ZHANG, Baifa and LIN, Zhi-Cheng. Application of Functional Data Analysis in Complex Human Movement Analysis. Quality in Sport. 2025;40:60004. eISSN 2450-3118.
https://doi.org/10.12775/QS.2025.40.60004
https://apcz.umk.pl/QS/article/view/60004

The journal has been 20 points in the Ministry of Higher Education and Science of Poland parametric evaluation. Annex to the announcement of the Minister of Higher Education and Science of 05.01.2024. No. 32553.
Has a Journal's Unique Identifier: 201398. Scientific disciplines assigned: Economics and finance (Field of social sciences); Management and Quality Sciences (Field of social sciences).
Punkty Ministerialne z 2019 - aktualny rok 20 punktów. Załącznik do komunikatu Ministra Szkolnictwa Wyższego i Nauki z dnia 05.01.2024 r. Lp. 32553. Posiada Unikatowy Identyfikator Czasopisma: 201398.
Przypisane dyscypliny naukowe: Ekonomia i finanse (Dziedzina nauk społecznych); Nauki o zarządzaniu i jakości (Dziedzina nauk społecznych).

The authors declare that there is no conflict of interests regarding the publication of this paper.
Received: 02.03.2025. Revised: 21.03.2025. Accepted: 26.04.2025 Published: 30.04.2025.

Application of Functional Data Analysis in Complex Human Movement Analysis

Bai-Fa Zhang[1]*, Zhi-Cheng Lin[1]
https://orcid.org/0000-0002-6746-0066

[1] School of Sports, Southwest University, Chongqing 400715, China

Abstract

Functional data analysis, pioneered by Ramsay, former president of the Statistical Society of Canada, is a novel statistical method widely applied in the field of human movement analysis. This study clarifies relevant concepts and basic processes of functional data analysis, outlines the computational framework of functional principal component analysis, and focuses on exploring its applications in sports science, clinical rehabilitation, and motor development. Compared to traditional cross-sectional statistics, functional data analysis demonstrates unique advantages in human movement research and is expected to become a reliable analytical techique for exploring the laws of human mechanical movement in the future.

Keywords: Functional Data Analysis; Movement Analysis; Time series data;

Human movement analysis is a cross-disciplinary field aimed at quantifying human movement characteristics and mechanical performance through scientific technical means. It reveals the external mechanical and physical movement patterns of the human body, thereby promoting improvements in athletic performance（Dariush, 2003）. Human movement analysis has broad applications, involving various research areas such as sports science, clinical rehabilitation, and human-computer interaction. In recent years, with advances in data acquisition and storage methods, an increasing amount of functional data with high-dimensional features have been collected and archived, such as joint angles, ground reaction forces, joint moments, and surface electromyography. The underlying nature of such data represents complete continuous processes. However,traditional cross-sectional statistics not only overlook the sequential nature of samples but also rely on researchers' understanding of the movements themselves and subjective experience when selecting feature values, easily resulting in the loss of potentially important information（Richter et al., 2014）. For example, in gait analysis, most studies only extract parameter values at a few characteristic moments (such as peak knee flexion) for analysis, missing subtle transitions in overall movement patterns（Dannenmaier et al., 2020）. Consequently, there is an urgent need to introduce statistical techniques —Functional Data Analysis (FDA)—that better align with the attributes of functional data and help uncover deeper insights and dynamic change patterns.

FDA was recognized as a novel statistical method, is widely applied in the field of human movement analysis, such as motor development（Harrison et al., 2007）, sports injury diagnosis（Son et al., 2017a）, and rehabilitation effect evaluation（Baumgart et al., 2017）. The central concept of FDA is to view observed data functions as unified entities, rather than as discrete sequences of individual observations. Compared with traditional cross-sectional statistics, FDA both reflects the continuity of the system itself and satisfies the principle of finite energy. At the same time, as an important tool connecting continuous dynamic analysis and digital signal processing, it can be easily combined with regression analysis and pattern recognition technology. This paper reviews the application of FDA in the field of complex human movement analysis, clarifies relevant concepts and computational frameworks, in order to provide new statistical methods for future research in this area.

1 Introduction to Functional Data Analysis

1.1 Concept of Functional Data Analysis

In 1982, Professor Ramsay, an honorary professor at McGill University in Canada, first published the concept of functional data（Ramsay, 1982）. Functional data refers to data that varies with one or more continuous sets (time, space, etc.) and is presented in functional form. This type of data has functional characteristics, and its external manifestation can be different mathematical geometric objects such as curves, planes, or three-dimensional images. For instance, lower limb joint angles of subjects in gait analysis, ground reaction force curves, and electromyography signals of athletes in movement technique diagnosis—all these data with mathematical functional curve characteristics can be called functional data. According to the different number of individual observations and time interval characteristics in functional data, functional data can be divided into dense functional data and sparse functional data. If observations are made only at a few random irregular time points, such data is called sparse functional data. Longitudinal data in biology and medicine are typical sparse functional data. Under normal circumstances, functional data refers to dense functional data, where individual observations are regular and dense.

