

# The Functional Regression With Reconstructed Functions From Hybrid Principal Components Analysis: With EEG-fMRI Application

Mohammad Fayaz<sup>1,\*</sup>, Alireza Abadi<sup>2,3</sup>, Soheila Khodakarim<sup>4</sup>

<sup>1</sup> PhD in Biostatistics, Department of Biostatistics, School of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran. (<https://orcid.org/0000-0002-5643-9763>)

<sup>2</sup> Professor of Biostatistics, Department of Community Medicine, Faculty of medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran. (<http://orcid.org/0000-0003-2653-6623>)

<sup>3</sup> Social Determinants of Health Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran

<sup>4</sup> Soheila Khodakarim, Associate Professor, Department of Biostatistics, School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran. (<https://orcid.org/0000-0002-5473-999X>)

**Abstract** Objective: In this article, we reconstruct the hybrid data with hybrid principal component analysis (HPCA) as a feature extraction step and modeling them with functional regression as a modeling step, and comparing models and choosing the best number of HPCA based on the prediction accuracy as an evaluation step. Method: We decompose the hybrid data to the eigencomponents with HPCA. The reconstructed data from HPCA were divided into the training and testing dataset. The function-on-function signal compression and bayesian scalar-on-function regressions were used. Three simulation scenarios and their applications in the neuroimaging datasets (EEG-fMRI) were studied. The number of HPCA was selected with the mean squared prediction error (MSPE). Result: One of the simulations shows that the raw data, reconstructed from the first and all HPCAs for the training dataset has a median MSPE of 0.1001, 0.0028, and 0.0174 respectively, and the testing dataset has 0.3207, 0.1118, and 0.2484 respectively. The EEG-fMRI suggests that in both auditory and visual tasks and standard and target stimuli for a different region of the brain the first few HPCA has the smallest MSPE. Conclusions: We conclude this method improves the prediction accuracy of the experiments with the EEG datasets. And we recommend that instead of using the functional PCA on the desired dimension, reconstruct the data with HPCA and average it on the other two dimensions for functional regression models.

**Keywords** Functional Data Analysis, Hybrid Principal Component, Multiple Functional Covariates, Functional Regression, EEG, fMRI

AMS 2010 subject classifications 62M10 , 62F15

DOI: 10.19139/soic-2310-5070-1437

## 1. Introduction

The principal component regression (PCR) was based on the principal component analysis (PCA) and it was used in the regression settings to tackle the collinearity problem among covariates[1, 2, 3]. By an introduction to the functional data[4], the functional extension of Principal Component Analysis (FPCA) captured the underlying curves of data and reduced their dimensionality[5]. We count different methods to estimate the number of FPCA such as the fraction of variance explained (FVE), comparing the log-likelihood of different models with different numbers of FPCAs[6], Likelihood Ratio Test (LRT)[7],

---

\*Correspondence to: Mohammad Fayaz (Email: [Mohammad.Fayaz.89@gmail.com](mailto:Mohammad.Fayaz.89@gmail.com)). PhD in Biostatistics, Department of Biostatistics, School of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran. (<https://orcid.org/0000-0002-5643-9763>)

the mean square error (MSE), mean square prediction error (MSPE) [8] and Bayesian information criteria (BIC)[9]. FPCA methods extend in different ways to deal with; for example, longitudinal dimension[10, 11], multilevel structure[12], multidimensional[13], and spatially dimension [7, 14, 15, 16].

The scalar-on-function (SOF), function-on-scalar (FOS), and function-on-function (FOF) regression and different supervised learning models use FPCA scores instead of functional data. [4, 5, 17]. Some of them are adapted for predictions such as signal compression estimations for modeling EEG and many regions of interest (ROI) of Functional magnetic resonance imaging (fMRI) [18, 19] and Bayesian estimations with applications to the  $\gamma$ -power of electroencephalography (EEG) data and intraday physical activity with many covariates in the sleep study [20, 21].

The hybrid PCA (HPCA) was introduced for situations that exist in both scalar and functional data, called mixed or hybrid data, [4]. For example, the (EEG) data were analyzed with HPCA to extract the functional, longitudinal, and regional dimensions: frequency or time domain ERPs, the repeating stimuli, and the brain regions, respectively. The first two of them have functional characteristics, and the last one is scalar. HPCA extracts the subject-specific scores and eigencomponents of hybrid data [22]. Among the regression models which can deal with hybrid data, we can count covariate-adjusted HPCA [23], Semi-functional partial linear regression [24, 25], FOF regression with signal compression [18], and functional additive regression models [26].

In this research, first, we decompose the hybrid data with HPCA. Second, we reconstruct the hybrid data with a minimum number of eigenfunctions and eigenvectors in each dimension (as a feature selection extractions step). Third, we calculate the functional average along the desired dimension of reconstructed hybrid data. Fourth, we use two functional regression models including FOF regression with signal compression and Bayesian SOF regression (as a supervised learning step). Fifth, we calculate the prediction accuracy (ACC) and mean square prediction error (MSPE) of the model in the training and test dataset. Sixth, we repeat the first to fifth steps by adding one by one the number of eigenfunctions and eigenvectors of each dimension. Seventh, we compare all models with ACC and MSE and we choose the best number of HPCA that has the minimum MSPE in the FOF regression and we also compute Deviance Information Criteria(DIC) and log pseudo-marginal likelihood (LPML) indices in the Bayesian SOF regression. (as the model evaluation steps). Finally, we applied our methodology to the EEG-fMRI dataset. The paper has the following sections, 1-introduction contains the problem definitions with some background, 2- The functional regression with HPCA contains HPCA decomposition, FOF with multiple functional covariates and Bayesian SOF, 3- The number of HPCA section contains a methodology to choose the best number of HPCA according to the correlation, MSPE, etc., 4- Simulation section contains three simulation scenarios, 5- Application to EEG-fMRI dataset, 6- Conclusion contains the comparative study with some comments, and 7- References. Moreover, we provide supplementary materials, including further analysis and R codes.

## 2. Functional Regression with HPCA

### 2.1. Functional PCA

The PCA is a multivariate method in statistics that is used for dimension reduction and it estimates the eigenvectors and related scores. In recent advances, from one point by introducing high-dimensional datasets in the application such as neuroimaging datasets, remote sensing, weather stations, etc., and from other points developing the statistical methodologies, the multivariate methods are not well-adapted to work with infinite dimension datasets. In this regard, the functional data analysis (FDA) proposed a methodology that first interpolates (without measurement noise) or smoothing (with measurement noise) functional data with basis functions such as B-Spline, Fourier ,etc. They estimate the coefficients and use them in other methods and they introduce FPCA, Functional Regression, Functional ANOVA, Functional Canonical Correlation Analysis and also provide to work with the derivatives of the smoothed datasets like Phase-Plane plots and Principal Differential Analysis (PDA). The FPCA is estimated with different

methodologies and it decomposes the dataset into eigenfunction and related scores. An adequate number of FPCA is obtained with different methods such as the fraction of variance explained (FVE), Akaike information criteria (AIC), and Bayesian information criteria (BIC) and it is a dimension reduction methodology for functional data. [4].

