The effect of Hybrid Principal Components Analysis on the Signal Compression Functional Regression: With EEG-fMRI Application

Mohammad Fayaz\textsuperscript{a}, Alireza Abadi\textsuperscript{b,c,*} and Soheila Khodakarim \textsuperscript{d}

\textsuperscript{a} PhD Student of Biostatistics, Department of Biostatistics, School of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

\textsuperscript{b} Professor of Biostatistics, Department of Community Medicine, Faculty of medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

\textsuperscript{c} Social Determinants of Health Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

\textsuperscript{d} Associate Professor, School of Allied Medical Sciences, School of Public Health and Safety, Shahid Beheshti University of Medical Sciences, Tehran, Iran

* Corresponding Author: Dr. Alireza Abadi (alirezaabadi@gmail.com)

Professor of Biostatistics, Department of Community Medicine, Faculty of medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran. Social Determinants of Health Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

Koodakyar Ave., Daneshju Blvd., Velenjak, Tehran, IRAN.
1- Introduction

The principal component regression (PCR) was used in the regression settings to tackle the collinearity problem between covariates [1-3]. With an introduction to the functional data [4], the functional extension of the Principal Component Analysis (FPCA) captured the underlying curves of the data and reduce their dimensionality [5]. We count different methods for estimating the number of FPCAs such as fraction of variance explained (FVE), comparing the log-likelihood of different models with different numbers of FPCAs [6], Likelihood Ratio Test (LRT) [7], the mean square error (MSE), mean square prediction error (MSPE) [8] and Bayesian information criteria (BIC) [9]. The FPCA methods extend in different ways to deal with for example longitudinal dimension [10, 11], multilevel structure [12], multidimensional [13], and spatially dimension [7, 14-16].

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 the functional data.[4, 5, 17]. Some of them are adapted for predictions such as signal compression estimations [18, 19] for modeling the EEG and many regions of interest (ROI) of the Functional magnetic resonance imaging (fMRI) and Bayesian estimations [20, 21] with applications to the δ-power of the electroencephalography (EEG) data and intraday physical activity with many covariates in the sleep study.

The hybrid PCA (HPCA) was introduced for situations that exist 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 that they are frequency or time domain ERPs, the repeating stimuli, and the regions of the brain, respectively. The first two of them have functional characteristics and the last one is scalar. The HPCA extracts the subject-specific scores and eigencomponents of hybrid data [22]. Among the regression models for the hybrid data, we can count covariate-adjusted HPCA [23], the Semi-functional partial linear regression [24, 25], FOF regression with signal compression [18], and functional additive regression models [26].

In this research, we study the effects of HPCA decomposition of hybrid data on the prediction accuracy of the FOF regression with signal compression. We choose the number of HPCA based on the MSPE. Finally, we applied our methodology to the EEG-fMRI dataset. The paper has the following parts, 1-introduction contains the problem definitions with some background, 2- The functional regression with HPCA contains HPCA decomposition and FOF with multiple functional covariates, 3- The number of
HPCA contains a methodology for choosing the best number of HPCA for prediction accuracy, 4- *Simulations* contain two simulation scenarios, 5- *Application to the EEG-fMRI dataset*, 6- *Conclusion* contains the comparative study with some comments and 7- *References*. We also provide supplementary materials including the R codes.

2- The Functional Regression with HPCA

2.1 – The HPCA Decomposition

With the same notation and methodology of [22], we mention the HPCA decomposition briefly in this section. Assume that $Y_i(r, \omega, s)$ has three dimensions of the de-mean observation for the subject $i, \{i = 1, \ldots, n\}$, the first dimension has scalar domain $r, r = 1, \ldots, R$, the second dimension has functional domain $\omega, \omega \in \Omega$, and the third and desired dimension has functional domain $s, s \in S$. We want to decompose $Y_i(r, \omega, s)$ with the 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:

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

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

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

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

2.2 – The Function-on-multiple-Functions Regression

In this part, we use the FOF regression [19], the response and covariates are functional data. For example Evoked Related Potential (ERP) curves on time or frequently
domains. We first decompose hybrid data with HPCA and then reconstruct them with different number of HPCAs. The reconstructed hybrid data have three dimensions and we merge them to 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 usually only interested in values of the ERP at the time domain, \( s \). And the other two dimensions are the regional with the scalar domain, \( r \) and the longitudinal dimension, \( \omega \), with the functional domain.

