Partial Functional Kriging Regression Models Based on LASSO And Group LASSO

Junnan Chen *, Xingzhan Lu

School of Science, Beijing Forestry University, Beijing, China, 100083

* Corresponding Author Email: chenjunnan0603@outlook.com

Abstract. For the regression problem containing continuous type response variables and mixed type explanatory variables (vector-valued and functional), this study proposes a partial function linear model incorporating random effects. The model assumes that the two types of explanatory variables are linearly related to the response variable, and their random effects obey a Gaussian process a priori. Methodological level: Functional Principal Component Analysis is used for basis function expansion of functional variables, while Lasso is combined for feature selection of vector-type variables, and Group Lasso is applied to realize group structure screening of functional variables. Parameter estimation is realized by the great likelihood method fusing L1 and Group L1 penalty terms. Numerical simulations show that the method can accurately identify the key features of the two types of variables, and demonstrates higher prediction accuracy than the traditional method in real data applications.

Keywords: Partial Function Linear Modeling, Kriging, Lasso, Group Lasso.

1. Introduction

With the development of sensor technology and continuous monitoring means, the object of data analysis gradually changes from scalar data to high-dimensional functional data. This concept was firstly proposed by Ramsay[1] , and its basic theory, analysis method and practical application were systematically summarized by Ramsay and Silverman[2] , which laid the basic framework of Functional Data Analysis. In recent years, Wang[3] et al. and Morris[4] have elaborated the basic forms and application paths of functional linear regression models from the review and methodological perspectives, respectively, and Greven and Scheipl[5] have constructed a unified framework for functional regression modeling, which shows the systematic development of the field. Although, a mature regression modeling framework has been developed for functional-type data analysis, in real life, the observed objects often have both functional-type features and discrete scalar features, such as the dynamic ECG signals of patients in clinical medicine with static indicators such as age and gender, and the absorption spectrum data of meat in different infrared bands with static indicators such as meat fat content, moisture, protein content, and so on. The prevalence of such Partially Functional Data poses a challenge to traditional modeling methods.

To address the dilemma of mixed data analysis, researchers often face three limiting strategies: one is to reduce functional variables to scalar treatment; the other is to force functionalization on non-functional variables; and the third is to model functional and scalar variables separately. These methods tend to weaken the model's ability to express the intrinsic smoothness and temporal correlation of functional data, and are more likely to lead to an imbalance between model complexity and predictive efficacy, deviation of fitting results from the true pattern, and bias in the inference system. In order to break through these limitations, Partially Functional Linear Model emerged and gradually developed into an important solution. Xue Zhang[6] et al. introduce a presmoothing estimation method to improve the consistency and stability of coefficient estimation. Aneiros and Vieu[7] extend the model to handle multiple functional covariates simultaneously to enhance its modeling flexibility. Kong[8] et al. extend the PFLM to a high-dimensional setting and propose a feasible variable selection mechanism. Ling[9] et al. further developed a k- nearest neighbor estimation method that combines linear and nonlinear components of the function, which significantly improves the prediction accuracy and adaptability of the model. Xiao[10] et al. proposed a generalized partially functional linear model, which can be used to deal with nonnormal response variables. Song-Xuan Li[11] et al. achieved functional covariate dimensionality reduction by principal component analysis, and combined maximum likelihood estimation and local linear regression to improve the modeling accuracy and adaptability. Xiao[12] et al. introduced the interaction between functional variables to enhance the model's ability to express the complex data structure. Sun[13] et al. introduced Gaussian process and integrated learning into PFLM to achieve better prediction performance. achieved better prediction performance.

However, traditional functional-type models usually assume that the random effects obey a specific parameterized distribution, which is often difficult to adequately portray the potential complex spatio-temporal heterogeneity in the data in practical applications. To break through this limitation, this paper proposes a partial functional linear model that integrates Gaussian process random effects with sparse cluster penalty. On the one hand, the method utilizes the Gaussian process to model the random effects nonparametrically, which enhances the model's ability to portray complex individual differences and dynamic structures; on the other hand, it realizes the automatic selection of high-dimensional individual variables and structural identification through the sparse clustering penalty strategy, which effectively improves the modeling efficiency and prediction accuracy while maintaining model interpretability. The model can be expressed as

$$Y = Z^T\beta + \sum_{j=1}^{T} \int X_j(t)\gamma_j(t)dt + f(t) + \varepsilon \tag{1}$$

where, Y is the response variable, Z^T is the linear predictor variable, β is its regression coefficient to be estimated, $X_j(t)$ is the functional type predictor variable, $\gamma_j(t)$ is its regression coefficient to be estimated, $f(t)$ is the nonlinear random effect term obeying a Gaussian process, and ε is the error term.

Referring to the literature[11], this paper uses functional principal component analysis to downscale the functional independent variable part, estimates the regression coefficients through the great likelihood estimation method, and penalizes the regression coefficients with the L1 paradigm as well as the group Lasso penalty to make them sparse, so as to achieve the purpose of variable selection. The results of the comparison between the prediction accuracy of this model and other basic methods are obtained through data simulation and real data analysis.

2. Theory and Methods

The aim of this paper is to construct a regression model that integrates linear and nonlinear features and possesses the ability to handle functional-type explanatory variables, in order to adapt to the problem of prediction and feature selection under high-dimensional and complex structures. To this end, a partially functional linear model is proposed, whose basic idea is to introduce the nonlinear modeling capability on the basis of the partially functional linear model and to combine the sparse modeling technique for variable screening, so as to enhance the model's expressive and explanatory capabilities.