The book "Functional Data Analysis"（Ramsay et al., 2009）, co-published by Ramsay and Silverman in 1997, systematically introduces the basic concepts of FDA methods such as smoothing, interpolation, and registration, as well as examples of using common multivariate statistical methods for functional data, such as functional principal component analysis (FPCA), functional canonical correlation analysis (FCCA), and functional linear models. Unlike traditional econometric data types such as cross-sectional data, time series data, and panel data, FDA views the discrete, finite observation sequences in traditional data points as continuous, infinite functional curves, and treats them as separate data objects with the following advantages:

First, since functions are continuous and infinite, we can consider that the sample points of FDA come from an infinite-dimensional space. Therefore, FDA not only enables the analysis of existing data points but also allows for the prediction of unobserved points through the functional patterns identified. Second, through differential, derivative, and other processing of functional curves, more potential information than static data can be mined, such as the velocity and acceleration of the research object as the independent variable changes. Finally, due to the widespread application and improvement of the FDA method itself and statistical software MATLAB and R language, more reasonable and intuitive geometric interpretations can be given to the analysis results.

1.2 Steps of Functional Data Analysis

1.2.1 Functionalizing Discrete Data

The primary task of FDA is to transform discrete data into smooth functions. To convert discrete data into functional form, the measured time series data must first be standardized to a uniform length, forming an equal number of equidistantly distributed nodes (knots). Subsequently, K basis functions, $\emptyset_k$,$k = 1,2, \dots, K$, are linearly combined to form a functional curve that approximates the measured data curve. The larger the value of K, the number of basis functions, the closer the transformed functional curve is to the measured data; the smaller the value of K, the smoother the transformed functional curve and the larger the residual value between it and the measured data（Eilers et al., 1996）. The formula expression of the function formed by the linear combination of basis functions is as follows:

$$x(t) = \sum_{k=1}^{K} c_k \emptyset_k(t)$$

In the above formula, $\emptyset_k(t)$is the k-th basis function at time t, and $c_k$is the coefficient vector（Ramsay et al., 2009）. Currently, there are multiple types of basis functions that can functionalize discrete data, such as Fourier basis, B-spline basis, polynomial basis, exponential power basis, step function basis, empirical distribution function basis, and piecewise linear function basis. Among these many basis functions, the Fourier basis and B-spline basis are the most common. No specific basis function is universal for all types of data; the appropriate basis function needs to be chosen based on the characteristics of the data. The B-spline basis is suitable for non-periodic data, and due to its combination of compact support and flexibility, it, along with M-spline basis, I-spline basis, and truncated power basis, is an outstanding spline function. It has been applied by scholars in research on human movement analysis such as race walking（Dona et al., 2009）, vertical jump（Harrison et al., 2007 ; Ryan et al., 2006）, running（Coffey et al., 2011）, and rowing（Coffey et al., 2011）.

Spline basis functions are piecewise polynomials containing multiple independent internal observation values, with the boundary points at both ends of each subsegment called break points. Data nodes are usually placed equidistantly at break points. The order of a spline basis function equals the degree of its polynomial plus one; for example, if the order of a spline basis is default 4, then the degree of its polynomial is 3. In human movement analysis research, scholars are more concerned with the first-order derivative (velocity) and second-order derivative (acceleration) of the measured data. The order of the basis function needs to be at least 2 orders higher than the order of the derivative. For example, to find the second-order derivative (acceleration) of the original data, the order of the basis function should be at least 4（Ramsay et al., 2009）. The number of basis functions K is usually the sum of the internal nodes of the measured data and the order of the basis function. Internal nodes refer to the nodes excluding the two ends; for example, with 100 nodes, there are 98 internal nodes.

1.2.2 Smoothness

In real-life measurements, data usually contains signal noise and measurement errors, which are assumed to follow a normal distribution. This can be expressed by the following formula:

$$y_j = x(t_j) + \varepsilon_j$$

In this formula, $y_j$represents the measured data, $x(t_j)$represents the true signal data, and ε_j represents noise or measurement error. After defining the basis functions$\emptyset_k$, to eliminate noise and measurement errors in the signal, the coefficient vector of the basis functions c_k needs to be adjusted. Currently, there are two common methods to calculate the coefficient vector of basis functions: the least squares method and the penalty function. The least squares estimation or minimum residual sum of squares based on B-spline basis expansion, also known as smoothing spline regression, is suitable for solving simple linear problems. The formula is as follows:

$$\mathrm{SSE}(x) = \sum_{j=1}^{n} \left[y_j - x(t_j)\right]^2$$

Converting the function in the above formula through basis function expansion:

$$\mathrm{SSE}(x) = \sum_{j=1}^{n} \left[y_j - \sum_{k=1}^{K} c_k \emptyset_k(t)\right]^2$$

The standard deviation of the total residual sum of squares, also known as root mean square error, is used to evaluate the difference between estimated $\mathrm{y_j}$and observed $\mathrm{x(t_j)}$ values. It is expressed by the following formula:

$$\mathrm{RMSE}(x) = \sqrt{\frac{\sum_{j=1}^{n}\left[y_j - x(t_j)\right]^2}{n}}$$

The roughness penalty function is a regularization approach that effectively approximates discrete data through functions and is usually considered a more comprehensive smoothing technique. Compared to the smoothing spline regression method, the roughness penalty method selects a larger K value to estimate the measured data. At the same time, to prevent overfitting, a penalty term is set to adjust the curvature of the estimated function. The formula is as follows:

$$\begin{aligned}\mathrm{PENSSE} &= \sum_{j=1}^{n} \left[y_j - x(t_j)\right]^2 + \lambda \times PEN_m \\ &= \sum_{j=1}^{n} \left[y_j - x(t_j)\right]^2 + \lambda \times \int [D^m x(t)]^2 dt\end{aligned}$$