## 2.2. HPCA Decomposition

By the same notation and methodology of [22], we mention HPCA decomposition briefly in this section. Assume  $Y_i(r, \omega, s)$  has three dimensions of the de-mean observation for the subject  $i, \{i = 1, \dots, n\}$ , the first dimension has the scalar domain  $r, \{r = 1, \dots, R\}$ , the second and third dimensions have the functional domain  $\omega, \omega \in \Omega$  and  $s, s \in S$ , respectively. We want to decompose  $Y_i(r, \omega, s)$  with HPCA to the eigencomponents and subject-specific scores. In this regard, we first estimate the covariance matrices. Because there are three-dimensional matrices, we can estimate the eigencomponents of each dimension by marginal covariance and product them under the weak separability assumptions of covariances. In practice, we have equation 1 :

$$Y_i(r, \omega, s) = Z_i(r, \omega, s) + \epsilon_i(r, \omega, s) \approx \sum_{k=1}^K \sum_{l=1}^L \sum_{m=1}^M \xi_{i,klm} V_k(r) \phi_l(\omega) \psi_m(s) + \epsilon_i(r, \omega, s) \quad (1)$$

$V_k(r)$  is the marginal eigenvector of the first dimension which has a scalar domain,  $\phi_l(\omega)$  is the marginal eigenfunction of the second dimension which has a functional domain and  $\psi_m(s)$  is the marginal eigenfunction of three dimensions which has a functional domain.  $\xi_{i,klm}$  is the subject-specific scores obtained using the following inner product, equation 2:

$$\xi_{i,klm} = \langle Z_i(r, \omega, s), V_k(r) \phi_l(\omega) \psi_m(s) \rangle = \sum_{r=1}^R \iint Z_i(r, \omega, s) V_k(r) \phi_l(\omega) \psi_m(s) d\omega ds \quad (2)$$

$K$ ,  $L$ , and  $M$  are the number of eigencomponents that marginally have at least 90% fraction of variation (FVE) of each dimension. Therefore, it is truncated to the maximum number of KLM eigencomponents.  $\epsilon_i(r, \omega, s)$  are error terms.[22]

## 2.3. Function-on-multiple-Functions Regression

In this section, we use FOF regression [19]; the response and covariates are functional data. For example, Evoked Related Potential (ERP) curves on time or frequency domains. We first decompose hybrid data with HPCA and then reconstruct them with different numbers of HPCAs. The reconstructed hybrid data have three dimensions, and we merge them into one dimension with a functional domain. Many applications work only with one dimension. For example, in the EEG dataset,  $Y_i(r, \omega, s)$ , we are usually only interested in values of ERP at the time domain,  $\omega$ . The other two dimensions are the regions with the scalar domain,  $r$ , and the longitudinal dimension,  $s$ , with the functional domain.

Therefore, we first reconstruct the dataset in equation 3 with the specified number of HPCA,  $Y_i(r, \omega, s)_{HPCA\_KLM}$ . We use hat notation because it estimates the real observed value,  $Y_i(r, \omega, s)$  and we subscript  $HPCA\_KLM$  to show the number of HPCAs; for example, KLM. In this step, we use the most variations of three dimensions (usually at least 90%), and we reconstruct the data based on them; therefore, many unimportant variations were omitted. Second, we merge  $Y_i(r, \omega, s)_{HPCA\_KLM}$  on other two dimensions; for example, on the first,  $r$ , and third,  $s$ , dimensions. As stated in equation 4, we average point-wise on desired dimension domain, for example, the second dimension,  $\omega$ .

$$\sum_{k=1}^K \sum_{l=1}^L \sum_{m=1}^M \xi_{i,klm} V_k(r) \phi_l(\omega) \psi_m(s) = \hat{Y}_i(r, \omega, s)_{HPCA\_KLM}. \quad (3)$$

$$\frac{1}{R} \sum_{r=1}^R \int_{-\infty}^{\infty} \hat{Y}_i(r, \omega, s)_{HPCA\_KLM} ds = \hat{W}_i(\omega)_{KLM}. \quad (4)$$

The advantage of this method instead of using only FPCA on the desired dimension,  $\omega$ , is that our method contains the most important variations of the other two dimensions in the dataset and it avoids

considering the unimportant variations on the data which help reduce its complexity. The functional variable,  $\hat{W}_i(\omega)_{KLM}$ , with the functional domain,  $\omega$ , and the number of HPCs,  $KLM$ , is calculated by equation 4 for regression analysis in the next steps.

We assume that  $\hat{W}_i(\omega)_{KLM}$  in equation 4 is a response function in the following FOF regression [19]:

$$\hat{W}_i(\omega)_{KLM} = m(\omega) + \sum_{j=1}^p \int_0^1 \chi_j(g)\beta_j(g, \omega) dg + \varepsilon(\omega) \tag{5}$$

$\chi_1(g), \dots, \chi_p(g)$  are predictive functions and  $\beta_1(g, \omega), \dots, \beta_p(g, \omega)$  are p-coefficient functions,  $m(\omega)$  and  $\varepsilon(\omega)$  are intercept function and noise function. The focus of this model with signal compression methods is predicting the response [19].

### 2.4. Scalar-on- Functions Bayesian Regression

In this section, we use the Bayesian scalar-on-function regression [20]. The response is from the exponential distribution family and the functional covariate was obtained from HPCA decomposition. We assume that  $[U_i, \hat{W}_i(\omega), \omega \in \Omega]$  is a classical SOF regression in which  $U_i$  is the outcome and has an exponential family with the linear predictor  $v_i$  and dispersion parameter  $a$ ,  $EF(v_i, a)$ . The functional covariates  $\hat{W}_i$  are random curves that are obtained from merging  $\hat{Y}_i(r, \omega, s)_{HPCA\_KLM}$  on  $r$  and  $\omega$ , equation 4.

We assume that  $\hat{W}_i(\omega)_{KLM}$  have three parts in equation 6: the first is  $\hat{\mu}(\omega)$  which is the estimation of population average, the second is the estimation of the underlying curve which is  $\hat{X}_i(\omega)$ , and the third is  $\hat{\varepsilon}_i(\omega)$  which is an error term. We de-mean  $\hat{W}_i(\omega)_{KLM}$  by  $\hat{W}_i(\omega)_{KLM} = \hat{W}_i(\omega)_{KLM} - \hat{\mu}(\omega)$ ,  $\hat{\mu}(\omega) = \frac{1}{I} \sum_{i=1}^I \hat{W}_i(\omega)$ .

$$\hat{W}_i(\omega)_{KLM} = \hat{\mu}(\omega) + \hat{X}_i(\omega) + \hat{\varepsilon}_i(\omega) \tag{6}$$

The linear predictor has the following form:

$$v_i = \int_0^1 \hat{X}_i(\omega)\beta(\omega) d\omega \tag{7}$$

The  $\beta(\cdot) \in L_2[0, 1]$  is a functional parameter. We assume that  $\{\eta_d, d \geq 1\}$  is an orthonormal basis in  $L_2[0, 1]$  then we have  $\hat{X}_i(\omega) = \sum_{d \geq 1} \zeta_{id}\eta_d(\omega)$ ,  $\beta_i(\omega) = \sum_{d \geq 1} \beta_d\eta_d(\omega)$  then formula 7 is  $v_i = \sum_{d \geq 1} \zeta_{id}\beta_d$ . Therefore, we write them as the following mixed-effects model, equation 8:

$$\left\{ \begin{array}{l} U_i \sim EF(v_i, a); \\ v_i = \sum_{d \geq 1} \zeta_{id}\beta_d; \\ \hat{W}_i(\omega)_{KLM} = \sum_{d \geq 1} \zeta_{id}\eta_d(\omega) + \hat{\varepsilon}_i(\omega) \\ \zeta_{id} \sim N(0, \lambda_k) \\ \varepsilon_i(\omega) \sim N(0, \sigma_\varepsilon^2) \end{array} \right. \tag{8}$$

The parameters were estimated with the Markov Chain Monte Carlo (MCMC) samplers.