Let the response variable be Y, the scalar covariate be $Z^T \in \mathbb{R}^p$, and the functional covariate be $X(t) \in L^2(T)$, and consider the following modeling form:

$$Y = Z^T\beta + \sum_{j=1}^{T} \int X_j(t)\gamma_j(t)dt + f(t) + \varepsilon \tag{2}$$

where $\beta \in \mathbb{R}^p$ is the vector of regression coefficients corresponding to the scalarcovariates, $\gamma(s) \in L^2(T)$ is the functional coefficient function, $f(t)\sim\mathcal{GP}(0, k(t,t'))$ denotes the nonlinear stochastic term modeled by a Gaussian process to capture complex variations that cannot be expressed by a linear structure, and $\varepsilon \sim N(0, \sigma^2)$ is the Gaussian white noise error term, independently and identically distributed.

In order to effectively deal with the functional covariates $X(t)$ and realize the parameter estimation, we use Functional Principal Component Analysis to downscale them and expand them into a set of linear combinations of orthogonal basis functions. Let $X_j(t)$ be expanded as follows:

$$X_j(t) \approx \sum_{k=1}^{K_j} a_{jk}\, \phi_{jk}(t) \tag{3}$$

where $\phi_{jk}(t)$ is the first k principal component basis function and a_{jk} is the principal component score. Further, the function type term may be expressed as:

$$\int X_j(t)\gamma_j(t)dt \approx \sum_{k=1}^{K_j} a_{jk}\, b_{jk} \tag{4}$$

where $b_{jk} = \int \phi_{jk}(t)\gamma_j(t)dt$. Combining all function terms and defining $A = [a_{11}, \dots, a_{1K_1}, \dots, a_{T1}, \dots, a_{TK_T}]$ as the principal component score matrix for all samples and $B = [b_{11}, \dots, b_{1K_1}, \dots, b_{T1}, \dots, b_{TK_T}]^T$ as the principal component coefficients of the function to be estimated, the overall model can be transformed into:

$$Y = Z^T\beta + Ab + f(t) + \varepsilon \tag{5}$$

Since $f(t) \sim \mathcal{GP}(0, k(t, t'))$, then Y as a whole also obeys a Gaussian process, the response variable Y obeys the following joint distribution:

$$Y \sim \mathcal{N}(Z\beta + Ab, K + \sigma^2 I).$$

where $K \epsilon \mathbb{R}^{n\times n}$ is the covariance matrix generated by the kernel function $k(t, t')$. Thus the log-likelihood function is:

$$\begin{aligned}\mathcal{L}(\beta, b, \theta) = -\frac{1}{2}(Y - Z\beta - Ab)^T(K + \sigma^2 I)^{-1}(Y - Z\beta - Ab)\\ -\frac{1}{2}log|K + \sigma^2 I| - \frac{n}{2}log2\pi\end{aligned} \tag{6}$$

θ is the parameter of the Gaussian process kernel function.

Since the scalar covariates Z and functional covariates $\mathrm{X}(t)$ often have high dimensionality and only some of them are significantly correlated with the response Y in real problems, it is necessary to introduce a variable selection mechanism in order to avoid overfitting and improve the model interpretability. In this paper, the sparse regularization strategy is used to penalize the above maximum likelihood function, and the following regularization objective function is constructed:

$$\begin{aligned}\mathcal{L}(\beta, b) = \frac{1}{2}(Y - Z\beta - Ab)^T(K + \sigma^2 I)^{-1}(Y - Z\beta - Ab)\\ +\lambda_1 \parallel \beta \parallel_1 + \lambda_2 \sum_{j=1}^{T} \parallel b_j \parallel_2\end{aligned} \tag{7}$$

where λ_1 controls the L1 penalty on the scalar coefficients, thus inducing some of the β_j contractions to zero for the purpose of variable selection; λ_2 controls the Group Lasso penalty for the overall selection of functional covariates by principal component groups, avoiding sparsity only on individual principal components.

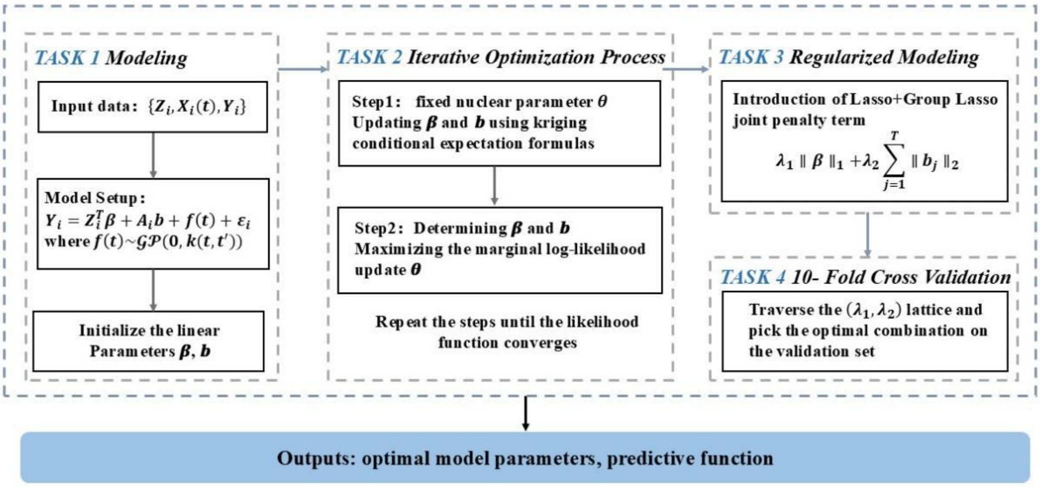

Figure 1 Model Flow Diagram

The solution process is based on Kriging as the core, nonlinear component fitting under Gaussian process modeling, and combined with the joint regularization strategy of Lasso and Group Lasso to achieve the enhancement of feature selection and model generalization ability. As shown in Figure 1, the whole method is iteratively updated between the kernel parameters and linear coefficients through an alternating optimization method, and the optimal regularization strength is finally determined through cross-validation, which achieves a dual balance between modeling accuracy and variable screening capability.