In the above formula, $\lambda \times \mathrm{PEN_m}$ is the penalty term, and λ is the smoothing parameter of the penalty term used to balance the degree of fitting and noise. Common methods for determining the value of λ include cross-validation (CV) and generalized cross-validation (GCV)（Craven et al., 1979）. The formula for determining the value of λ using cross-validation is as follows:

$$\mathrm{CV}(x) = \sum_{i=1}^{N} \left[y_j - \alpha_\lambda^{(-i)} - \int x_i(t) \beta_\lambda^{(-i)}\, dt\right]^2$$

In the above formula, $\alpha_\lambda^{(-i)}$ and $\beta_\lambda^{(-i)}$ represent the estimated values of α and β after removing the i observation. Generalized cross-validation is developed from cross-validation and is more computationally simple and reliable compared to the latter. The formula is as follows:

$$\text{GCV}(\lambda) = \left(\frac{n}{n = df(\lambda)}\right)\left(\frac{SSE}{n - df(\lambda)}\right)$$

In the above formula, $df(\lambda)$ is the degrees of freedom of the estimated function, and SSE is the total residual sum of squares. The parameter λ is usually set in the interval $[10^{-12}, -10^{-2}]$（Warmenhoven et al., 2021）. However, some scholars（Zin et al., 2020）have found that few human movement analysis studies using FDA methods make an objective choice of the smoothing parameter. Researchers usually subjectively determine an appropriate value but do not describe the specific method of determination. Therefore, Zin（Zin et al., 2020） compared the smoothing effects of CV, GCV, and subjective selection methods by collecting a dataset of lower limb hip joint sagittal plane angles from 20 male subjects performing American kettlebell swing movements. The study found that CV and GCV did not calculate the most appropriate value of λ. Zin believed that the selection of smoothing parameters should combine subjective selection and objective calculation methods (CV and GCV).

1.2.3 Registration

Before formally analyzing functional data, there is still an urgent problem to be solved, namely, the need to eliminate phase variation in the measured data. Usually, the measured data curves not only include amplitude variation but also phase variation, as shown in Figure 1. Phase variation comes from the asynchrony between the internal clock of the system and the sampling clock, that is, the physical time scale is not directly related to the internal dynamics of the system. If the phase variation between data is not eliminated, abruptly performing cross-sectional statistics on the data will mislead researchers to make incorrect judgments. Therefore, scholars have introduced the Curve Registration algorithm.

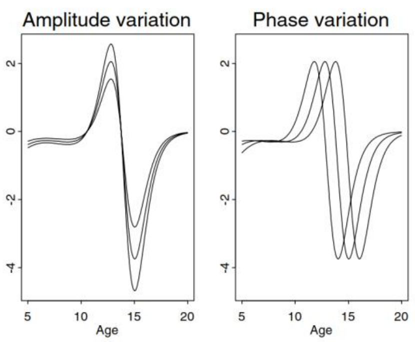

Figure 1. Schematic diagram of amplitude variation and phase variation（Tuddenham et al., 1954）

Curve registration refers to an algorithm that uses a warping function to capture the phase variation in the original data curve, and subsequently applies monotone transformations to the sample function domain, thereby aligning their functional features (such as peaks and valleys) （Kneip et al., 2008）. After registration processing, the sample curves only show amplitude variation, highlighting the characteristics of amplitude changes in the curves. Common registration methods include shift registration, landmark registration, and continuous registration. Shift registration is a relatively simple method that achieves registration by determining shift parameters. The formula is as follows:

$$x_i^*(t) = x_i(t + \delta_i)$$

Taking Ramsay's hand movement control experiment as an example（Ramsay et al., 1995）, scholars required subjects to pinch a sensor simultaneously with their index finger and thumb and reach a target force value (10N). The data results are shown in Figure 2. Since the time for each subject to pinch the sensor is not fixed, shift registration is needed to align the curve landmarks. The method for determining the shift parameter is to find the minimum sum of squares error (SSE) between the registration function and the target function amplitude difference. The formula is as follows:

$$\text{REGEER} = \sum_{i=1}^{N} \int_{\mathcal{T}} [x_i(t+\delta_i) - \hat{\mu}(t)]^2 \, ds$$
$$= \sum_{i=1}^{N} \int_{\mathcal{T}} [x_i^*(t) - \hat{\mu}(t)]^2 \, ds$$

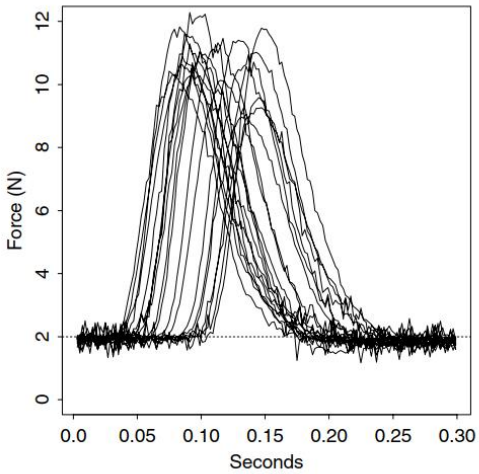

Figure 2. Force value curves collected from hand movement experiments（Ramsay et al., 1995）

