The original MRI data are preprocessed using advanced normalization tools (Avants et al., 2011). The preprocessing procedure consists of several steps: N4 bias correction, registration-based brain extraction, and a prior-based N4-Atropos 6 tissue segmentation using the oasis template. By performing multi-atlas cortical parcellation, we obtain the brain local volumetric measures of 101 ROIs defined by the manually edited labels of the publicly available MindBoggle-101 dataset (Klein and Tourville, 2012).

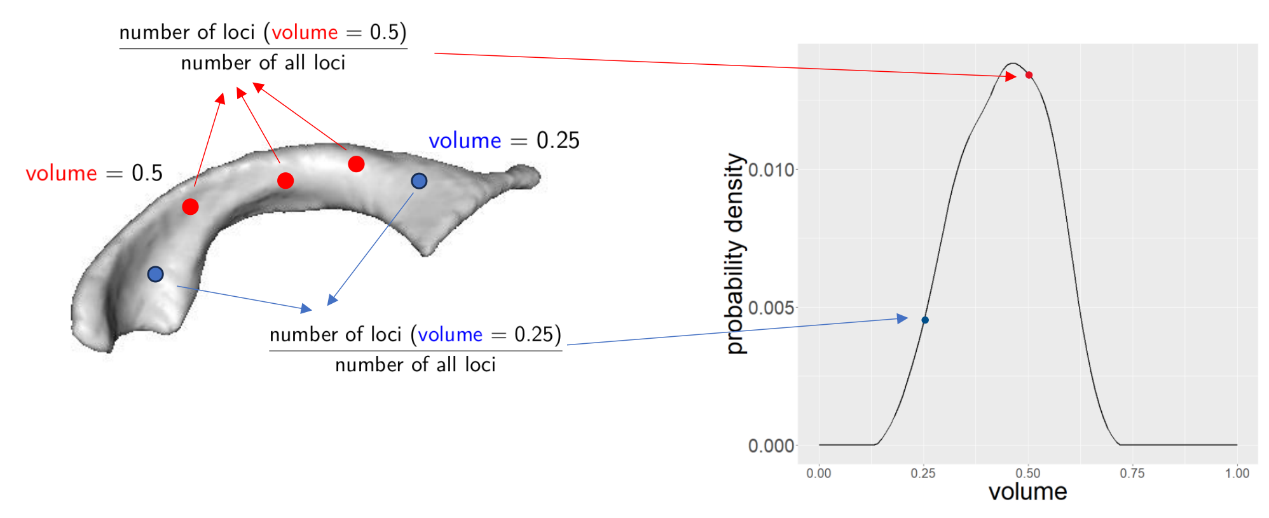
We illustrate how to convert a single MRI data into a volume density curve.

As shown in the following figures, the left figure displays the regions corresponding to the left hippocampus in an MRI scan. By uniformly partitioning the regions and counting the loci of the volume, we convert the MRI scan into the

density curve of the volume shown in the right figure, which records the frequency of each volume appearing in the partitioned left figure.

Then, to deal with the restrictions that density functions do not live

in a linear space, following Petersen and Müller (2016); Li et al. (2023), we employ the log quantile density transformation and take the log quantile density functions as functional variables. Concretely, let *f, F, Q* =*F-1* represent the density function, the cumulative distribution function and the quantile functions of the volume. The log quantile density transformation is given by log{*dQ*(*t*)*/dt*}= log{*d F-1*(*t*)*/dt*}=-log[*f* {*Q*(*t*)}].



Left: the distribution of certain loci on the left hippocampus; Right: the volume density curve of the left hippocampus for an individual. Blue and Red represent loci with different volumes (Blue = 0.25, Red = 0.5). By calculatingthe frequency, MRI data is converted into volume density curves.

Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., and Gee, J. C. (2011). A reproducible evaluation of ants similarity metric performance in brain image registration. *NeuroImage*, 54(3):2033–2044.

Klein, A. and Tourville, J. (2012). 101 labeled brain images and a consistent human cortical labeling protocol. *Front. Neurosci.*, 6(8):171.

Petersen, A. and Müller, H. G. (2016). Functional data analysis for density functions by transformation to a Hilbert space. *Ann. Stat.*, 44(1):183–218.

Li, T., Zhu, H., Li, T., and Zhu, H. (2023). Asynchronous functional linear regression models for longitudinal data in reproducing kernel Hilbert space. *Biometrics*, 79(3):1880–1895.

This file enumerates the code provided in this article and specifically explains how to reproduce the figures and tables provided. Some codes needs to load data from <https://github.com/LinhzLab/SFFPCA>.

----------------------------------------------------------------------------------------------------------------------------------------------------------

**Table 1**: prediction error of X in the realdata analysis

You need to load packages fda, Matrix and load ADNI\_realdata.RData. One only need to run realdata\_Xpredict.R and can obtain the results of the proposed method in the real data analysis. (Recorded in mean(PE) and sd(PE))

The results of SF-FPCA\_f are recorded in mean(PE1) and sd(PE1)

The results of SF-FPCA\_B are recorded in mean(PE3) and sd(PE3)

The results of SF-FPCA\_{B+f} are recorded in mean(PE4) and sd(PE4)

**Table 2**: prediction error of Y in the realdata analysis

The prediction error of Y of the proposed method are the same in Table 2 and Figure 3.

You need to load packages fda, Matrix, MASS,glmnet and load ADNI\_realdata.RData. One only need to run realdata\_Y.R and can obtain the results of the proposed method in the real data analysis. (Recorded in mean(PE) and sd(PE))

**Table 3**: prediction error of X in the simulations

Simulation\_X1.R and Simulation\_X2.R are used to reproduce the results of the proposed method in Simulation Scenario X1 and Simulation Scenario X2 in the first row of Table 3.