Therefore, we first reconstruct the dataset in equation (3) with the specified number of the HPCA, \( \hat{Y}_i(r, \omega, s)_{HPCA,KLM} \). We use hat notation because it is an estimation of 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 \( \hat{Y}_i(r, \omega, s)_{HPCA,KLM} \) on the two other dimensions, for example, on the first, \( r \), and second, \( \omega \), dimensions. And as stated in the (4), we average point-wise on their domain of our desired dimension, for example, the third dimension, \( s \).

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

\[
\frac{1}{R} \sum_{r=1}^{R} \int_{r-\infty}^{+\infty} \hat{Y}_i(r, \omega, s)_{HPCA,KLM} d\omega = W_i(s)_{KLM} \quad (4)
\]

The advantage of this method instead of using only FPCA on the desired dimension, \( s \), is that our method contains the most important variations of the other two-dimension in the dataset and it avoids considering the unimportant variations on the data and help to reduce the complexity of it. The functional covariate, \( W_i(s)_{KLM} \), with the functional domain, \( s \), and the number of HPCs, \( KLM \), is calculated by (4) for regression analysis in the next steps.

We assume that \( W_i(s)_{KLM} \) (4) is a response function in the following FOF regression [19]:

\[
W_i(s)_{KLM} = m(s) + \sum_{j=1}^{p} \int_{0}^{1} \chi_j(g) \beta_j(g, s) dg + \epsilon(s) \quad (5)
\]

The \( \chi_1(g), ..., \chi_p(g) \) are predictive functions and \( \beta_1(g, s), ..., \beta_p(g, s) \) are \( p \) –coefficient functions, \( m(s) \) and \( \epsilon(s) \) are intercept function and noise function. The focus of this model with signal compression methods is predicting the response [19].
3- The number of HPCA

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

$$\xi_{i,1}V_1(r)\phi_1(\omega)\psi_1(s) = \hat{Y}_i(r, \omega, s)_{HPCA,1}$$

$$\xi_{i,1}V_1(r)\phi_1(\omega)\psi_1(s) + \xi_{i,2}V_1(r)\phi_1(\omega)\psi_2(s) = \hat{Y}_i(r, \omega, s)_{HPCA,2}$$

$$...$$

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

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

| Algorithm 1 – FOF Regression (6) |
|----------------------------------|

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

2- for $q = 1$ to $q = KLM$

2-1 The reconstruction based on the HPCA from 1 to $q$:

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

2-2 Merging on two dimensions:

$$\frac{1}{R} \sum_{r=1}^{R} \int_{-\infty}^{+\infty} \hat{Y}_i(r, \omega, s)_{HPCA,q} d\omega = W_i(s)_{HPCA,q}$$

2-3 Fit FOF based on the $W_i(s)_{HPCA,q}$

$$W_i(s)_{HPCA,q} = m(s) + \sum_{j=1}^{p} \int_{0}^{1} \chi_j(g)\beta_j(g,s)dg + \epsilon(s)$$

2-4 Predict the $\widehat{W}_i(s)_{HPCA,q}$ within (2-3)
2- Calculate the MSPE

\[
\text{Mean Squared Prediction Error} = MSPE(q) = \frac{1}{n \times M} \sum_{i=1}^{n} \sum_{m=1}^{M} (W_i(s_m)_{HPCA,q} - \hat{W}_i(s_m)_{HPCA,q})^2
\]

\{i, 1 \leq i \leq n, number of subject\}, \{m, 1 \leq m \leq M, number of observation in each curve\}

3- Find the minimum \(MSPE(q_{min}) = \min\{q = 0 \ or \ MSPE(q), 1 \leq q \leq KLM\}\).