Landmarks refer to certain feature points of the curve, such as maximum values, minimum values, and zero-crossing points. Usually, the same landmark points are located at the same moment of the system's internal clock, and landmarks can be better identified with the help of the derivative form of the curve sample. Landmark registration requires capturing the independent variable $\mathrm{t_{if}}, \mathrm{f} = 1, \ldots, \mathrm{F}$corresponding to the landmark points of each curve, thereby constructing a warping function $\mathrm{h_i}$ for each curve. The specific formula is as follows:

$$x^*(t) = x_i[h_i(t)]$$

Continuous registration is improved from the time-warping algorithm. The specific calculation formula is as follows:

$$\mathrm{h}(t) = C_0 + C_1 \int_0^t expW(u)du$$

In the above formula, $\mathrm{C_0}$ and $\mathrm{C_1}$are constants, corresponding to the amplitude scaling factor and phase shift scaling factor, respectively. $\mathrm{h(t)}$is consistent with the upper and lower limits on the interval [0, T]. When $\mathrm{W(u)} = 0$, $\mathrm{h(t)} = \mathrm{t}$, the system clock is synchronized with the sampling clock; when $\mathrm{W(u)} > 0$, $\mathrm{h(t)} > \mathrm{t}$, the system clock is slower than the sampling clock, and the warping time plays an accelerating role; when $\mathrm{W(u)} < 0$, $\mathrm{h(t)} < \mathrm{t}$, the system clock is faster than the sampling clock, and the warping time plays a decelerating role. The smoothness of $\mathrm{h(t)}$is controlled by the roughness penalty mentioned above. The formula is as follows:

$$\mathrm{MINEIG}_{\lambda}(h) = MINEIG(h) + \lambda \int \left\{W^{(m)}(t)\right\}^2 dt$$

1.2.4 Functional Principal Component Analysis

Functional Principal Component Analysis (fPCA), as the most common analysis technique for handling functional multivariate problems in human movement analysis research, plays a pioneering role in the development of FDA, dating back to 1982（Dauxois et al., 1982）. This algorithm explores the typical variation patterns of data curves by reducing the dimensions of the original dataset containing a large amount of correlated variable information（Ullah et al., 2013）. According to the number of analysis indicators, functional principal component analysis can be divided into univariate functional principal component analysis or simple functional principal component analysis, and multivariate functional principal component analysis. Essentially, functional principal component analysis converts the coordinate system of the original data, establishing a new coordinate system based on the direction of maximum variance. The new coordinate axes are uncorrelated and ordered, with a few leading coordinate axes or principal components reflecting most of the information in the original data. The formula is as follows:

$$\mathrm{f_i} = \int \xi \mathrm{x_i} = \int \xi(\mathrm{s})\mathrm{x_i}(\mathrm{s})\mathrm{ds}$$

In this formula, $\mathrm{f_i}$ is the principal component score, $\xi(\mathrm{s})$ is the weight function or eigenfunction, and $\mathrm{x_i}(\mathrm{s})$ is the observed value. Substituting the observed value into the above formula, we can solve for the principal component score of the eigenfunction on the observed value. The process of extracting functional principal components is essentially finding the eigenfunction that can maximize the variance of the functional principal component scores ξ_p while satisfying unit norm regularization. The formula is as follows:

$$\begin{cases} \max \sum_{\mathrm{i}} \left(\mathrm{f_{ip}}\right)^2 \\ \mathrm{s.t.} \int \xi_\mathrm{p}(\mathrm{s})^2 \mathrm{ds} = 1, \int \xi_\mathrm{p}(\mathrm{s})\xi_\mathrm{m}(\mathrm{s})\mathrm{ds} = 0, \mathrm{m} = 1,2,\dots,\mathrm{p}-1 \end{cases}$$

Similar to multivariate statistics, functional principal component analysis also extracts eigenvalues ρ through the covariance matrix $\upsilon(\mathrm{s},\mathrm{t})$. The formula is as follows:

$$\upsilon(\mathrm{s},\mathrm{t}) = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{x_i}(\mathrm{s})\mathrm{x_i}(\mathrm{t})$$

$$\int \upsilon(\mathrm{s},\mathrm{t})\xi(\mathrm{t})\mathrm{dt} = \rho\,\xi(\mathrm{s})$$

Applying integral transform to the left side of the above formula, we obtain the covariance operator V, expressed as follows:

$$\mathrm{V}\xi = \int \nu(.,\mathrm{t})\xi(\mathrm{t})\mathrm{dt} = \rho\xi$$

Functional principal component analysis has achieved good results in many research areas of human movement analysis. Ullah（Ullah and Finch, 2013）conducted a systematic review of applied research on FDA published between 1995 and 2010. The authors found that 51 (69.7%) of the 84 selected studies used functional principal component analysis, such as gait analysis research for Parkinson's patients. Scholars usually adopt general statistical hypothesis testing methods（Warmenhoven et al., 2018a）(such as analysis of variance) or multivariate statistical analysis methods（Warmenhoven et al., 2018c ; Warmenhoven et al., 2017 ; Liebl et al., 2014）(such as discriminant analysis, multivariate linear regression models, clustering) to analyze functional principal component scores, and then associate each principal component with real physical meanings. In traditional multivariate statistics, principal component analysis usually selects sample curves corresponding to the 5th percentile and 95th percentile of each principal component score, that is, two representative extreme curves, and explains the physical meaning of each component by comparing their differences（Landry et al., 2007）. For example, when at a certain moment, the sample curve corresponding to the 5th percentile has a larger amplitude, while the sample curve corresponding to the 95th percentile has a smaller amplitude, it can be considered that this principal component captures the amplitude change at this moment（Wrigley et al., 2005）. However, this method can easily lead to interference between the characteristics of each principal component. Therefore, Ramsay added and subtracted fixed multiples of eigenfunctions based on the mean function. The formula is as follows:

$$s_{+} = \bar{u} + \sqrt{\rho} * \xi$$
$$s_{-} = \bar{u} - \sqrt{\rho} * \xi$$

In the above formula, $\bar{u}$ is the mean function, ξ is the eigenfunction, and $\sqrt{\rho}$ is the standard deviation of the principal component eigenvalue. Through the region enclosed by s_{+} and s_{-}, we can clearly see where the variability occurs, and thus reasonably explain the physical meaning represented by this principal component.

2 Application of Functional Data Analysis in Complex Human Movement Analysis

2.1 Applications in Sports Science

With the continuous development of FDA methods, this technology has gained increasingly widespread application in the field of sports science, such as movement analysis research in weightlifting（Dalla Bernardina et al., 2021 ; Kipp et al., 2012）, rowing （Warmenhoven et al., 2019b ; Warmenhoven et al., 2018b ; Warmenhoven et al., 2018a ; Warmenhoven et al., 2018c ; Warmenhoven et al., 2017）, race walking（Dona et al., 2009）, running（Willwacher et al., 2016 ; Liebl et al., 2014）, cycling（Soares et al., 2021）, swimming（Leroy et al., 2018）, and other sports.

For weightlifting events, Kipp （Kipp et al., 2012）used functional principal components analysis (fPCA) to analyze the movement patterns of the lower limbs of 10 college weightlifters performing snatch lifts at 85% of 1RM. The study found that the PC1 scores of hip joint flexion-extension angle curves and the PC2 scores of knee joint flexion-extension moment were positively correlated with normalized weightlifting performance. At the same time, they corresponded to the variability in the first pull phase and the second pull phase, respectively. The study suggested that maintaining trunk stability during the first pull phase and rapid knee extension during the second pull phase help improve weightlifting performance. Bernardina（Dalla Bernardina et al., 2021）used functional analysis of variance (fANOVA) to analyze the displacement velocity changes of the barbell when 10 Paralympic weightlifters completed bench press movements at 50% and 90% of 1RM. The study found that at 90% maximum load, barbell displacement velocity was asymmetric during 10-20% and 90-100% of the upward pushing phase. However, traditional ANOVA did not find statistical differences in the average velocity at the two ends. The study suggested that the asymmetry was due to slower initial velocities amplifying the imbalance in muscle strength between the two sides, while at the end of the push, the non-dominant limb increased pushing speed to maintain the barbell level. The results indicated that fANOVA more accurately reflects the movement patterns in Paralympic weightlifting events.

For rowing events, Professor Warmenhoven from the Department of Exercise and Sport Science at the University of Sydney demonstrated and explained the application of FDA methods such as fPCA（Warmenhoven et al., 2019b ; Warmenhoven et al., 2018a）, bivariate functional principal components analysis (BfPCA)（Warmenhoven et al., 2019a ; Warmenhoven et al., 2018c ; Warmenhoven et al., 2017）, and functional t-test (ft-test)（Warmenhoven et al., 2018b） in rowing. In 2017, Warmenhoven（Warmenhoven et al., 2017）applied BfPCA to analyze the propulsive force and entry angle changes of the oar during the water entry phase for 27 Australian female rowers (national level n=14, international level n=13) rowing at a stroke rate of 32 strokes/minute on water. The study found that the greater the loading rate of oar propulsive force in the first half of the water entry phase, the faster the boat speed. The study concluded that the BfPCA method is beneficial for extracting features from the oar propulsive force-entry angle change diagram, thereby analyzing the force application patterns of high-level rowers. Subsequently in 2018, Warmenhoven（Warmenhoven et al., 2018c） also used the BfPCA method to explore whether gender would affect the oar propulsive force-entry angle change diagram, i.e., the rowing technique patterns of athletes. The study found that gender had a statistically significant effect on rowers' oar propulsive force-entry angle changes, suggesting that coaches need to consider gender differences when analyzing rowers' movement patterns. In 2019, Warmenhoven（Warmenhoven et al., 2019a） used the complete rowing cycle oar propulsive force-entry angle change data from two excellent Australian female rowers as an example, summarized data processing experiences from previous research, and systematically discussed the practical workflow of BfPCA, precautions, and how to connect data results with real physical meanings. The study suggested that when two variables have inconsistent units, consideration should be given to converting variables into dimensionless quantities to improve the interpretability of results. For the fPCA method, Warmenhoven（Warmenhoven et al., 2019b）used the oar propulsive force change data during the complete rowing cycle of two excellent Australian female rowers as an example to explain the practical workflow of fPCA, the interpretation of results, and reviewed previous research. The study suggested that appropriate use of registration, maximum variance rotation, and other methods can effectively improve the interpretability of results.