3. Data analysis

3.1. Simulation experiments

3.1.1 Data generation and experimental setup

In order to test the modeling and prediction ability of the proposed partially functional linear model with the addition of Gaussian process perturbation terms, this section constructs a simulation data environment for controlling the accuracy and complexity through Monte Carlo simulation and carries out a series of experiments for evaluating the parameter estimation and model prediction performance.

Assume the number of samples $n = 100$, and construct a time index by uniformly dividing 100 time points on the time region $T = [0,1]$. Utilizing the property of orthogonal Fourier basis expansion, the number of basis functions is set to 99, and two functional covariates $X^{(1)}(t)$ and $X^{(2)}(t)$ are generated. In order to effectively analyze the variable selection effect of the model, it is assumed that the first functional variable $X^{(1)}(t)$ is a valid covariate with real correlation with the response variable Y , and the second functional variable $X^{(2)}(t)$ participates in the modeling process as an irrelevant variable. In order to ensure the smoothness of the functional data in the high-frequency part and to avoid the interference of too much high-frequency noise, the first functional independent variable draws 99-dimensional Gaussian random coefficients for each sample, and the variance of each dimension is decayed at the rate of k^{-3} . The second functional independent variable is generated in the same way, but the coefficients decay at a slower rate ($k^{-3/2}$), making it more morphologically complex but not useful in the response variable.

In the construction of the functional coefficient function, the same Fourier basis expansion is used and the coefficients are artificially set to k^{-2}, thus ensuring that the functional coefficients themselves have a smooth, sparse structure. This coefficient function is then inner product with $X^{(1)}(t)$ to form the functional part of the response variable.

In the linear term part, a 10-dimensional independent and identically distributed Gaussian random variable is designed as a scalar covariate, whose covariance matrix is the unit matrix I_{10}, and the mean vector is 0. The true regression coefficients are set as a vector with a sparse structure, where the coefficients of the first three variables are set to 0, and the last seven variables are assigned positive values varying from 0.05 to 0.2, respectively. This setting allows the model to have both variable redundancy and effective signals in the true structure, which facilitates the variable selection ability of the assessment method.

In the design of the Gaussian process random effect term, we obtain the intermediate variable $S_i \sim \mathcal{N}(0,1)$, by sampling the standard normal distribution and set the random effect term to $f_i = \sin(\frac{2}{3}\pi S_i)$. This functional form ensures the nonlinear structure of the random effect term, which introduces a complex form of nonparametric noise that approximates the nonlinear or unknown disturbances that may exist in the actual data.

Ultimately, the response variable was constructed according to the following model:

$$Y_i = Z_i^T \beta + 0.01\langle X_i^{(1)}, \gamma \rangle + f_i + \varepsilon_i, \varepsilon_i \sim \mathcal{N}(0, 0.01^2) \quad (8)$$

Where $\langle X_i^{(1)}, \gamma \rangle$ is realized by an approximation of the inner product of the functions exhibited by the principal components and ε_i is the small perturbation error.

At the same time, we use functional principal component analysis (FPCA), by smoothing and mapping the functional data to the Fourier basis function space, then calculating the eigenfunctions and eigenvalues of the covariance operator, and retaining the first five principal component scores as the low-dimensional representation of the functional variables. In this way, the original high-dimensional function-type data are effectively compressed into a finite-dimensional matrix input, which makes the subsequent modeling more stable and computationally efficient.

In order to comprehensively evaluate the predictive performance and variable selection ability of the proposed partially functional linear Gaussian process models in practical modeling, we introduced ten control models for comparison in the subsequent experiments. These models cover two main types of modeling strategies: one type uses functional principal component analysis as a dimensionality reduction tool for functional data, and constructs a model in the form of "linear term + FPCA principal component + Gaussian process perturbation term", followed by XGBoost, random forest (RF), support vector regression (SVR), ElasticNet, and Gaussian Process Regression (GPR) as regressors for modeling; the other category combines the linear term, functional data (without separate dimensionality reduction), and Gaussian process term as the overall input variables, and applies the traditional Principal Component Analysis (PCA) for overall dimensionality reduction of the entire input space, and then regresses it using the same five machine learning models. modeling, constituting five combined models such as PCA+XGBoost and PCA+RF.

To ensure the fairness of the comparison and the robustness of the results, we generated data by repeated independent sampling in simulation experiments (the test set was fixed at 20% of the total dataset) and conducted multiple experiments using Monte Carlo methods to comprehensively compare the performance of each model in predicting the response variables in the test set. We selected three typical prediction error metrics for evaluation: mean square error (MSE), mean absolute error (MAE), and mean relative error (MRE). Among them, the mean error reflects the prediction accuracy of the model, while the standard deviation measures the stability of the prediction results. The combination of fixed test set proportion and random sampling is designed to avoid the impact of data delineation bias on the evaluation results and to ensure the benchmark consistency of model comparison. Ultimately, by comparing the performance of different models on various indicators horizontally (observe the box plots), the advantages of the proposed method in terms of accuracy and robustness can be objectively reflected.