4- The best number of HPCA is \(q_{min}\).

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

4- Simulations

4-1- Simulation Scenario 1

The function-on-function regression: The data simulation process has two steps: We first generate the \(Y_i(r, \omega, s)\) with two sample size, 1) \(\{i, i = 1, ... , 20\}\) and 2) \(\{i, i = 1, ... , 140\}\) and two status for the 1) complete and 2) sparse observation for \(\omega\) as mentioned in the [22], then we merge each of them with (1) to (5), as \(W_i(s)_{HPCA,0}\).

In the second step, we generate the fof regression data as mentioned at [19] with 10 covariates functions, \(\theta(g, s)\), and two status for coefficients 1) all 10 coefficients, \(\theta(g, s)\), have a value greater than 0 and 2) only 50% of \(\theta(g, s)\) have the non-zero value. Therefore, we have \(A_i(s) = m(s) + \sum_{j=1}^{10} \int_{0}^{1} x_j(g) \theta_j(g, s) dg + \epsilon(s)\).

We combine the first and second steps with \(W_i(s)_{HPCA,0} + A_i(s) = D_i(s)_{HPCA,0}\) and we fit the fof regression with the following equation, as (6):

\[
D_i(s)_{HPCA,0} = m(s) + \sum_{j=1}^{p} \int_{0}^{1} x_j(g) \beta_j(g, s) dg + \epsilon(s)
\]

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 the coefficients:

\[
\text{MSE } \beta_1(s, t) = \frac{1}{n \times S \times T} \sum_{i=1}^{n} \sum_{s=1}^{S} \sum_{t=1}^{T} (\beta_1(s, t) - \hat{\beta}_1(s, t))^2
\]

\[
\text{MSE } \beta_{10}(s, t) = \frac{1}{n \times S \times T} \sum_{i=1}^{n} \sum_{s=1}^{S} \sum_{t=1}^{T} (\beta_{10}(s, t) - \hat{\beta}_{10}(s, t))^2
\]

\ldots
For the prediction accuracy of model in the train and test, separately:

\[ MSPE_{\text{pred}}(\text{train}) = \frac{1}{n_{\text{train}}} \sum_{i=1}^{n_{\text{train}}} \sum_{j=1}^{l} (D_{i,\text{train}}(s_j)_{HPCA,0} - \tilde{D}_{i,\text{train}}(s_j)_{HPCA,0})^2 \]

\[ MSPE_{\text{pred}}(\text{test}) = \frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} \sum_{j=1}^{l} (D_{i,\text{test}}(s_j)_{HPCA,0} - \tilde{D}_{i,\text{test}}(s_j)_{HPCA,0})^2 \]

\[ COR_{\text{pred}}(\text{train}) = COR(D_{i,\text{train}}(s)_{HPCA,0}, \tilde{D}_{i,\text{train}}(s)_{HPCA,0}) \]

\[ COR_{\text{pred}}(\text{test}) = COR(D_{i,\text{test}}(s)_{HPCA,0}, \tilde{D}_{i,\text{test}}(s)_{HPCA,0}) \]

We also 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.

We explain briefly the result: in the table 1, the \( MSE \beta_{1,...,10}(s,t) \) are greater for the complete rather than 50 % sparsity of \( \beta \). The \( MSPE_{\text{pred}}(\text{train}) \) are not very different in all situations with each other, but \( MSPE_{\text{pred}}(\text{test}) \) are different. The first difference is the \( MSPE_{\text{pred}}(\text{test}) \) for complete \( \beta \) are greater than the 50 % sparsity of \( \beta \). The second difference is with increasing the sample size, their values are smaller than big sample size. But its value for the sparsity for \( \omega \) are not different with other. The \( COR_{\text{pred}}(\text{train}) \) are negligibly bigger than \( COR_{\text{pred}}(\text{test}) \), but they are not very different between scenarios. The elapsed time for 50 % sparsity of \( \beta \) are greater than the complete \( \beta \) and they increase significantly with the number of sample size. But this behaviour are not observed for the system time. Therefore, the sparsity of \( \beta \) will affect the model performance and the sparsity of the \( \omega \) don’t have any significant effect on them.