For race walking events, Dona （Dona et al., 2009）used the fPCA method to analyze the sagittal plane joint angles and net joint moments of the knee in 7 Italian national race walkers at self-selected speeds. The study found that PC2 of the knee flexion-extension angle curve reflects the technical characteristics of race walkers during the foot landing buffer phase. High-level race walkers maintain knee joint stability, while low-level race walkers increase knee extension amplitude. At the same time, low-level athletes have larger knee flexion moments during this phase. The study suggested that maintaining this movement pattern for an extended period would increase the risk of knee joint injury in low-level race walkers. The research recommended that fPCA helps extract movement characteristics of athletes at different skill levels that traditional statistical methods cannot identify, but attention should be paid to eliminating differences unrelated to sports performance.

For running events, Liebl （Liebl et al., 2014）combined fPCA with clustering algorithms to analyze the ankle joint flexion-extension moments of the dominant lower limb during the support phase of 119 amateur runners at a speed of 3.5m/s, thereby classifying runners into forefoot and rearfoot runners. The study found clear grouping of runners, with forefoot runners having statistically significantly higher maximum ankle flexion strength than rearfoot runners. Willwacher （Willwacher et al., 2016）also used a combination of fPCA and clustering algorithms to analyze the relationship between free moments, lower limb biomechanical characteristics, and injury risk during the support phase of 222 amateur runners at a speed of 3.5m/s, using free moment as an indicator for runners' choice of shoe type. The study found that runners' Free Moment (FM) had two PCs: PC1 functions to counteract the interference of the body's angular momentum during the flight phase, while PC2 functions to compensate for the moment difference between the upper and lower limbs during the support phase to ensure body stability. The study suggested that runners' FM PC scores could serve as an indicator for classifying runners according to injury risk.

For cycling events, Soares（Soares et al., 2021）used the fANOVA method to analyze the pedaling torque of 20 amateur cyclists riding power bicycles at 60%, 80%, and 95% maximum power loads. The study found that traditional ANOVA did not identify statistical differences in pedaling torque peaks between the two sides of the lower limbs, while fANOVA showed statistical differences in the non-peak portions of the pedaling torque on both sides. The study concluded that pedaling torque peaks cannot be used alone as an indicator to judge whether the force on both lower limbs is symmetrical. When analyzing whether cyclists' lower limb force is symmetrical, the overall trend of change needs to be considered. Therefore, fANOVA can be used as a tool to analyze differences in overall change trends applied in the biomechanical analysis of cycling.

For swimming events, Leroy（Leroy et al., 2018）used functional clustering analysis to analyze tracking data of 100m freestyle performance during the 12-20 year training period for 1468 French youth swimmers published by the French Swimming Federation. The study identified five different patterns of athletic ability growth, showing that athletes with similar levels in the early stages may have significant differences in athletic level in adulthood. At the same time, the critical period for athletes' performance improvement is before the age of 16. The study suggested that functional clustering analysis can effectively predict the growth potential of swimmers and can be a powerful tool for coaches in talent selection.

2.2 Applications in Clinical Rehabilitation

In the field of clinical rehabilitation, most applications of FDA research focus on the biomechanical mechanisms of lower limb joint injuries, such as anterior cruciate ligament (ACL) injuries（Stephens et al., 2020；Baumgart et al., 2017；Hebert-Losier et al., 2015）, knee pain（Son et al., 2017a）, ankle instability or sprains（Son et al., 2017b）, and chronic Achilles tendon injury (AT)（Donoghue et al., 2008）. After an injury occurs, patients develop compensatory phenomena to complete functional movement tasks, manifested as abnormal movement patterns. Observing and analyzing these patterns helps rehabilitation therapists make clinical decisions and avoid the risk of re-injury. FDA treats the movement parameters (joint angle curves, ground reaction force curves, etc.) of the patient's complete movement process as geometric objects, which has inherent advantages in revealing the differences in movement processes of patients with different severity compared to traditional methods.

For example, Baumgart（Baumgart et al., 2017）divided 40 post-ACL surgery patients into high-score and low-score groups according to the International Knee Documentation Committee (IKDC) scoring standards. The study combined ft-test and traditional discrete indicators to analyze the GRF data of the two groups of subjects performing vertical jump tasks. The ft-test results showed that compared to the high-score group, the GRF of the affected lower limb in the low-score group was smaller during the vertical jump squat phase; however, no difference in GRF peaks was found between the two groups. The study suggested that FDA indicators are more sensitive than traditional discrete indicators and help clinicians discover differences in movement patterns between patients at different rehabilitation levels. Similarly, Stephens（Stephens et al., 2020）combined fPCA and traditional discrete indicators to analyze the ground reaction force (GRF) during single-leg landing tasks performed by the healthy leg and the affected leg of an ACL-injured athlete. The study found that PC4 of the GRF curve could effectively identify changes in the movement pattern of the affected lower limb at the end of the rapid buffering phase, while there were no differences in traditional discrete indicators (GRF peak, landing stability duration) between the two lower limbs. The study suggested that patients adjust their movement strategies according to their disease constraints to ensure the stability of their movements, and traditional discrete indicators have difficulty reflecting such process-related changes. Additionally, Donoghue（Donoghue et al., 2008）used the fPCA method to study the kinematics of AT patients during running. The study found that compared to the normal control group, AT patients had smaller PC1 amplitude variation in the frontal plane lower limb joint angles during the first half of the running support phase. The study considered that AT patients have highly similar movement patterns during the foot loading phase, consistent with the dynamic systems theory view that after injury, patients' movements become stereotyped and energy-consuming, i.e., movement variability decreases. This suggests that the characteristics of patients' movement patterns after injury are not related to the actual numerical values of movement parameters, but to the overall change trend.

