (2008). Robust functional estimation using the median and spherical principal components. *Biometrika*, **95**(3), 587–600.

Hall, P., Müller, H.-G., and Wang, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *The Annals of Statistics*, **34**(3), 1493–1517.

Imberti, R., Picetti, E., Rossi, S., Capaccio, E., Accetta, G., Klersy, C., Lafe, E., Pietrobono, L., Cimino, F., Frattini, L., et al. (2021). Intracranial pressure monitoring in poor-grade patients with aneurysmal subarachnoid hemorrhage treated by coiling. *World Neurosurgery*, **156**, e206–e214.

Kim, K. (2010). The recent history functional linear model and its extension to sparse longitudinal data. *Communications in Statistics - Theory and Methods*, **39**(3), 368–388.

Ramsay, J. O. (1997). Functional data analysis. *Wiley Series in Probability and Statistics*.

Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Springer Verlag.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer Series in Statistics. Springer.

Ramsay, J., Hooker, G., and Graves, S. (2009). *Functional Data Analysis with R and MATLAB*. Use R! Springer.

Silverman, B. W. (1995). Incorporating parametric effects into functional principal components analysis. *Journal of the Royal Statistical Society: Series B (Methodological)*, **57**(4), 673–689.

Svedung Wettervik, T., Howells, T., Hånell, A., Ronne-Engström, E., Lewén, A., and Enblad, P. (2021). Low intracranial pressure variability is associated with delayed cerebral ischemia and unfavorable outcome in aneurysmal subarachnoid hemorrhage. *Journal of Clinical Monitoring and Computing*, 1–10.

Ullah, S. and Finch, C. F. (2013). Applications of functional data analysis: A systematic review. *BMC Medical Research Methodology*, **13**, 1–12.

Yao, F. and Lee, T. C. (2006). Penalized spline models for functional principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **68**(1), 3–25.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, **100**(470), 577–590.

Zoerle, T., Caccioppola, A., D'Angelo, E., Carbonara, M., Conte, G., Avignone, S., Zanier, E. R., Birg, T., Ortolano, F., Triulzi, F., et al. (2020). Optic nerve sheath diameter is not related to intracranial pressure in subarachnoid hemorrhage patients. *Neurocritical Care*, **33**, 491–498.