3.1.2 Results

After completing the training and model fitting on the simulated data, we further evaluate the performance of the proposed partially functional linear model from several perspectives.

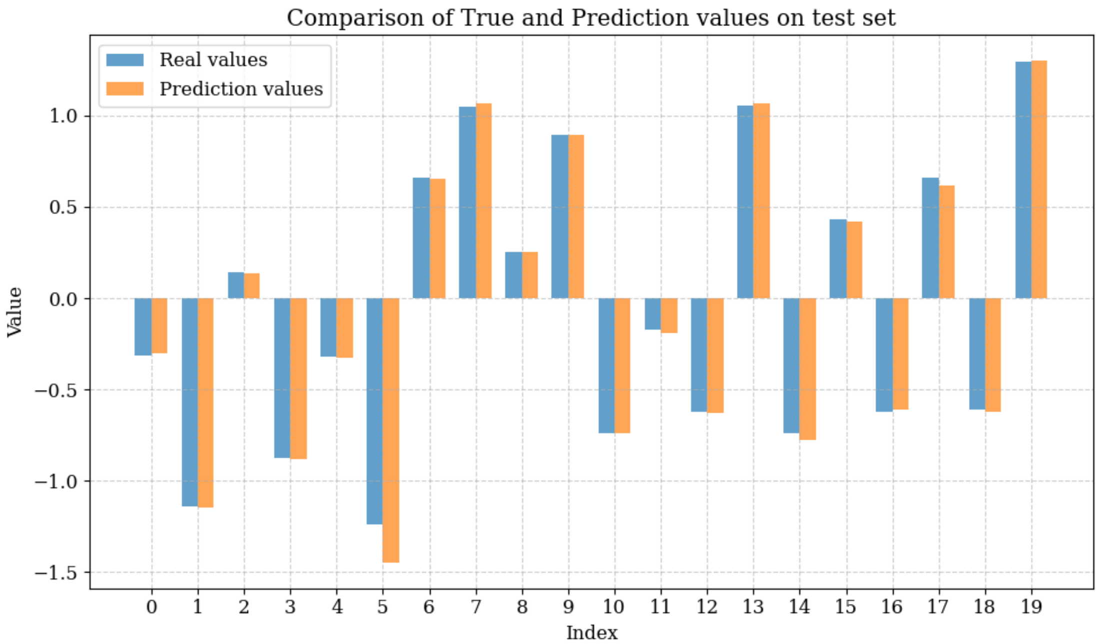

Figure 2 Comparison results between predicted and true values on the test set

In Figure 2, we show the comparison results between the predicted and true values on the test set. From the figure, we can see that the overall trend between the predicted values and the true values matches well, especially in most sample points, the two are highly overlapped, and there is only a certain deviation in some extreme points, which indicates that the model has a good performance in terms of the overall trend capture and specific value fitting, and it has a strong generalization ability.

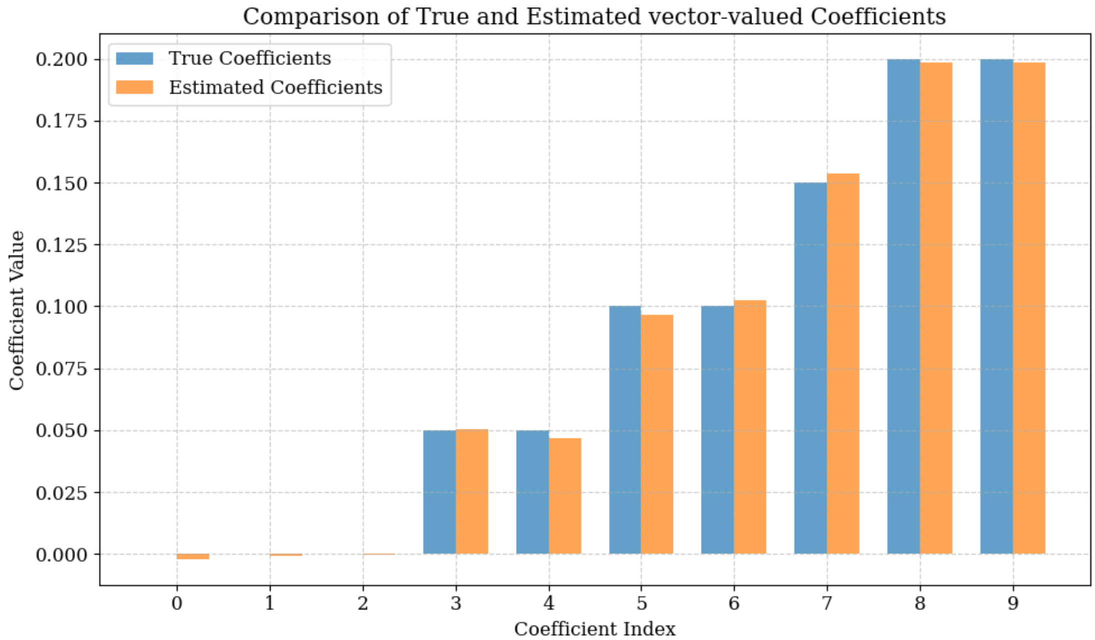

Figure 3 Coefficients β Comparison

Figure 3 demonstrates the comparison between the linear term coefficients $\hat{\beta}$ obtained from the model estimation and the true coefficients β from the data generation process. It can be observed that at the positions of non-zero coefficients, the estimated values are very close to the true values, indicating that the model has strong parameter estimation ability; while at the positions where the original true coefficients are zero (indexed 0, 1, 2), the model also accurately estimates the corresponding coefficients to near-zero values, demonstrating a good variable selection ability. This further verifies that the model in this paper can effectively identify the linear independent variables that have significant influence onYY after embedding the sparse regularization term, and at the same time eliminate the redundant features to enhance the model interpretability.