2.3 Applications in Movement Development

Early movement development scholars mostly adopted qualitative descriptive methods to summarize movement pattern information of basic movement skills at different developmental stages, such as whole-sequence and part-sequence methods, but these are easily influenced by the observer's subjectivity. In the later period, to objectively and quantitatively evaluate the staged changes in children's movement patterns, scholars introduced FDA methods. Initially, Ryan（Ryan et al., 2006）combined fPCA with stepwise discriminant analysis to study the developmental characteristics of children's vertical jump movements. The study found that PC3 of knee flexion angle could be used as a sensitive indicator to distinguish different developmental stages, reflecting children's ability in muscle stretch-shortening cycles of the lower limbs. At the same time, registration could effectively improve the discrimination of this PC. Soon after, Harrison（Harrison et al., 2007）used the BfPCA method to further study the relationship between children's vertical jump movement maturity and lower limb joint coordination. The study found that PC3 of the knee-hip angle-angle diagram could effectively distinguish vertical jump movements at different developmental stages, suggesting that higher-order vertical jump movements are characterized by greater knee flexion angles and smaller hip flexion angles.

3 Summary

During human movement, the motion parameters of limb segments have continuity and periodicity, which are typical functional data. Using FDA techniques can more precisely identify subtle transitions in movement patterns and supplement important information overlooked by traditional cross-sectional statistics. This study clarified the relevant concepts and analytical steps of FDA, outlined the computational framework of fPCA, and provided a detailed review of the applications of FDA in complex human movement analysis research, offering a reference for the widespread use of FDA. As seen in this paper, FDA demonstrates superior advantages in human movement analysis research, but many challenges still exist. These include the setting of smoothing parameters in pre-processing, the necessity of registration and maximum orthogonal rotation in fPCA, and the robustness of results. To enable this method to become a reliable technical means for human movement analysis research as soon as possible, it is recommended that subsequent research focus on the impact of additions, deletions, and parameter settings in each analytical step on computational results, in order to form specific and feasible operational process standards.

Author’s contributions

B-FZ designed the study and wrote the manuscript; Z-CL searched for literature.

Funding

This work was supported by special fund project for basic research business expenses of the central universities of Southwest University in China (grant number SWU2309726).

Acknowledgments

We would like to thank staff and faculty at the school of sports at Southwest University.

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

BAUMGART C, HOPPE M W, FREIWALD J, 2017. Phase-Specific Ground Reaction Force Analyses of Bilateral and Unilateral Jumps in Patients With ACL Reconstruction [J]. Orthopaedic Journal of Sports Medicine, 5(6): 1-9.

COFFEY N, HARRISON A J, DONOGHUE O A, et al., 2011. Common functional principal components analysis: A new approach to analyzing human movement data [J]. Human Movement Science, 30(6): 1144-1166.

CRAVEN P, WAHBA G, 1979. Smoothing noisy data with spline functions – estimating the correct degree of smoothing by the method of generalized cross-validation [J]. Numerische Mathematik, 31(4): 377-403.

DALLA BERNARDINA G R, DOS SANTOS M D M, RESENDE R A, et al., 2021. Asymmetric velocity profiles in Paralympic powerlifters performing at different exercise intensities are detected by functional data analysis [J]. Journal of Biomechanics, 123(23): 1-5.

DANNENMAIER J, KALTENBACH C, KOLLE T, et al., 2020. Application of functional data analysis to explore movements: walking, running and jumping - A systematic review [J]. Gait & posture, 77(3): 182-189.

DARIUSH B, 2003. Human motion analysis for biomechanics and biomedicine [J]. Machine Vision and Applications, 14(4): 202-205.

DAUXOIS J, POUSSE A, ROMAIN Y, 1982. Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference [J]. Journal of Multivariate Analysis, 12(1): 136-154.

DONA G, PREATONI E, COBELLI C, et al., 2009. Application of functional principal component analysis in race walking: An emerging methodology [J]. Sports Biomechanics, 8(4): 284-301.

DONOGHUE O A, HARRISON A J, COFFEY N, et al., 2008. Functional data analysis of running kinematics in chronic Achilles tendon injury [J]. Medicine and science in sports and exercise, 40(7): 1323-1335.

EILERS P H C, MARX B D, 1996. Flexible smoothing with B-splines and penalties [J]. Statistical Science, 11(2): 89-102.

HARRISON A J, RYAN W, HAYES K, 2007. Functional data analysis of joint coordination in the development of vertical jump performance [J]. Sports Biomechanics, 6(2): 199-214.

HEBERT-LOSIER K, PINI A, VANTINI S, et al., 2015. One-leg hop kinematics 20 years following anterior cruciate ligament rupture: Data revisited using functional data analysis [J]. Clinical Biomechanics, 30(10): 1153-1161.

KIPP K, REDDEN J, SABICK M B, et al., 2012. Weightlifting performance is related to kinematic and kinetic patterns of the hip and knee joints [J]. Journal Of Strength And Conditioning Research, 26(7): 1838-1844.

KNEIP A, RAMSAY J O, 2008. Combining Registration and Fitting for Functional Models [J]. Journal of the American Statistical Association, 103(483): 1155-1165.