### 3. The number of HPCA

The number of appropriate HPCA for  $Y_i(r, \omega, s)$  in equation 1 is discussed at [22] by estimating the fraction of variance explained (FVE) in the mixed-effects models and the variance components estimations. In this section, we propose a new method to select the number of HPCA due to the mean squared prediction error (MSPE) of regression models. In equation 1, we have KLM eigencomponents, but instead of using all of them in equation 3, we add each of them at each step in equation 9:

$$\left\{ \begin{array}{l} \xi_{i,1}V_1(r)\phi_1(\omega)\psi_1(s) = \hat{Y}_i(r, \omega, s)_{HPCA\_1} \\ \xi_{i,1}V_1(r)\phi_1(\omega)\psi_1(s) + \xi_{i,2}V_2(r)\phi_2(\omega)\psi_2(s) = \hat{Y}_i(r, \omega, s)_{HPCA\_2} \\ \dots \\ \sum_{k=1}^K \sum_{l=1}^L \sum_{m=1}^M \xi_{i,klm}V_k(r)\phi_l(\omega)\psi_m(s) = \hat{Y}_i(r, \omega, s)_{HPCA\_KLM}. \end{array} \right. \tag{9}$$

Therefore,  $\{\hat{Y}_i(r, \omega, s)_{HPCA\_q}, 1 \leq q \leq KLM\}$  are estimations due to the first  $q$  eigencomponents. We transform each of them with equation 4 to  $\{\hat{W}_i(\omega)_{HPCA\_q}, 1 \leq q \leq KLM\}$ . For equation 5, we fit the regression model to all KLM models with  $\hat{W}_i(\omega)_{HPCA\_q}$  as a response curve and we predict  $\{\hat{W}_i(\omega)_{HPCA\_q}, 1 \leq i \leq n, 1 \leq q \leq KLM\}$  (Algorithm-1) (Table 1)

Algorithm 1 –FOF Regression	
1-	The HPCA decomposition $Y_i(r, \omega, s) \approx \sum_{k=1}^K \sum_{l=1}^L \sum_{m=1}^M \xi_{i,klm}V_k(r)\phi_l(\omega)\psi_m(s) + \epsilon_i(r, \omega, s)$ (eq. 1 )
2-	for $q = 1$ to $q = KLM$
2-1	The reconstruction due to HPCA from 1 to $q$ : $\sum_{k=1}^K \sum_{l=1}^L \sum_{m=1}^M \xi_{i,klm}V_k(r)\phi_l(\omega)\psi_m(s) \approx \hat{Y}_i(r, \omega, s)$ (eq. 3 )
2-2	Merging on two dimensions: $\frac{1}{R} \sum_{r=1}^R \int_{-\infty}^{\infty} \hat{Y}_i(r, \omega, s)_{HPCA\_KLM} ds = \hat{W}_i(\omega)_{KLM}$ . (eq. 4 )
2-3	Fit FOF due $\hat{W}_i(\omega)_{KLM}$ $\hat{W}_i(\omega)_{KLM} = m(\omega) + \sum_{j=1}^p \int_0^1 \chi_j(g)\beta_j(g, \omega) dg + \varepsilon(\omega)$ (eq. 5 )
2-4	Predict $\hat{W}_i(\omega)_{HPCA\_q}$ within (2-3)
2-5	Calculate MSPE Mean Squared Prediction Error $MSPE(q) = \frac{1}{(n \times M)} \sum_{i=1}^n \sum_{m=1}^M (\hat{W}_i(\omega_m)_{HPCA\_q} - \hat{W}_m(\omega_m)_{HPCA\_q})^2$ $\{i, 1 \leq i \leq n, \text{number of subject}\}$ $\{m, 1 \leq m \leq M, \text{number of observation in each curve}\}$
3-	Find the minimum $MSPE(q_{min}) = \min\{q = 0 \text{ or } MSPE(q), 1 \leq q \leq KLM\}$
4-	The best number of HPCA is $q_{min}$

Table 1. Algorithm 1 - FOF Regression

For equation 8, we fit the regression model to all KLM models with  $\hat{W}_i(\omega)_{HPCA\_q}$  as predictive curves and we estimates  $\{\hat{U}_{i,HPCA\_q}, 1 \leq i \leq n, 1 \leq q \leq KLM\}$  (Algorithm-2) (Table 2)

This methodology can be applied to any other regression problem which deals with the hybrid data.

#### 4. Simulations

##### 4.1. Simulation Scenario 1

In this scenario, we evaluate the effect of HPCA on the function-on-function regression in two small and big sample size and complete and sparse observations for observed functions. We simulate dataset in two steps: We first generate  $Y_i(r, \omega, s)$  with two sample size, 1)  $\{i, i = 1, \dots, 20\}$  and 2)  $\{i, i = 1, \dots, 140\}$  and two status for the 1) complete and 2) sparse observations for  $s$  as mentioned in the [22], then we merge each of them with equations 1 to 5, as  $\hat{W}_i(\omega)_{HPCA\_0}$ . In the second step, we generate the fof regression data as mentioned at [19] with 10 covariates functions,  $\theta(g, \omega)$ , and two status for coefficients 1) all 10 coefficients,  $\theta(g, \omega)$ , have a value greater than 0 and 2) only 50% of  $\theta(g, \omega)$  have non-zero value. Therefore, we have  $A_i(\omega) = m(\omega) + \sum_{j=1}^{10} \int_0^1 \chi_j(g)\theta_j(g, \omega)dg + \varepsilon(\omega)$ . We combine the first and

Algorithm 2 –Bayesian SOF Regression	
1-The HPCA decomposition $Y_i(r, \omega, s) \approx \sum_{k=1}^K \sum_{l=1}^L \sum_{m=1}^M \xi_{i,klm} V_k(r) \phi_l(\omega) \psi_m(s) + \epsilon_i(r, \omega, s)$ (eq. 1)	
2- for $q = 1$ to $q = KLM$	
2-1	The reconstruction due to HPCA from 1 to $q$ : $\sum_{k=1}^K \sum_{l=1}^L \sum_{m=1}^M \xi_{i,klm} V_k(r) \phi_l(\omega) \psi_m(s) \approx \hat{Y}_i(r, \omega, s)$ (eq. 3)
2-2	Merging on two dimensions: $\frac{1}{R} \sum_{r=1}^R \int_{-\infty}^{\infty} \hat{Y}_i(r, \omega, s)_{HPCA\_KLM} ds = \hat{W}_i(\omega)_{KLM}$ . (eq. 4)
2-3	Fit Bayesian SOF based on the $\hat{W}_i(s)_{HPCA\_q}$ $\begin{cases} U_i \sim EF(v_i, a); \\ v_i = \sum_{d \geq 1} \zeta_{id} \beta d; \\ \{W_i\} : \hat{W}_i(\omega)_{HPCA\_q} = \sum_{d \geq 1} \zeta_{id} \eta d(\omega) + \epsilon_i(\omega) \text{ (eq.8)} \\ \zeta_{id} \sim N(0, \lambda_k) \\ \epsilon_i(s) \sim N(0, \sigma_\epsilon^2) \end{cases}$
2-4	Predict $\hat{U}_{i,HPCA\_q}$ within (2-3)
2-5	Calculate MSPE Mean Squared Prediction Error $MSPE(q) = \frac{1}{n} \sum_{i=1}^n (U_{i,HPCA\_q} - \hat{U}_{i,HPCA\_q})^2$
3- Find the minimum $MSPE(q_{min}) = \min\{q = 0 \text{ or } MSPE(q), 1 \leq q \leq KLM\}$	
4- The best number of HPCA is $q_{min}$	