4-2- Simulation Scenario 2

This simulation scenario is very similar to the simulation scenario 1, but in this one we study the number of HPCA. Again, we first generate the \( Y_i(r, \omega, s) \) \{\( i, i = 1, ..., 20 \} \) as mentioned in the [22], then we merge them with (4) in three ways: 1) Without HPCA decomposition, as \( W_i(s)_{HPCA,0} \), 2) With only the first HPCA decomposition, as \( W_i(s)_{HPCA,1} \), 3) With All HPCA decomposing, as \( W_i(s)_{HPCA,All} \) (All HPCA has 90% FVE on each dimension). We also generate the fof regression data as mentioned at [19] with 10 covariates functions, \( \theta(g,s) \), and two status for coefficients 1) all 10 coefficients, \( \theta(g,s) \), have the value greater than 0 and 2) only 50% of \( \theta(g,s) \) have the non-zero
value. Therefore, we have \( A_i(s) = m(s) + \sum_{j=1}^{10} \int_0^1 \chi_j(g)\theta_j(g,s)\,dg + \varepsilon(s) \). We define:

\[
W_i(s)_{HPCA,0} + A_i(s) = D_i(s)_{HPCA,0}
\]

\[
D_i(s)_{HPCA,0} = m(s) + \sum_{j=1}^{p} \int_0^1 \chi_j(g)\beta_j(g,s)\,dg + \varepsilon(s)
\]

\[
W_i(s)_{HPCA,1} + A_i(s) = D_i(s)_{HPCA,1}
\]

\[
D_i(s)_{HPCA,1} = m(s) + \sum_{j=1}^{p} \int_0^1 \chi_j(g)\beta_j(g,s)\,dg + \varepsilon(s)
\]

\[
W_i(s)_{HPCA,All} + A_i(s) = D_i(s)_{HPCA,All}
\]

\[
D_i(s)_{HPCA,All} = m(s) + \sum_{j=1}^{p} \int_0^1 \chi_j(g)\beta_j(g,s)\,dg + \varepsilon(s)
\]

In each model, \( D_i(s)_{HPCA,0}, D_i(s)_{HPCA,1} \) and \( D_i(s)_{HPCA,All} \), we split the data to the train (70%) and test (30%) and we fit the models on the train dataset. We calculate the summary statistics as mentioned in the scenario 2. We repeat the iteration 100 times. The median (the first quartile, the third quartile) in all simulations were reported.

We explain briefly the result: in the table 2, the most of MSE \( \beta_{1,...,10}(s,t) \) are greater for the complete rather than 50% sparsity of \( \beta \). The 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 the 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 the MSPE\(_{Pred(test)}\). The COR\(_{Pred(train)}\) and COR\(_{Pred(test)}\) are high (more than 90%) and the differences are negligible. The elapsed time are lowest for the first HPCA model among others.

5- Application to the EEG-fMRI dataset

In this application, we use the EEG-fMRI dataset which was analysed with FOF regression previously[19]. The novelty of this new analysis are: 1) We analyse the event-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 internal attentions in the brain with 17 participants (6 females, mean age is 27.7 years) [28]. There are two tasks for each
experiment including auditory and visual stimuli. And 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. The EEG was captured based on 64 channels and the preprocessing steps are artifact removal, baseline correction and epoching (-200 to 800 milliseconds of each stimuli) that they were done with EEGLAB[29]. The fMRI dataset was captured with 3 Tesla Philips scanner and the dataset was preprocessed 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/). The ERP waves are response functions and the ROI time series for different Brodmann Areas in the brain are functional covariates. We conduct two analyses: the first one is choosing the best number of HPCA and the second is FOF regression for EEG-fMRI by the region of the brain.