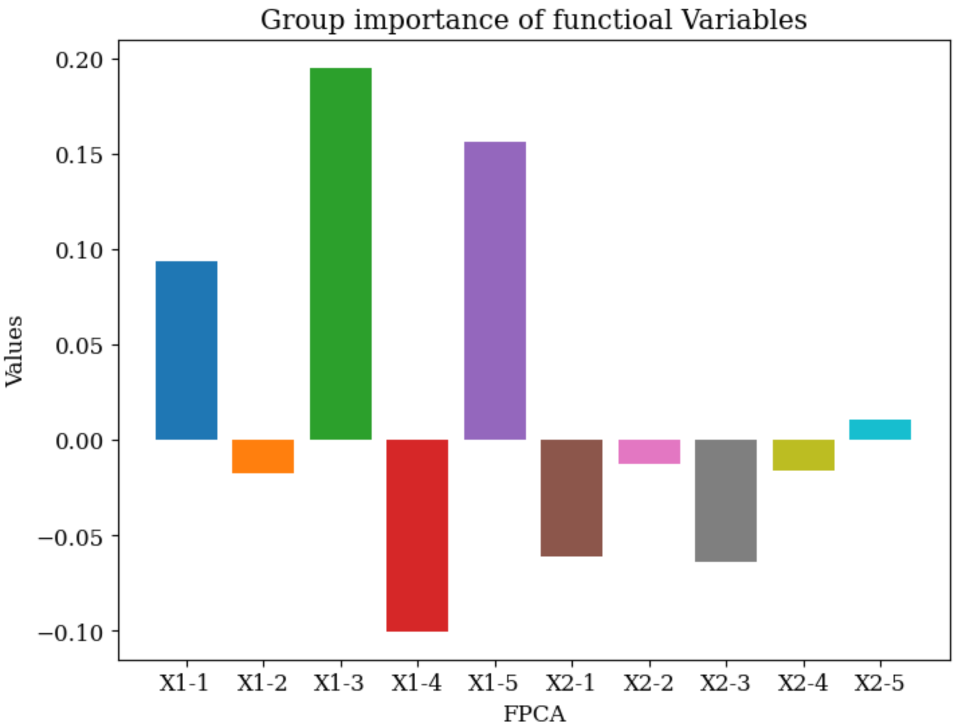

Figure 4 Degree of influence of principal components on Y

Figure 4 reflects the degree of influence of the principal components of some functional variables on the response variable Y after FPCA treatment. In the figure, the overall weight of the principal components of the $X1$ series (corresponding to the functional variables related to Y) is significantly higher than that of the $X2$ series (functional variables not related to Y), especially the influence of the $X1-3$, $X1-5$ principal component is the most prominent.This shows that FPCA can still retain the important information related to the response variables while downscaling the high-dimensional functional data, and at the same time make the influence of the principal components of the irrelevant variables close to zero, so as to achieve the purpose of feature screening and downscaling modeling. This also indirectly supports the effectiveness of the Group Lasso regular term in the model in the selection of group variables.

The mean and standard deviation of MSE, MAE and MRE for each model were calculated as shown in the table below and plotted as box line plots.

Table 1 MSE, MAE, MRE Means and Standard Deviations

	MSE	**SD-1**	**MAE**	**SD-2**	**MRE**	**SD-3**
LASSO_FSGPR	0.0017	0.0037	0.0181	0.0073	0.0564	0.0428
FPCA+XGBoost	0.2008	0.1462	0.3133	0.0771	0.8779	0.2842
FPCA+RF	0.1353	0.0738	0.2890	0.0466	0.8271	0.3374
FPCA+SVR	0.3182	0.1299	0.4680	0.0800	0.9433	0.2335
FPCA+ElasticNet	0.5677	0.1045	0.6605	0.0872	1.0297	0.0600
FPCA+GPR	0.3307	0.1675	0.4538	0.0718	0.9839	0.2567
PCA+XGBoost	0.7280	0.1158	0.7152	0.0955	1.4003	0.3000
PCA+RF	0.6071	0.0687	0.6774	0.0498	1.1900	0.2962
PCA+SVR	0.6401	0.0524	0.6721	0.0533	1.2022	0.3547
PCA+ElasticNet	0.5769	0.0908	0.6695	0.0778	1.0419	0.0910
PCA+GPR	0.5560	0.1106	0.6524	0.0922	1.0000	0.0000