Table 2. Algorithm 2 - Bayesian Regression

second steps with  $\hat{W}_i(\omega)_{HPCA\_0} + A_i(\omega) = D_i(\omega)_{HPCA\_0}$  and we fit the fof regression with the following equation in equation 10 (like equation 6):

$$D_i(\omega)_{HPCA\_0} = m(\omega) + \sum_{j=1}^p \int_0^1 \chi_j(g) \beta_j(g, \omega) dg + \epsilon(\omega) \tag{10}$$

In each model, we split the data to the train (70%) and test (30%) and we fit the model on the training dataset. We calculate the following summary statistics for coefficients ( equations 11 , 12 and 13 ):

$$\begin{cases} MSE\beta_1(\omega, t) = \frac{1}{n \times \Omega \times T} \sum_{i=1}^n \sum_{\omega=1}^{\Omega} \sum_{t=1}^T (\beta_1(\omega, t) - \hat{\beta}_1(\omega, t))^2 \\ \dots \\ MSE\beta_{10}(\omega, t) = \frac{1}{n \times \Omega \times T} \sum_{i=1}^n \sum_{\omega=1}^{\Omega} \sum_{t=1}^T (\beta_{10}(\omega, t) - \hat{\beta}_{10}(\omega, t))^2 \end{cases} \tag{11}$$

For the prediction accuracy of model in the train and test, respectively:

$$\begin{cases} MSPE_{Pred(train)} = \frac{1}{n_{train}} \sum_{i=1}^{n_{train}} \sum_{j=1}^J (D_{i,train}(\omega_j)_{HPCA\_0} - \hat{D}_{i,train}(\omega_j)_{HPCA\_0})^2 \\ MSPE_{Pred(test)} = \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} \sum_{j=1}^J (D_{i,test}(\omega_j)_{HPCA\_0} - \hat{D}_{i,test}(\omega_j)_{HPCA\_0})^2 \end{cases} \tag{12}$$

$$\begin{cases} COR_{Pred(train)} = COR(D_{i,train}(\omega)_{HPCA\_0}, \hat{D}_{i,train}(\omega)_{HPCA\_0}) \\ COR_{Pred(test)} = COR(D_{i,test}(\omega)_{HPCA\_0}, \hat{D}_{i,test}(\omega)_{HPCA\_0}) \end{cases} \tag{13}$$

Moreover, we compare the elapsed, system, and user time for computations. We repeat the iteration 100 times. The median (the first quartile, the third quartile) in all simulations were reported.

$s$	Number of Subjects							
	Complete, (n=20)		Sparsity, (n=20)		Complete, (n=140)		Sparsity, (n=140)	
	Complete	50% Sparsity	Complete	50% Sparsity	Complete	50% Sparsity	Complete	50% Sparsity
$\beta$	Complete	50% Sparsity	Complete	50% Sparsity	Complete	50% Sparsity	Complete	50% Sparsity
$MSE\beta_1(\omega, t)$	20.61 (9.94,33.4)	1.96 (0.03,22.18)	17.61 (8.59,34.25)	5.02 (0.08,19.23)	29.39 (18.05,46.26)	0.0132 (0.22,4816)	33.29 (19.93,48.03)	8.83 (0.003,32.637)
$MSE\beta_2(\omega, t)$	1.5 (1.03,2.89)	0.5 (0.1,1.08)	1.47 (0.88,2.42)	0.58 (0.1,1.32)	0.18 (0.14,0.23)	0.0895 (0.0032,0.1596)	0.19 (0.15,0.22)	0.098 (0.001,0.186)
$MSE\beta_3(\omega, t)$	2.75 (1.73,4.66)	0.99 (0.08,2.82)	2.86 (1.93,4.92)	1.14 (0.06,2.4)	1.46 (1.04,2.32)	0.018 (0.0001,1.1931)	1.59 (1.02,2.42)	0.042 (0.003,1.148)
$MSE\beta_4(\omega, t)$	2.96 (1.86,5.29)	1.32 (0.13,2.85)	3.55 (1.95,5.54)	0.76 (0.04,2.59)	2.4 (1.71,3.7)	0.9192 (0.006,2.2836)	2.76 (1.83,4.06)	0.047 (0.003,2.845)
$MSE\beta_5(\omega, t)$	4.03 (2.83,6.35)	1.23 (0.07,3.88)	3.3 (2.11,5.14)	1.15 (0.1,3.47)	2.94 (2.13,4.72)	0.8649 (0.0005,3.3766)	3.15 (2.01,4.3)	0.654 (0.009,3.227)
$MSE\beta_6(\omega, t)$	3.93 (2.43,6.27)	0.57 (0.05,4.31)	3.7 (2.17,5.03)	1.17 (0.11,3.18)	3.33 (1.84,5.19)	1.1584 (0.0052,3.4304)	3.29 (2.2,4.68)	1.436 (0.004,4.286)
$MSE\beta_7(\omega, t)$	3.73 (2.36,5.79)	0.45 (0.07,3.06)	3.93 (2.29,5.84)	1.09 (0.1,3.92)	2.99 (1.82,4.48)	1.0422 (0.0028,3.8262)	2.74 (1.87,5.23)	1.809 (0.012,3.895)
$MSE\beta_8(\omega, t)$	3.56 (2.04,5.43)	0.91 (0.08,3.9)	3.91 (2.61,6.84)	0.56 (0.03,3.47)	3.41 (1.94,5.47)	0.0603 (0.0025,3.5484)	4.02 (2.5,93)	0.037 (0.001,2.989)
$MSE\beta_9(\omega, t)$	4.2 (2.76,6.67)	0.4 (0.03,3.55)	4.02 (2.44,5.95)	1.07 (0.07,3.53)	3.41 (2.29,5.51)	1.7362 (0.0086,3.9769)	4.1 (2.14,6.39)	0.051 (0.003,3.411)
$MSE\beta_{10}(\omega, t)$	3.4 (2.1,6.17)	2.05 (0.1,5.22)	3.72 (2.12,6.06)	1.13 (0.05,3.74)	4.1 (2.52,6.52)	0.0398 (0.001,3.2849)	3.75 (2.32,6.28)	0.671 (0.004,4.251)
$MSPE_{Pred(train)}$	0.09 (0.07,0.12)	0.09 (0.08,0.11)	0.09 (0.07,0.12)	0.09 (0.07,0.1)	0.09 (0.09,0.1)	0.0975 (0.0933,0.1012)	0.1 (0.09,0.1)	0.098 (0.096,0.102)
$MSPE_{Pred(test)}$	0.37 (0.25,0.5)	0.19 (0.14,0.32)	0.33 (0.25,0.54)	0.21 (0.15,0.33)	0.12 (0.11,0.13)	0.1101 (0.1045,0.1182)	0.12 (0.11,0.13)	0.114 (0.106,0.12)
$COR_{Pred(train)}$	0.98 (0.98,0.99)	0.97 (0.96,0.98)	0.98 (0.98,0.99)	0.97 (0.96,0.98)	0.98 (0.98,0.99)	0.9613 (0.9559,0.9784)	0.98 (0.98,0.99)	0.97 (0.957,0.983)
$COR_{Pred(test)}$	0.94 (0.9,0.96)	0.94 (0.92,0.96)	0.94 (0.9,0.97)	0.94 (0.91,0.96)	0.98 (0.97,0.98)	0.9564 (0.9511,0.971)	0.98 (0.97,0.98)	0.964 (0.952,0.979)
$Time_{user}$	48.33 (35.55,70.02)	101.98 (55.02,133.85)	51.64 (37.22,74.64)	102.54 (53.44,136.38)	95.5 (79.82,120.1)	322.955 (105.965,381.245)	97.94 (79.59,123.5)	169.76 (88.53,385.28)
$Time_{system}$	0.07 (0.05,0.09)	0.09 (0.07,0.13)	0.06 (0.04,0.09)	0.07 (0.05,0.11)	0.24 (0.18,0.3)	0.415 (0.32,0.51)	0.11 (0.07,0.14)	0.11 (0.08,0.15)
$Time_{elapsed}$	48.4 (35.59,70.1)	102.11 (55.1,133.98)	51.66 (37.26,74.71)	102.69 (53.56,136.51)	95.94 (80.28,120.36)	323.425 (106.585,381.8525)	98.27 (79.93,123.84)	170.01 (88.78,385.73)