We first decompose the ERP curves with the HPCA from first to the total number of HPCAs for the visual standard task. The KML is 14 that captures 90% of FVE in each dimension. Then we fit model (7) on the training samples and compare with the test. We do this procedure 100 times with different training/testing samples. Finally, we compare the \( \text{MSPE}_{\text{Pred(train)}} \), \( \text{MSPE}_{\text{Pred(test)}} \), \( \text{COR}_{\text{Pred(train)}} \), \( \text{COR}_{\text{Pred(test)}} \) and \( \text{Time}_{\text{elapsed}} \) for different number of HPCAs.

We see the first HPCA has the lowest MSPE in both train/test. With increasing number of HPCA, the MSPE increases. Notably, the model without HPCA has the largest MSPE. (figure 1) The same pattern are seen in the correlation and the elapsed time. (figure 2 and 3)

Therefore, we use the first HPCA and analyses the whole dataset. In this regard, we analyse the whole brain and by the regions and compare the MSPE and Correlation in the train/test for the auditory and visual tasks. The 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 are also provided by the different regions of the brain. (table 3)

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). All R codes with their instructions are attached in the supplementary files.
6- Conclusion

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

Note that this functional regression works with one-dimensional function, for example, it is common to use the ERP over time or frequency in the EEG dataset. We propose a method to convert the hybrid data to one-dimensional data by pooling and averaging two other dimensions as mentioned in the functional regression models. Note that instead of using the observed hybrid data, we use the reconstructed data from the HPCA eigencomponents. We also find the best number of HPCA based on the functional regression MSPE. Surprisingly, the simulations and real-world application with the EEG-fMRI show that the best performance of the regression was achieved by only the first eigencomponents. We can count the following reasons to explain this situation, first, the simulations are designed to show the EEG structure of the data, which has three dimensions, and most of the variations from two other dimensions are captured with the first HPCA; second, the higher number of HPCA were used in the reconstructed data, it is getting closer to the observed data, therefore their complexity increases that this matter affects the prediction accuracy in the functional regression. The future direction of the proposed method is using other indices than MSPE of prediction to choose the number of HPCA, using different types of functional regressions, and applying the methodology on their applications other than the EEG structure. Finally, we conclude that our methodology improves the prediction of the experiments with the EEG datasets. And we recommend that instead of using the functional PCA on the desired dimension, for example, ERP over time or frequency, reconstruct the data with HPCA and average it on the other two dimensions for functional regression models.
7- References:

1. Dunn Lii, W., D. Scott, and W. Glen, *Principal components analysis and partial least squares regression*. Tetrahedron computer methodology, 1989. 2(6): p. 349-376.

2. Johnson, R.A. and D.W. Wichern, *Applied multivariate statistical analysis*. Vol. 5. 2002: Prentice hall Upper Saddle River, NJ.

3. Mandel, J., *Use of the singular value decomposition in regression analysis*. The American Statistician, 1982. 36(1): p. 15-24.

4. Silverman, J.O.R.B.W., *Functional Data Analysis*. 2005.

5. Ferraty, F. and P. Vieu, *Nonparametric functional data analysis: theory and practice*. 2006: Springer Science & Business Media.

6. James, G.M., T.J. Hastie, and C.A. Sugar, *Principal component models for sparse functional data*. Biometrika, 2000. 87(3): p. 587-602.

7. Staicu, A.-M., C.M. Crainiceanu, and R.J. Carroll, *Fast methods for spatially correlated multilevel functional data*. Biostatistics, 2010. 11(2): p. 177-194.