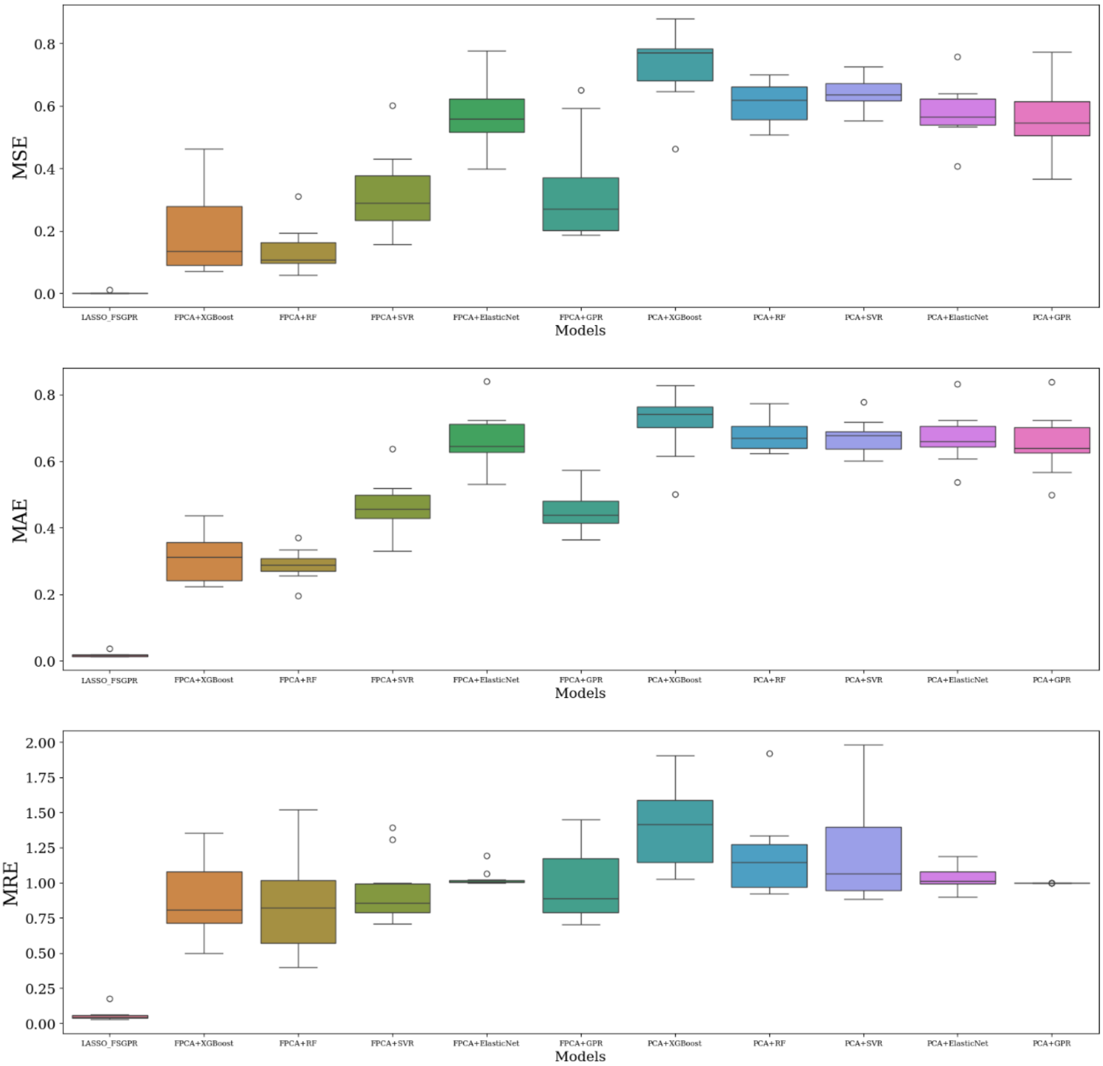

Figure 5 Comparison of MSE, MAE, and MRE boxplots across models

From the boxplot, the LASSO_FSGPR model has the lowest box position on the MSE, the MAE, and the MRE, the values are much smaller than those of other models, the box is compact, and the upper and lower quartile distances are extremely small. This indicates that the model has low sensitivity to outliers and stable prediction results; more consistent prediction accuracy on different samples and stronger robustness to noise; very high prediction accuracy, stable prediction error on different magnitude data, strong generalization ability, and significant advantages and high reliability in practical applications.

3.2. Real data analysis

A publicly available infrared spectral dataset of meat samples is used as an example. This data was obtained from the Carnegie Mellon University Department of Statistics Data Warehouse (http://lib.stat.cmu.edu/datasets/tecator) and contains 100 wavelength-point spectral profiles for a total of 215 meat samples, as well as corresponding moisture, fat, and protein content metrics. The moisture content of this sample was predicted based on the absorption spectral profile, and fat and protein content. Since it was not possible to determine whether the protein and fat content had only a linear or non-linear relationship with moisture, both components were considered for inclusion in the model. Denote Y as the moisture content, Z as the fat and protein content, $X(t)$ as the infrared absorption spectrum profile, and $f(t)$ as the fat and protein content.

In the data preprocessing stage, the original data were firstly transformed into functional data by Fourier basis expansion to realize smoothing, and then the high-dimensional functional features were downscaled by FPCA. Based on the processed data, a regression model was built to predict the moisture content of the samples and systematically compared with the other 10 benchmark models. The analysis process is consistent with Section 3.1.2: the same parameter configurations are used for model training, and box-and-line plots and error tables are used to demonstrate the prediction performance of the different methods.