Table 3. Percentiles 50% (25%, 75%) of the simulation scenario 1 of  $MSE$  of coefficients,  $MSPE$  and correlation of prediction and running time group by the number of subject ( $n$ ), the longitudinal status  $s$  and the coefficients status  $\beta$ .

We briefly explain the result: in Table 3, the  $MSE\beta_1(\omega, t)$ , ...,  $MSE\beta_{10}(\omega, t)$  is greater for the complete rather than 50 % sparsity of  $\beta$ . The  $MSPE_{Pred(train)}$  are not very different in all situations with each other, but  $MSPE_{Pred(test)}$  are different. The first difference is  $MSPE_{Pred(test)}$  for complete  $\beta$  are greater than the 50 % sparsity of  $\beta$ . The second difference is by increasing the sample size, their values are getting smaller than the big sample size. But its value for the sparsity for  $s$  is not different from others. The  $Corr_{Pred(train)}$  are negligibly bigger than  $Corr_{Pred(test)}$  and they are not very different among scenarios. The elapsed time for 50 % sparsity of  $\beta$  is greater than complete  $\beta$  and they increase significantly with the number sample size. This behavior is not observed for system time. Therefore, the sparsity of  $\beta$  will affect the model performance, and the sparsity of  $s$  don't have any significant effect on them.

#### 4.2. Simulation Scenario 2

This simulation scenario is very similar to simulation scenario 1, but in this one, we study the number of HPCA. Furthermore, we first generate  $Y_i(r, \omega, s), i = 1, \dots, 20$  as mentioned in the [22], then we merge them with equation 4 in three ways: 1) Without HPCA decomposition, as  $\hat{W}_i(\omega)_{HPCA_0}$ , 2) by only the first HPCA decomposition, as  $\hat{W}_i(\omega)_{HPCA_1}$ , 3) by All HPCA decomposing, as  $\hat{W}_i(\omega)_{HPCA_{All}}$  (All HPCA has 90% FVE on each dimension). Furthermore, we generate fof regression data as mentioned at [19] with 10 covariates functions,  $\theta(g, \omega)$ , and two statuses for coefficients 1) all 10 coefficients,  $\theta(g, \omega)$ ,

have the value greater than 0 and 2) only 50% of  $\theta(g, \omega)$  have the non-zero value. Therefore, we have  $A_i(\omega) = m(\omega) + \sum_{j=1}^{10} \int_0^1 \chi_j(g)\theta_j(g, \omega), dg + \epsilon(\omega)$ . We define in equation 14:

$$\left\{ \begin{array}{l} \hat{W}_i(\omega)_{HPCA\_0} + A_i(\omega) = D_i(\omega)_{HPCA\_0} \\ D_i(\omega)_{HPCA\_0} = m(\omega) + \sum_{j=1}^p \int_0^1 \chi_j \beta_j(g, \omega), dg + \epsilon(\omega) \\ \hat{W}_i(\omega)_{HPCA\_1} + A_i(\omega) = D_i(\omega)_{HPCA\_1} \\ D_i(\omega)_{HPCA\_1} = m(\omega) + \sum_{j=1}^p \int_0^1 \chi_j \beta_j(g, \omega), dg + \epsilon(\omega) \\ \hat{W}_i(\omega)_{HPCA\_ALL} + A_i(\omega) = D_i(\omega)_{HPCA\_ALL} \\ D_i(\omega)_{HPCA\_ALL} = m(\omega) + \sum_{j=1}^p \int_0^1 \chi_j \beta_j(g, \omega), dg + \epsilon(\omega) \end{array} \right. \quad (14)$$

In each model,  $D_i(\omega)_{HPCA\_0}$ ,  $D_i(\omega)_{HPCA\_1}$  and  $D_i(\omega)_{HPCA\_ALL}$ , we split the data to the train (70%) and test (30%), and we fit the models on the training dataset. We calculate the summary statistics as mentioned in scenario 2. We repeat the iteration 100 times. The median (the first quartile, the third quartile) in all simulations were reported.

We briefly explain the result: in (Table 4), the most of  $MSE\beta_1(\omega, t)$ , ...,  $MSE\beta_{10}(\omega, t)$  are greater for the complete rather than 50 % sparsity of  $\beta$ .  $MSPE_{Pred(train)}$  are 0.1001 (0.084, 0.1294) for without HPCA, 0.0028 (0.0001,0.0176) with the first HPCA and 0.0174 (0.0024,0.0409) for all HPCA in the complete  $\beta$  settings. It stated that the first HPCA has the lowest MSPE among others (2% and 16% of without HPCA and with all HPCA, respectively). The same pattern exists for  $MSPE_{Pred(test)}$ .

$COR_{Pred(train)}$  and  $COR_{Pred(test)}$  are high (more than 90%), and the differences are negligible. The elapsed time is the lowest for the first HPCA model among others.

### 4.3. Simulation Scenario 3

The Bayesian Scalar-on-Function: in this scenario, we first generate the  $Y_i(r, \omega, s)$  with  $i, i = 1, , 15$  as mentioned in the [22], then we follow equations 1 to 5 to decompose, reconstruct and merge them, in three ways: 1) Without HPCA decomposition, as  $\hat{W}_i(s)_{HPCA\_0}$ , 2) With only the first HPCA decomposition ,as  $\hat{W}_i(s)_{HPCA\_1}$ , 3) With All HPCA decomposition, as  $\hat{W}_i(s)_{HPCA\_All}$ .