8. Febrero-Bande, M., 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, 2017. 85(1): p. 61-83.

9. Hullait, H., et al., *Robust Function-on-Function Regression*. Technometrics, 2020: p. 1-14.

10. Chen, K. and H.-G. Müller, *Modeling repeated functional observations*. Journal of the American Statistical Association, 2012. 107(500): p. 1599-1609.

11. Greven, S., et al., *Longitudinal functional principal component analysis*, in Recent Advances in Functional Data Analysis and Related Topics. 2011, Springer. p. 149-154.

12. Di, C.-Z., et al., *Multilevel functional principal component analysis*. The annals of applied statistics, 2009. 3(1): p. 458.

13. Hasenstab, K., et al., *A multi-dimensional functional principal components analysis of EEG data*. Biometrics, 2017. 73(3): p. 999-1009.

14. Kuenzer, T., S. Hörmann, and P. Kokoszka, *Principal component analysis of spatially indexed functions*. Journal of the American Statistical Association, 2020: p. 1-13.

15. Liu, C., S. Ray, and G. Hooker, *Functional principal component analysis of spatially correlated data*. Statistics and Computing, 2017. 27(6): p. 1639-1654.

16. Zhou, L., et al., *Reduced rank mixed effects models for spatially correlated hierarchical functional data*. Journal of the American Statistical Association, 2010. 105(489): p. 390-400.

17. Reimherr, P.K.M., *Introduction to Functional Data Analysis*. 2017.

18. Luo, R. and X. Qi, *Function-on-function linear regression by signal compression*. Journal of the American Statistical Association, 2017. 112(518): p. 690-705.

19. Qi, X. and R. Luo, *Function-on-function regression with thousands of predictive curves*. Journal of Multivariate Analysis, 2018. 163: p. 51-66.

20. Crainiceanu, C.M. and A.J. Goldsmith, *Bayesian functional data analysis using WinBUGS*. Journal of statistical software, 2010. 32(11).

21. Kowal, D.R. and D.C. Bourgeois, *Bayesian function-on-scalars regression for high-dimensional data*. Journal of Computational and Graphical Statistics, 2020: p. 1-10.

22. Scheffler, A., et al., *Hybrid principal components analysis for region-referenced longitudinal functional EEG data*. Biostatistics, 2020. 21(1): p. 139-157.

23. Scheffler, A.W., et al., *Covariate-adjusted region-referenced generalized functional linear model for EEG data*. Statistics in Medicine, 2019. 38(30): p. 5587-5602.

24. Aneiros-Pérez, G. and P. Vieu, *Semi-functional partial linear regression*. Statistics & Probability Letters, 2006. 76(11): p. 1102-1110.

25. Liebl, D. and F. Walders, *Parameter regimes in partial functional panel regression*. Econometrics and Statistics, 2019. 11: p. 105-115.

26. Febrero-Bande, M., W. González-Manteiga, and M.O. de la Fuente, *Variable selection in functional additive regression models*. Computational Statistics, 2019. 34(2): p. 469-487.

27. Christensen, R., et al., *Bayesian ideas and data analysis: an introduction for scientists and statisticians*. 2011: CRC Press.

28. Walz, J.M., et al., *Simultaneous EEG-fMRI reveals temporal evolution of coupling between supramodal cortical attention networks and the brainstem*. Journal of Neuroscience, 2013. 33(49): p. 19212-19222.
29. Delorme, A. and S. Makeig, EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of neuroscience methods, 2004. 134(1): p. 9-21.
Table 1 – Percentiles 50% (25%, 75%) of the simulation scenario 1.