Table 2 Real Data MSE, MAE, MRE Means and Standard Deviations

	MSE	SD-1	MAE	SD-2	MRE	SD-3
LASSO_FSGPR	0.0856	0.0149	0.2117	0.0104	0.2117	0.0104
FPCA+XGBoost	0.1573	0.0356	0.2636	0.0240	0.2636	0.0240
FPCA+RF	0.1333	0.0304	0.2619	0.0239	0.2619	0.0239
FPCA+SVR	0.2016	0.0566	0.3010	0.0330	0.3010	0.0330
FPCA+ElasticNet	0.6414	0.1022	0.6851	0.0537	0.6851	0.0537
FPCA+GPR	0.8464	0.1029	0.7339	0.0450	0.7339	0.0450
PCA+XGBoost	0.1777	0.0717	0.2719	0.0385	0.2719	0.0385
PCA+RF	0.1671	0.0668	0.2705	0.0295	0.2705	0.0295
PCA+SVR	0.2459	0.0643	0.3327	0.0382	0.3327	0.0382
PCA+ElasticNet	0.5620	0.0838	0.6301	0.0490	0.6301	0.0490
PCA+GPR	0.3288	0.0786	0.3710	0.0453	0.3710	0.0453

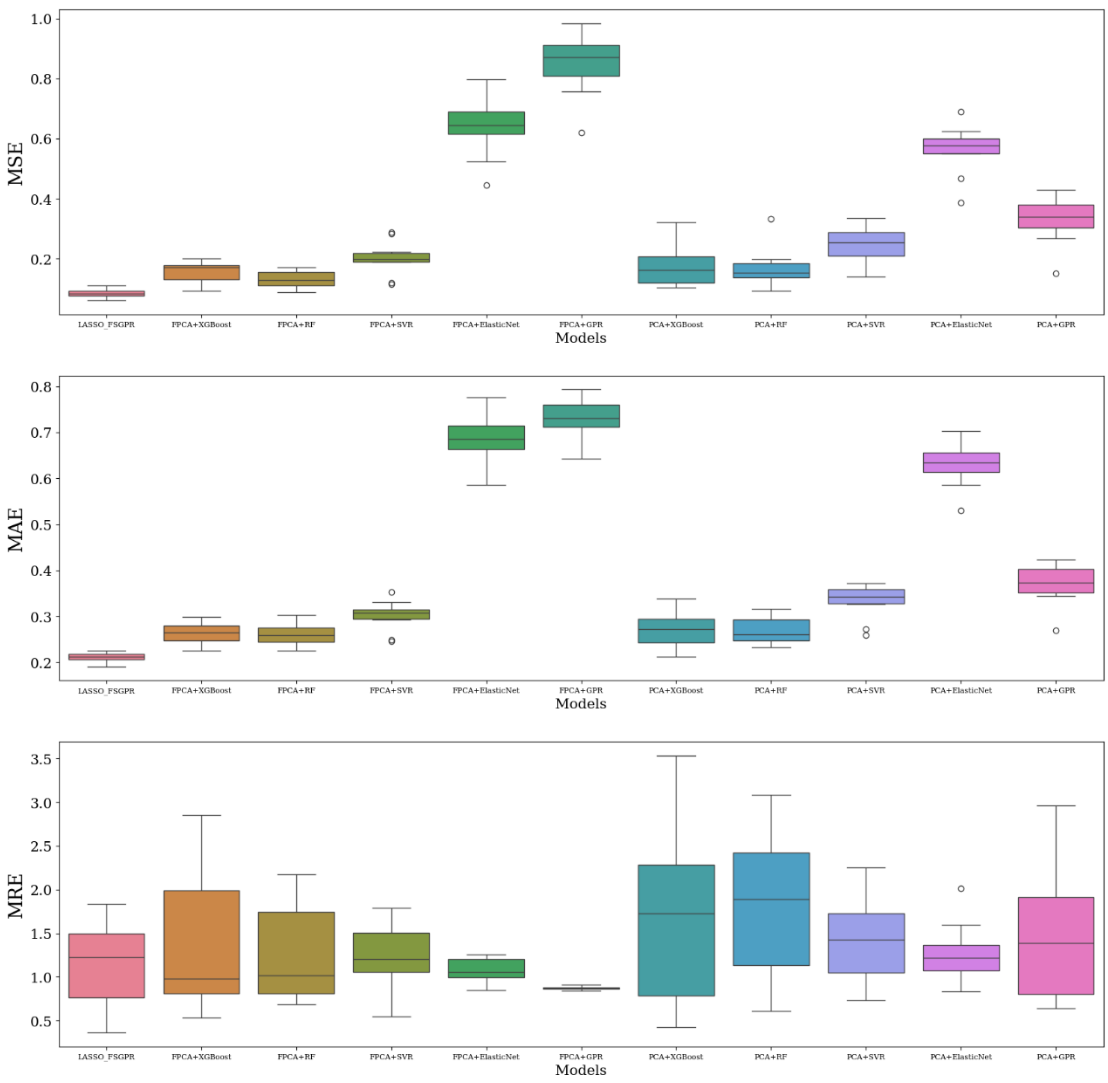

Figure 6 Comparison of MSE, MAE and MRE box plots across models for real data

From the box plot, in terms of MSE, the box of the LASSO_FSGPR model is at the bottom of all models, with values significantly lower than those of the other comparison models and small upper and lower interquartile spacing, indicating a low squared average of prediction error, low data dispersion, low sensitivity to outliers, stable prediction results, and a good fitting effect. In terms of MAE, the LASSO_FSGPR model has a low box position and a relatively small MAE value, which means that the predicted value deviates from the true value to a low degree, and the interquartile spacing is narrow, the prediction error fluctuation range is narrow, and the prediction accuracy is more consistent on different samples. In the MRE index, the LASSO_FSGPR model box position is in the middle position, the MRE value is in the middle, and it has some generalization ability.