We generate the beta coefficient with the following function,  $\beta(\omega) = \cos(\sqrt{\frac{\pi \times \omega}{3}})$  and error term  $a_i$  Normal( $\mu = 0, \sigma = 0.4$ ). We calculate the real scalar response as:  $U_i = \int \beta(\omega) \times \hat{W}_i(\omega)_{HPCA\_0} \times 0.01 + a_i$ . Then, we run the Bayesian Scalar-on-Function in three settings: the response scalar  $U_i$  and the functional covariates 1)  $\hat{W}_i(\omega)_{HPCA\_0}$ , 2)  $\hat{W}_i(\omega)_{HPCA\_1}$  and 3)

$\hat{W}_i(\omega)_{HPCA\_All}$ . We also compare ten models in each of them from  $v_i = \zeta_{i1}\beta(1)$  to  $v_i = \sum_{d=1}^{10} \zeta_{id}\beta(d)$ . In each model, we calculate the following summary statistics [27], Deviance, Deviance Information Criteria (DIC), log pseudo-marginal likelihood (LPML) which contains Conditional predictive ordinate (CPO), and pseudo-marginal likelihood is  $\hat{f}(y|M)$ , Mean Square Error for Mean ( $MSE_{mean}$ ) and correlation ( $COR_{mean}$ ) ( equation 15 ):

$$\left\{ \begin{array}{l} D(\theta) \equiv -2\log[L(\theta)] + C(\#Parameters) \\ DIC \equiv 2E[D(\theta)|y] - D[\hat{\theta}] \\ CPO_i = f_i(y_i|y_{-i}, M), ModelM \\ (y |M) = \prod_{i=1}^n f_i(y_i|y_{-i}, M) \\ LPML = \sum_{i=1}^n \log(CPO_i) \\ MSE_{mean} = \frac{1}{n} \sum_{i=1}^n (U_i - \hat{U}_{i,mean})^2 \\ COR_{mean} = COR(U_i, \hat{U}_{i,mean}) \end{array} \right. \quad (15)$$

We also compare the elapsed, system, and user time for computations. The MCMC algorithm has 5,000 samples with the first 1,000 are burn-out and the thin is 1. We repeat the simulations 100 times and report



HPCA	NO HPCA		First HPCA		All HPCA	
	Bet	Complete	50% Sparsity	Complete	50% Sparsity	Complete
MSE Beta 1	14.5112 (8.3463,27.4955)	19.1862 (10.6082,27.8363)	14.6201 (7.2417,25.9006)	16.317 (10.498,27.322)	15.925 (8.7129,30.6679)	19.0602 (11.577,28.7054)
MSE Beta 2	1.4757 (0.8491,2.6039)	0.6611 (0.0946,1.3561)	3.0187 (1.6888,4.1153)	0.94 (0.042,2.369)	2.0448 (1.0246,2.8579)	0.7309 (0.0696,1.5715)
MSE Beta 3	2.6643 (1.6905,4.3293)	0.783 (0.0967,2.6223)	3.4721 (2.1247,5.1566)	0.873 (0.026,3.042)	2.8849 (1.6589,4.6093)	0.8535 (0.1357,2.6239)
MSE Beta 4	3.0949 (2.0827,6.0484)	0.4074 (0.0523,2.5258)	3.2804 (2.2413,5.7897)	0.134 (0.003,2.313)	3.0776 (2.0219,5.9766)	0.4004 (0.0257,2.015)
MSE Beta 5	3.9203 (2.3426,5.7395)	1.0133 (0.1678,3.2986)	4.0012 (2.4556,6.741)	0.439 (0.094,3.171)	3.9086 (2.6642,6.2414)	0.9969 (0.129,3.2651)
MSE Beta 6	4.221 (2.3879,6.6958)	1.4369 (0.122,4.3104)	3.8636 (2.4522,6.0266)	1.366 (0.085,4.27)	3.7442 (2.4925,6.9392)	1.592 (0.1351,4.2872)
MSE Beta 7	3.3584 (2.2642,5.8811)	0.426 (0.0373,2.8341)	3.3409 (2.1588,5.1684)	0.385 (0.007,3.052)	3.3756 (2.2225,4.9557)	0.5478 (0.0369,2.944)
MSE Beta 8	3.8073 (2.389,5.5397)	0.7336 (0.083,3.8915)	3.6125 (2.5294,5.5124)	0.423 (0.018,3.47)	3.7223 (2.4792,5.6815)	0.962 (0.0842,3.6444)
MSE Beta 9	3.48 (2.471,5.5648)	0.4391 (0.063,2.4887)	3.4562 (2.3211,5.3077)	0.236 (0.007,2.724)	3.5074 (2.3809,5.748)	0.4747 (0.1196,2.7758)
MSE Beta 10	3.7703 (2.2949,5.2103)	1.1676 (0.1598,3.8849)	3.5595 (2.3031,5.6235)	1.239 (0.034,3.666)	3.7815 (2.3975,5.914)	1.3054 (0.1028,4.0518)
MSPE Train	0.1001 (0.084,0.1294)	0.0953 (0.0753,0.1118)	0.0028 (0.0001,0.0176)	0.004 (0.001,0.009)	0.0174 (0.0024,0.0409)	0.0109 (0.0037,0.029)
MSPE Test	0.3207 (0.2452,0.5342)	0.2967 (0.2107,0.4604)	0.1118 (0.0439,0.2274)	0.087 (0.044,0.27)	0.2484 (0.1498,0.4314)	0.2529 (0.1223,0.4378)
COR Train	0.9801 (0.9691,0.9873)	0.9829 (0.9755,0.9884)	0.9994 (0.9962,1)	0.999 (0.998,1)	0.9971 (0.9922,0.9996)	0.9979 (0.9942,0.9994)
COR Test	0.9334 (0.9005,0.9647)	0.942 (0.915,0.9684)	0.9801 (0.9502,0.9928)	0.982 (0.958,0.993)	0.962 (0.9265,0.9765)	0.955 (0.9277,0.9792)
Time User	58.095 (42.135,83.1375)	55.82 (39.2425,82.46)	15.44 (14.305,17.685)	14.89 (12.83,17.863)	34.72 (30.61,41.42)	33.67 (28.37,40.185)
Time System	0.05 (0.03,0.08)	0.05 (0.03,0.06)	0.05 (0.03,0.0825)	0.04 (0.02,0.11)	0.08 (0.05,0.11)	0.05 (0.03,0.08)
Time Elapsed	58.14 (42.24,83.2425)	55.87 (39.2775,82.515)	15.61 (14.4425,17.815)	14.935 (12.868,18.265)	34.89 (30.765,41.7125)	33.75 (28.54,40.27)

Table 4. Percentiles 50% (25%, 75%) of the simulation scenario 2 of MSE of coefficients, MSPE and correlation of prediction and running time group by HPCA status and the coefficients status  $\beta$ . Number of iteration 100 times, \* has 96 iteration and \*\* has 99 iterations.

the median (first quantile – third quantile) of each parameter in the tables. (Tables A, B and C are in the Appendix)

We explain briefly the result: in table A, the  $MSPE_{mean}$  is decreased as FPCA grows and reach to 0 with 6 FPCAs with DIC -2884 (-2920.89,-2851.35). The  $MSPE_{mean}$  for 6 FPCA in table B and C are 0.000006 (0.000005,0.000007 ) and 0.000006 (0.000005,0.000007 ) with the DIC -4494.56 (-4523.2,-4463.96 ) and -4439.42 (-4478.33,-4411.17 ), respectively. The  $MSPE_{mean}$ , Elapsed time and user time are not very different in these tables, but the DIC values are the lowest and LPML are the greatest for the first HPCA (table-B in the Appendix). Therefore, we conclude that the model with the first HPCA has the best model performance (table-B in the Appendix) while their prediction accuracy are the same as the models with no HPCA (table-A in the Appendix) and with All HPCAs (table-C in the Appendix).