LANDRY S C, MCKEAN K A, HUBLEY-KOZEY C L, et al., 2007. Neuromuscular and lower limb biomechanical differences exist between male and female elite adolescent soccer players during an unanticipated run and crosscut maneuver [J]. American Journal of Sports Medicine, 35(11): 1901-1911.

LEROY A, MARC A, DUPAS O, et al., 2018. Functional Data Analysis in Sport Science: Example of Swimmers' Progression Curves Clustering [J]. Applied Sciences-Basel, 8(10): 18.

LIEBL D, WILLWACHER S, HAMILL J, et al., 2014. Ankle plantarflexion strength in rearfoot and forefoot runners: A novel clusteranalytic approach [J]. Human Movement Science, 35(06): 104-120.

RAMSAY J, HOOKER G, GRAVES S, 2009. Functional Data Analysis with R and MATLAB [M]. Springer New York.

RAMSAY J O, 1982. WHEN THE DATA ARE FUNCTIONS [J]. Psychometrika, 47(4): 379-396.

RAMSAY J O, WANG X, FLANAGAN R, 1995. A functional data analysis of the pinch force of human fingers [J]. Journal of the Royal Statistical Society: Series C (Applied Statistics), 44(1): 17-30.

RICHTER C, O′CONNOR N E, MARSHALL B, et al., 2014. Comparison of discrete-point vs. dimensionality-reduction techniques for describing performance-related aspects of maximal vertical jumping [J]. Journal of Biomechanics, 47(12): 3012-3017.

RYAN W, HARRISON A, HAYES K, 2006. Functional data analysis of knee joint kinematics in the vertical jump [J]. Sports Biomechanics, 5(1): 121-138.

SOARES J D, CARPES F P, GERALDO G D, et al., 2021. Functional data analysis reveals asymmetrical crank torque during cycling performed at different exercise intensities [J]. Journal of Biomechanics, 122(06): 1-5.
SON S J, KIM H, SEELEY M K, et al., 2017a. Efficacy of Sensory Transcutaneous Electrical Nerve Stimulation on Perceived Pain and Gait Patterns in Individuals With Experimental Knee Pain [J]. Archives Of Physical Medicine And Rehabilitation, 98(1): 25-35.
SON S J, KIM H, SEELEY M K, et al., 2017b. Movement Strategies among Groups of Chronic Ankle Instability, Coper, and Control [J]. Medicine and science in sports and exercise, 49(8): 1649-1661.
STEPHENS J M, CHAPMAN D W, TATE K, et al., 2020. A drop landing screening approach to monitor an individual using functional data analysis: An ACL injury case study [J]. Journal Of Science And Medicine In Sport, 23(3): 241-245.
TUDDENHAM R D, SNYDER M M, 1954. Physical growth of California boys and girls from birth to eighteen years [J]. Publications in child development. University of California, Berkeley, 1(2): 183-364.
ULLAH S, FINCH C F, 2013. Applications of functional data analysis: A systematic review [J]. BMC Medical Research Methodology, 43(13): 1-12.
WARMENHOVEN J, COBLEY S, DRAPER C, et al., 2017. Assessment of propulsive pin force and oar angle time-series using functional data analysis in on-water rowing [J]. Scandinavian Journal Of Medicine & Science In Sports, 27(12): 1688-1696.
WARMENHOVEN J, SMITH R, DRAPER C, et al., 2018a. Force coordination strategies in on-water single sculling: Are asymmetries related to better rowing performance? [J]. Scandinavian Journal Of Medicine & Science In Sports, 28(4): 1379-1388.
WARMENHOVEN J, HARRISON A, ROBINSON M A, et al., 2018b. A force profile analysis comparison between functional data analysis, statistical parametric mapping and statistical non-parametric mapping in on-water single sculling [J]. Journal Of Science And Medicine In Sport, 21(10): 1100-1105.
WARMENHOVEN J, COBLEY S, DRAPER C, et al., 2018c. How gender and boat-side affect shape characteristics of force-angle profiles in single sculling: Insights from functional data analysis [J]. Journal Of Science And Medicine In Sport, 21(5): 533-537.
WARMENHOVEN J, COBLEY S, DRAPER C, et al., 2019a. Bivariate functional principal components analysis: considerations for use with multivariate movement signatures in sports biomechanics [J]. Sports Biomechanics, 18(1): 10-27.
WARMENHOVEN J, COBLEY S, DRAPER C, et al., 2019b. Considerations for the use of functional principal components analysis in sports biomechanics: examples from on-water rowing [J]. Sports Biomechanics, 18(3): 317-341.
WARMENHOVEN J, BARGARY N, LIEBL D, et al., 2021. PCA of waveforms and functional PCA: A primer for biomechanics [J]. Journal of Biomechanics, 116(6.
WILLWACHER S, GOETZE I, FISCHER K M, et al., 2016. The free moment in running and its relation to joint loading and injury risk [J]. Footwear Science, 8(1): 1-11.
WRIGLEY A T, ALBERT W J, DELUZIO K J, et al., 2005. Differentiating lifting technique between those who develop low back pain and those who do not [J]. Clinical Biomechanics, 20(3): 254-263.
ZIN M A M, RAMBELY A S, ARIFF N M, et al., 2020. Smoothing and Differentiation of Kinematic Data Using Functional Data Analysis Approach: An Application of Automatic and Subjective Methods [J]. Applied Sciences-Basel, 10(7): 1-18.