Simulation\_X1.R includes the generation of X1 and the calculation flow of the proposed method. You only need to load packages MASS, fda, Matrix and load SimulationX1\_mu.RData. SimulationX1\_mu.RData includes the mean values. One only need to run Simulation\_X1.R and can obtain the results of the proposed method in Simulation Scenario X1. (Recorded in mean(PE) and sd(PE))

Simulation\_X2.R includes the calculation flow of the proposed method in X2.

First, you need to down load [SimulationX2\_mu1.RData](https://github.com/LinhzLab/SFFPCA/blob/main/SimulationX2_mu1.RData), [SimulationX2\_mu2.RData](https://github.com/LinhzLab/SFFPCA/blob/main/SimulationX2_mu1.RData), [SimulationX2\_mu3.RData](https://github.com/LinhzLab/SFFPCA/blob/main/SimulationX2_mu1.RData) from <https://github.com/LinhzLab/SFFPCA>. Then, you need to load packages MASS, fda, Matrix.

SimulationX2\_mu1.RData, SimulationX2\_mu2.RData, SimulationX2\_mu3.RData include the synthetic data in Simulation Scenario X2. One only need to run Simulation\_X2.R and can obtain the results of the proposed method in Simulation Scenario X2. (Recorded in mean(PE) and sd(PE))

**Table 4**: prediction error of Y in the simulations

Simulation\_Y1.R and Simulation\_Y2.R are used to reproduce the results of the proposed method in Simulation Scenario Y1

and Simulation Scenario Y2 in the first row of Table 4.

Simulation\_Y1.R includes the generation of Y1 and the calculation flow of the proposed method. You need to load packages MASS, fda, Matrix and load SimulationX1\_mu.RData. SimulationX1\_mu.RData includes the used covariates. One only need to run Simulation\_Y1.R and can obtain the results of the proposed method in Simulation Scenario Y1. (Recorded in mean(PE\_SFFPCA) and sd(PE\_SFFPCA))

Simulation\_Y2.R includes the generation of Y2 and the calculation flow of the proposed method.

First, you need to down load [SimulationX2\_mu1.RData](https://github.com/LinhzLab/SFFPCA/blob/main/SimulationX2_mu1.RData), [SimulationX2\_mu2.RData](https://github.com/LinhzLab/SFFPCA/blob/main/SimulationX2_mu1.RData), [SimulationX2\_mu3.RData](https://github.com/LinhzLab/SFFPCA/blob/main/SimulationX2_mu1.RData) from <https://github.com/LinhzLab/SFFPCA>. Then, you need to load packages MASS, fda, Matrix.

SimulationX2\_mu1.RData, SimulationX2\_mu2.RData, SimulationX2\_mu3.RData include the used covariates . One only need to run Simulation Y2.R and can obtain the results of the proposed method in Simulation Scenario Y2. (Recorded in mean(PE\_SFFPCA) and sd(PE\_SFFPCA))

**Figure S2**: consistency results

You need to load packages MASS, fda, Matrix and run Simulation\_Bpiece\_n100.R for the case n=100, Simulation\_consistency\_n200.R for the case n=200,

and Simulation\_consistency\_n400.R for the case n=400.

For Simulation\_consistency\_n200.R and Simulation\_consistency\_n400.R, the results are recorded in NNMI, FF\_result, PP\_result and SS\_result.

For Simulation\_Bpiece\_n100.R, the results are recorded in NMI, FF\_result[,2], PP\_result[,2], and SS\_result[,2].

After obtaining the results of the three R files, you need to load package ggplot2 and run consistency\_plot.R. The figures are recorded in pF, pP, pS.

**Table S1**: selection of number of factors

You need to load packages MASS, fda, Matrix and run Simulation\_qselect.R. The selection results are recorded in Q2,Q3,Q4,Q5,Q6.

**Table S2**: selection of number of eigenfunctions

You need to load packages MASS, fda, Matrix and run Simulation\_Kselect.R. The selection results are recorded in KKK2,KKK4,KKK6,KKK8,KKK10.

**Tables S3 and S4**: computational costs

You need to run Simulation\_npT\_computationl\_cost.R for Table S4 and Simulation\_qK\_computationl\_cost.R for Table S5. The results are calculated by R (version 4.4.0) on a 14-core computer with 32 GB of RAM. It should be noted that the specific calculation time on different computers will be different from Tables 8 and 9, so only pay attention to the relative size of the calculation time in each case.

**Figure 2**: prediction error of Y in the realdata analysis compared to the scalar volume

The prediction error of Y of the proposed method are the same in Table 2 and Figure 2.

You need to load packages fda, Matrix, MASS, glmnet, ggplot2 and load ADNI\_realdata. One only need to run realdata\_Y.R and can obtain the results of the proposed method (Fun-LR)in the real data analysis. (Recorded in mean(PE) and sd(PE)).

Vol-RR is recorded in mean(PE\_vol\_ridge\_2) and sd(PE\_vol\_ridge\_2).

Vol-F2-LR is recorded in mean(PE\_fac\_lm\_2) and sd(PE\_fac\_lm\_2).

Vol-F10-LR is recorded in mean(PE\_fac\_lm\_10) and sd(PE\_fac\_lm\_10).

Vol-F20-LR is recorded in mean(PE\_fac\_lm\_20) and sd(PE\_fac\_lm\_20).

Vol-F20-LR is recorded in mean(PE\_fac\_ridge\_20) and sd(PE\_fac\_ridge\_20).

The Figure 2 are recorded in pp.

**Figure 5**: the estimates of the coefficient functions

You need to load packages fda, Matrix, MASS,glmnet and load ADNI\_realdata. One only need to run realdata\_Y.R. The figures are recorded in a1-a4.