### 5. Application to the EEG-fMRI dataset

In this application, we use EEG-fMRI dataset, which was analyzed with HPCA and signal compression FOF regression separately, previously at [19, 28]. The novelty of this new analysis is included: 1) We

analyze the Evoked Related Potential (ERP) with the time instead of frequency domain as response function and 2) we also study the effect of HPCA on the prediction accuracy of FOF regression. The dataset is related to an experiment to study the brain's internal attention with 17 participants (6 females, mean age is 27.7 years) [29]. There are two tasks for each experiment, including auditory and visual stimuli. There are two stimuli in each task, standard and target. The auditory task includes standard stimuli (390 Hz tone) and target stimuli (laser gun sound). The visual task includes standard stimuli (small green circle on isoluminant grey background, 1.5-degree visual angle) and target stimuli (large red circle on isoluminant grey background, 3.45-degree visual angle). The dataset was downloaded from openneuro.org. EEG was captured due to 64 channels, and the pre-processing steps are artifact removal, baseline correction, and epoching (-200 to 800 milliseconds of each stimulus) that which were done with EEGLAB [30]. The fMRI dataset was captured with 3 Tesla Philips scanner and the dataset was pre-processed through slice timing correction (SCT), normalization, coregistering, smoothing, and region of interest (ROI) extraction with SPM 12 toolbox (<http://www.fil.ion.ucl.ac.uk/spm/>). ERP waves are response functions and ROI time series for different Brodmann Areas in the brain are functional covariates [19]. We conduct two analyses: the first one chooses the best number of HPCA based on the MSPE and Correlation, and the second is FOF regression for EEG-fMRI by brain region.

We first decompose ERP curves with HPCA from first to the total number of HPCAs for standard visual tasks. KML is 14 which captures 90% of FVE in each dimension. Then, we fit the model with equation 5 on the training samples and compare it to the test. We do this procedure 100 times with different training/testing samples. Finally, we compare  $MSPE_{Pred(train)}$ ,  $MSPE_{Pred(test)}$ ,  $COR_{Pred(train)}$ ,  $COR_{Pred(test)}$  and  $Time_{elapsed}$  for different number of HPCAs. We see the first HPCA has the lowest MSPE in both trains/tests. By increasing the number of HPCA, MSPE increases. The model without HPCA has the largest MSPE. The same pattern is seen in the correlation and the elapsed time. (Figure-1)

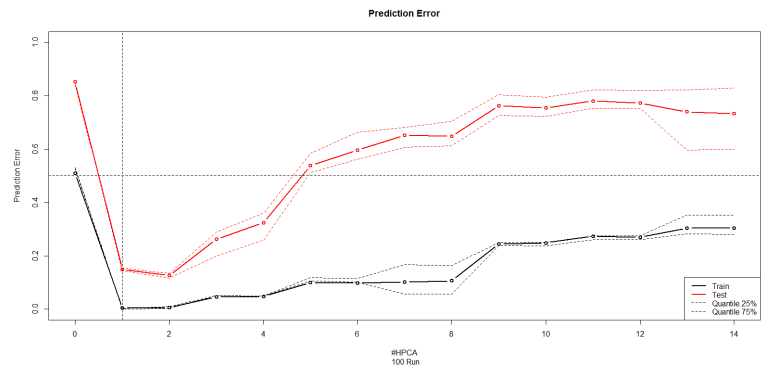
Therefore, we use the first HPCA and analyze the whole dataset. In this regard, we analyze the whole brain and different regions and compare MSPE and Correlation in the train/test for the auditory and visual tasks.  $MSPE$  for the auditory in all brain regions for standard and target stimuli are 0.00007 and 0.01337, and for visually are 0.00235 and 0.41549. The result is also provided by the different regions of the brain. (Table 5). The further analysis includes 10 iterations for each task and each stimuli group by brain regions and train/test were done in supplementary 1. The result of them showed the same pattern exists in most of them and the first HPCA has the lowest MSPE and highest correlation among others. This analysis doesn't have HPCA settings and starts from the first to the KLM HPCAs.

All computations are done with R x64 4.0.2 and RStudio version Version 1.2.5042 in the server with 24 GB RAM, 64-bit Operating System, Intel® Xeon® CPU, X5670 @ 2.93 GHz 2.93 GHz (2 processors) from Turin Cloud Services ([turin.ipm.ir](http://turin.ipm.ir)).

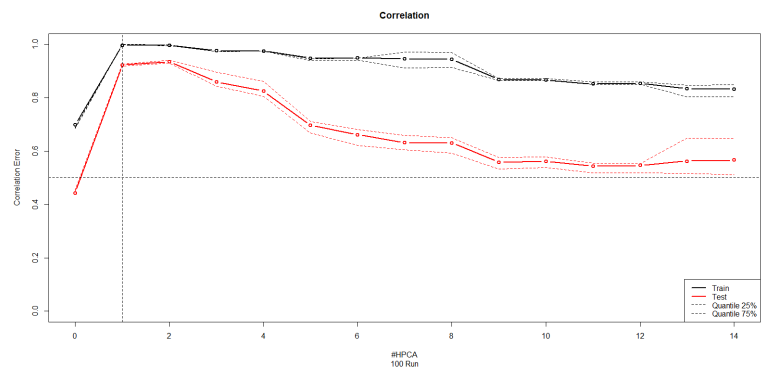
## 6. Conclusion

We propose a method to use HPCA in the functional regressions and choose the best number of HPCA. In this regard, we first decompose the hybrid data with HPCA, then we reconstruct the observation due to HPCA eigencomponents. This method helps us reconstruct the data with the most important eigencomponents of each dimension and eliminate noise and other negligible variations. Note that the various functional regressions due to the type of estimations, responses, and covariates are studied, and we use two of them which were previously published with the EEG-fMRI dataset [19].

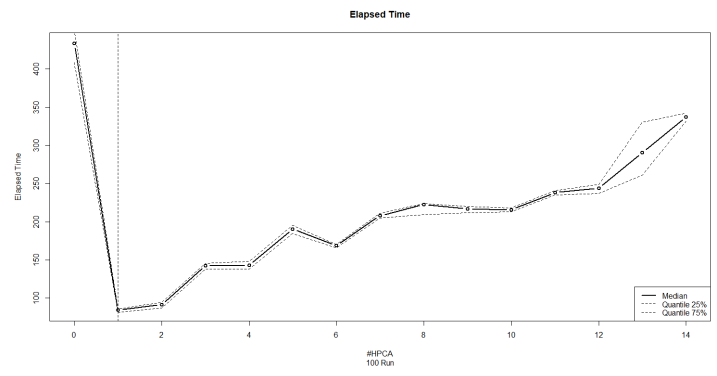
This functional regression works with a one-dimensional function; for example, it is common to use ERP over time or frequency in an EEG dataset. We propose a method to convert the hybrid data to one-dimensional data by pooling and averaging the other two dimensions of reconstructed data as mentioned in the functional regression models. Note that instead of using observed hybrid data, we use the reconstructed data from HPCA eigencomponents. Furthermore, we find the best number of HPCA due to functional regression MSPE.



(a)



(b)



(c)

Figure 1. (a)-(b): The 50% (25%-75%) quantile for train (black) and test (red) of MSPE and correlation by the number of HPCA in training/test samples in 100 iterations. (c): The 50% (25%-75%) quantile of elapsed time by the number of HPCA in training/test samples in 100 iterations.