| $\omega$ | $\beta$ | Complete | 50% Sparsity | Complete | 50% Sparsity | Complete | 50% Sparsity | Complete | 50% Sparsity |
|---|---|---|---|---|---|---|---|---|---|
| MSE $\beta_i(s,t)$ | 0.00 | 1.00 | 0.50 | 0.95 | 0.75 | 0.98 | 0.90 | 0.98 | 0.97 |
| MSE $\beta_2(s,t)$ | 0.04 | 0.76 | 0.64 | 0.84 | 0.70 | 0.82 | 0.80 | 0.83 | 0.81 |
| MSE $\beta_2(s,t)$ | 0.00 | 0.50 | 0.64 | 0.84 | 0.70 | 0.82 | 0.80 | 0.83 | 0.81 |
| MSE $\beta_2(s,t)$ | 0.04 | 0.76 | 0.64 | 0.84 | 0.70 | 0.82 | 0.80 | 0.83 | 0.81 |
| MSE $\beta_2(s,t)$ | 0.00 | 0.50 | 0.64 | 0.84 | 0.70 | 0.82 | 0.80 | 0.83 | 0.81 |
| MSE $\beta_2(s,t)$ | 0.04 | 0.76 | 0.64 | 0.84 | 0.70 | 0.82 | 0.80 | 0.83 | 0.81 |
| MSE $\beta_2(s,t)$ | 0.00 | 0.50 | 0.64 | 0.84 | 0.70 | 0.82 | 0.80 | 0.83 | 0.81 |
| MSE $\beta_2(s,t)$ | 0.04 | 0.76 | 0.64 | 0.84 | 0.70 | 0.82 | 0.80 | 0.83 | 0.81 |

Table 2 – Percentiles 50% (25%, 75%) of the simulation scenario 2.

| NO HPCA | HPCA | First HPCA | All HPCA |
|---|---|---|---|
| $\beta$ | $\beta$ | $\beta$ | $\beta$ |
| MSE $\beta_i(s,t)$ | 0.00 | 0.50 | 0.64 | 0.84 | 0.70 | 0.82 | 0.80 | 0.83 | 0.81 |
| MSE $\beta_2(s,t)$ | 0.04 | 0.76 | 0.64 | 0.84 | 0.70 | 0.82 | 0.80 | 0.83 | 0.81 |
| MSE $\beta_2(s,t)$ | 0.00 | 0.50 | 0.64 | 0.84 | 0.70 | 0.82 | 0.80 | 0.83 | 0.81 |
| MSE $\beta_2(s,t)$ | 0.04 | 0.76 | 0.64 | 0.84 | 0.70 | 0.82 | 0.80 | 0.83 | 0.81 |
| MSE $\beta_2(s,t)$ | 0.00 | 0.50 | 0.64 | 0.84 | 0.70 | 0.82 | 0.80 | 0.83 | 0.81 |
| MSE $\beta_2(s,t)$ | 0.04 | 0.76 | 0.64 | 0.84 | 0.70 | 0.82 | 0.80 | 0.83 | 0.81 |
| MSE $\beta_2(s,t)$ | 0.00 | 0.50 | 0.64 | 0.84 | 0.70 | 0.82 | 0.80 | 0.83 | 0.81 |
| MSE $\beta_2(s,t)$ | 0.04 | 0.76 | 0.64 | 0.84 | 0.70 | 0.82 | 0.80 | 0.83 | 0.81 |

Number of iteration 100 times, * has 96 iteration and ** has 99 iterations.
Table 3 – The FOF result from EEG-fMRI Dataset by the region of the brain.

| Tasks | Region | 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 | 0.98777 | 0.92011 |
|       | Left Temporal | Standard | 0.00116 | 0.02427 | 0.99942 | 0.98784 |
|       | | Target | 0.00005 | 0.00062 | 0.99988 | 0.99969 |
|       | Right Temporal | Standard | 0.00040 | 0.06260 | 0.99980 | 0.96833 |
|       | | Target | 0.02455 | 0.12669 | 0.98763 | 0.93528 |
| Visually | 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 |
Figure 1- The MSPE prediction error by the number of HPCA in training/test samples.

Figure 2- The Correlation by the number of HPCA in training/test samples.
Figure 3- The elapsed time by the number of HPCA in training/test samples.