A comprehensive box plot analysis of the MSE, MAE and MRE indicators shows that the LASSO_FSGPR model performs equally well in terms of predictive accuracy and stability in real-life data, and slightly outperforms the other compared models.

4. Conclusions

In this paper, a partial functional linear regression model based on LASSO, Group Lasso and Gaussian process (Kriging) is proposed to address the characteristics of response variables as scalars and explanatory variables containing mixed vector-type and function- type data. The study focuses on solving the limitations of traditional methods in dealing with mixed data types, such as model bias due to ignoring functional properties, dimensionality catastrophe under high-dimensional data, and insufficient modeling of nonlinear effects. The high-dimensional functional variables are downscaled by Functional Principal Component Analysis to preserve their key features; Lasso regularization is introduced to sparsify the scalar variables for screening, and Group Lasso regularization selects the principal component groups of functional variables as a whole, so as to achieve efficient variable selection while downscaling; and it is further combined with Gaussian process modeling to capture the complicated nonlinear random effects in the data to enhance the expressive power of the model. In the experiments with simulated data and real data (meat infrared spectral dataset), the prediction errors (e.g., MSE, MAE) of the proposed model (LASSO_FSGPR) are significantly lower than those of the comparison models, such as FPCA + XGBoost, PCA+RF, etc., and it can accurately identify the effective variables (e.g., zero coefficient estimation tends to be close to zero), which verifies its prediction accuracy, stability, and interpretive comprehensive advantages.

Despite the excellent performance of the model, there is still some room for improvement. First, the computational complexity of the Gaussian process is affected by the inverse operation of the kernel matrix, and its time complexity is $O(n^3)$, which may face a computational bottleneck in large-scale data scenarios, and the balance between computational efficiency and prediction accuracy can be achieved by using the sparsified Gaussian process regression model in the future. Second, the performance of Gaussian process depends on the selection of kernel function, although the feature selection module of the model is not highly related to the selection of kernel function, and the regenerative kernel Hilbert space of the RBF kernel is widely adaptable to smooth function data, but it still needs to be combined with flexible parameterization of kernel function such as Matern kernel to validate it for non-smooth or high-frequency oscillatory data, and the subsequent research can establish a kernel function Adaptive evaluation mechanism can be established to enhance the generalizability. In addition, the existing models have not fully explored the modeling of the interaction between functional variables and scalar variables, and the nonlinear interaction effects between multivariate variables can be explicitly portrayed by constructing an additive Gaussian process regression framework, introducing interaction terms in the kernel function, or adopting a hierarchical covariance structure. It is worth noting that Group Lasso imposes uniform penalty coefficients on all principal component groups of functional variables without considering the importance differences of different variable groups, which can be dynamically adjusted to enhance the selection flexibility. Finally, the direction of model extension can integrate integrated learning strategies and uncertainty quantification methods, such as coordinating multi-core Gaussian process models through the Stacking integration framework, or introducing Bayesian posterior sampling mechanism, which can provide reliable confidence interval estimation while improving the robustness of prediction, so as to expand the applicability boundaries of the model in complex application scenarios.

References

[1] Ramsay J O. When the data are function [J]. Psychometrika, 1982, 47 (3): 379 –396.

[2] Ramsay J O, Silverman B W. Functional Data Analysis (second edition) [M]. New York: Springer, 2005.

[3] Wang J L, Chiou J M, Müller H G. Functional data analysis[J]. Annual Review of Statistics and its application, 2016, 3(1): 257-295.

[4] Morris J S. Functional regression[J]. Annual Review of Statistics and Its Application, 2015, 2(1): 321-359.

[5] Greven S, Scheipl F. A general framework for functional regression modelling[J]. Statistical Modelling, 2017, 17(1-2): 1-35.

[6] Xue Zhang, Yuan Tian, Dehui Wang. Presmooth estimation for partial functional linear regression models[J]. Journal of Jilin University (Science Edition), 2014, 52(04): 715-719.

[7] Aneiros G, Vieu P. Partial linear modelling with multi-functional covariates[J]. Computational Statistics, 2015, 30: 647-671.

[8] Kong D, Xue K, Yao F, et al. Partially functional linear regression in high dimensions[J]. Biometrika, 2016, 103(1): 147-159.

[9] Ling, N., Aneiros, G. and Vieu, P. kNN Estimation in Functional Partial Linear Modeling[J]. Statistical Papers, 2020, 61: 423-444.

[10] Xiao W, Wang Y, Liu H. Generalized partially functional linear model[J]. Scientific reports, 2021, 11(1): 23428.

[11] Songxuan Li, Kejing Mao, Weiwei Xiao. Regression models and applications for partial functional data[J]. Advances in Applied Mathematics, 2023, 12(6): 2758-2764.

[12] Xiao W, Mao K, Liu H. Generalized Partially Functional Linear Model with Interaction between Functional Predictors[J]. Axioms, 2024, 13(9): 583.

[13] Sun W, Xu J, Liu T. Partially Functional Linear Regression Based on Gaussian Process Prior and Ensemble Learning[J]. Mathematics, 2025, 13(5): 853.