In the supplementary file, the simulations and real-world application with the EEG-fMRI show that only the first few eigenvectors in the most different scenarios achieved the best performance of the regression in the training and testing dataset. We can count the following reasons to explain this situation; first, the simulations are designed to show EEG structure of the data which has three dimensions, and most of the variations from other two dimensions are captured with the first few HPCA; second, the higher

Tasks	Regions	Stimuli	MSPE		Correlation	
			Train	Test	Train	Test
Auditory	All	Standard	0.00007	0.07061	0.99997	0.96444
		Target	0.01337	0.16030	0.99329	0.91817
	Frontal	Standard	0.00008	0.12564	0.99996	0.93640
		Target	0.00174	0.19295	0.99913	0.90111
	Central	Standard	0.00004	0.04605	0.99998	0.97686
		Target	0.00177	0.23528	0.99911	0.87700
	Occipital	Standard	0.00014	0.00123	0.99993	0.99939
		Target	0.00024	0.00178	0.99988	0.99911
	Parietal	Standard	0.00018	0.03787	0.99991	0.98092
		Target	0.02034	0.15549	0.98977	0.92011
	Left Temporal	Standard	0.00116	0.02427	0.99942	0.98784
		Target	0.00005	0.00062	0.99998	0.99969
	Right Temporal	Standard	0.00040	0.06260	0.99980	0.96833
		Target	0.02455	0.12669	0.98763	0.93528
Visual	All	Standard	0.00235	0.14560	0.99882	0.92580
		Target	0.41549	0.52133	0.79498	0.72420
	Frontal	Standard	0.01259	0.08244	0.99368	0.95795
		Target	0.06498	0.17451	0.96734	0.91379
	Central	Standard	0.00009	0.09191	0.99995	0.95308
		Target	0.53342	0.64774	0.71244	0.61246
	Occipital	Standard	0.00015	0.00149	0.99993	0.99926
		Target	0.00041	0.00186	0.99980	0.99907
	Parietal	Standard	0.00008	0.06213	0.99996	0.96845
		Target	0.32912	0.39895	0.83103	0.78537
	Left Temporal	Standard	0.00460	0.04253	0.99769	0.97848
		Target	0.02002	0.07749	0.98993	0.96081
	Right Temporal	Standard	0.00132	0.05436	0.99934	0.97249
		Target	0.11777	0.23887	0.93919	0.88745

Table 5. The MSPE and correlation of FOF result from EEG-fMRI Dataset group by tasks, stimuli and the region of the brain

number of HPCA was used in the reconstructed data, it gets closer to the observed data; therefore, their complexity increases which this matter affects the prediction accuracy in the functional regression.

The proposed method's future direction is 1) using other indices than MSPE of regression to choose the number of HPCA like Log pseudo-marginal likelihood (LPML) and Deviance information criteria (DIC) in Bayesian setting [27, 2] using different types of functional regressions, and applying the methodology on their applications other than EEG structure and 3) using in the offline and online estimates in Brain-Computer Interface (BCI) applications and comparing with the prediction accuracy in feature selections

extractions methods like common spatial pattern (CSP) and supervised models like Variational Bayesian logistic regression [31] .

Finally, we conclude that our methodology improves the prediction accuracy of the experiments with EEG datasets. We recommend that instead of using functional PCA on desired dimension; for example, ERP over time or frequency, reconstruct the data with HPCA and average it on other two dimensions for functional regression models.

#### REFERENCES

1. A. Beck, and M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.
2. Johnson, R.A. and D.W. Wichern, *Applied multivariate statistical analysis.*, Prentice hall Upper Saddle River, NJ Vol. 5. 2002.
3. Mandel, J., Use of the singular value decomposition in regression analysis., *The American Statistician*, 36(1): p. 15-24, 1982.
4. J.O. Ramsay and B.W. Silverman, *Functional Data Analysis, Second Edition*, Springer, 2005.
5. F. Ferraty and P. Vieu , *Nonparametric functional data analysis: theory and practice.*, Springer Science and Business Media, 2006.
6. G.M. James, T.J. Hastie and C.A. Sugar, Principal component models for sparse functional data., *Biometrika*, 87(3): p. 587-6025, 2000.
7. A.-M Staicu, C.M. Crainiceanu and R.J. Carroll , Fast methods for spatially correlated multilevel functional data, *Biostatistics*,11(2): p. 177-194, 2010.
8. M. Febrero-Bande, P. Galeano and W. González-Manteiga , Functional principal component regression and functional partial least-squares regression: An overview and a comparative study., *International Statistical Review.* 85(1): p. 61-83, 2017.
9. H. Hullahit et al., Robust Function-on-Function Regression., *Technometrics*, p. 1-14,2020.
10. K. Chen and H.-G. Müller , Modeling repeated functional observations., *Journal of the American Statistical Association*, 107(500): p. 1599-1609,2012.
11. S. Greven and et al. , Longitudinal functional principal component analysis, in *Recent Advances in Functional Data Analysis and Related Topics.*, Springer. p. 149-154,2011.
12. C.-Z Di and et al. , Multilevel functional principal component analysis., *The annals of applied statistics*,3(1): p. 458:2009.
13. K. Hasenstab and et al. , A multi-dimensional functional principal components analysis of EEG data, *Biometrics*. 73(3): p. 999-1009, 2017.
14. T. Kuenzer, S. Hörmann, and P. Kokoszka. , Principal component analysis of spatially indexed functions, *Journal of the American Statistical Association*, p. 1-13,2020.
15. C. Liu, S. Ray, and G. Hooker , Functional principal component analysis of spatially correlated data, *Statistics and Computing.* 27(6): p. 1639-1654,2017.
16. L. Zhou, et al., Reduced rank mixed effects models for spatially correlated hierarchical functional data., *Journal of the American Statistical Association.* 105(489): p. 390-400,2010.
17. P. Reimherr and P. Kokoszka, *Introduction to Functional Data Analysis*, CRC Press, Taylor and Francis Group,2017.
18. R. Luo and X. Qi, Function-on-function linear regression by signal compression, *Journal of the American Statistical Association.* 112(518): p. 690-705, 2017.
19. X. Qi and R. Luo, Function-on-function regression with thousands of predictive curves., *Journal of Multivariate Analysis*, 163: p. 51-66, 2018.
20. C.M. Crainiceanu and A.J. Goldsmith , Bayesian functional data analysis using WinBUGS, *Journal of statistical software*, 32(11), 2010.
21. D.R. Kowal and D.C. Bourgeois , Bayesian function-on-scalars regression for high-dimensional data, *Journal of Computational and Graphical Statistics*, p. 1-10, 2020.
22. A. Scheffler and et al. , Hybrid principal components analysis for region-referenced longitudinal functional EEG data, *Biostatistics*, 21(1): p. 139-157, 2020.
23. A.W. Scheffler and et al. , Covariate-adjusted region-referenced generalized functional linear model for EEG data, *Statistics in Medicine.* 38(30): p. 5587-5602,2019.
24. G. Aneiros-Pérez and P. Vieu , Semi-functional partial linear regression, *Statistics and Probability Letters*, 76(11): p. 1102-1110, 2006.
25. D. Liebl and F. Walders , Parameter regimes in partial functional panel regression., *Econometrics and Statistics*, 11: p. 105-115, 2019.
26. M. Febrero-Bande, W. González-Manteiga, and M.O. de la Fuente, Variable selection in functional additive regression models., *Computational Statistics.* 34(2): p. 469-487,2019.
27. R. Christensen et al., *Bayesian ideas and data analysis: an introduction for scientists and statisticians*, CRC Press, 2011.
28. M. Fayaz, A. Abadi, and S. Khodakarim , The Comparison between Visually and Auditory Oddball Tasks in the EEG Experiment with Healthy Subjects., *Frontiers in Biomedical Technologies.* 7(4): p. 249-258,2020.

29. J.M. Walz et al. , Simultaneous EEG–fMRI reveals a temporal cascade of task-related and default-mode activations during a simple target detection task., *Neuroimage*, 102: p. 229-239,2014.
30. A. Delorme and S. Makeig , EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis., *Journal of neuroscience methods*. 134(1): p. 9-21,2004.
31. C.A. Kothe and S. Makeig, BCILAB: a platform for brain–computer interface development. *Journal of neural engineering*, *Journal of neural engineering*, 10(5): p. 056014,2013.