



Published in final edited form as:

Stat (Int Stat Inst). 2016 ; 5(1): 108–118. doi:10.1002/sta4.109.

Interactive graphics for functional data analyses

Julia Wrobel^{1,*}, So Young Park², Ana Maria Staicu², and Jeff Goldsmith¹

¹Department of Biostatistics, Mailman School of Public Health, Columbia University

²Department of Statistics, North Carolina State University

Abstract

Although there are established graphics that accompany the most common functional data analyses, generating these graphics for each dataset and analysis can be cumbersome and time consuming. Often, the barriers to visualization inhibit useful exploratory data analyses and prevent the development of intuition for a method and its application to a particular dataset. The `refund.shiny` package was developed to address these issues for several of the most common functional data analyses. After conducting an analysis, the `plot_shiny()` function is used to generate an interactive visualization environment that contains several distinct graphics, many of which are updated in response to user input. These visualizations reduce the burden of exploratory analyses and can serve as a useful tool for the communication of results to non-statisticians.

Keywords

Functional principal component analysis; multilevel functional data; longitudinal functional data; function-on-scalar regression

1 Introduction

Functional data analysis (FDA) has become a popular and useful framework for applications in which the unit of measurement is a function, curve or image. Conceptually, FDA leverages the underlying data structure, often temporal or spatial, to improve understanding of patterns and variation. A wide array of tools have been developed for the functional data setting, for example, functional principal component analysis (FPCA) and regression models using functional responses (Ramsay and Silverman, 2005; Morris, 2015; Sørensen et al., 2013). The basic unit of observation is the curve $Y_i(t)$ for subjects $i \in \dots, I$ in the cross-sectional setting and $Y_{ij}(t)$ for subject i at visit $j \in \dots, J_i$ for the multilevel or longitudinal structure. Methods for functional data are typically presented in terms of continuous functions, but in practice data are observed on a discrete grid that may be sparse or dense at the subject level and that may be the same across subjects or irregular.

Many methods for FDA have standard visualization approaches that clarify the results of analyses; examples include scree plots for FPCA and coefficient function plots for function on scalar regression. Clear visualizations aid in exploratory analysis and help to

*jw3134@cumc.columbia.edu.

communicate results to non-statistical collaborators. However, creating useful plots is often time consuming and must be repeated each time a model is changed, and no software currently exists to facilitate this process.

The `refund.shiny` package (Goldsmith and Wrobel, 2015) creates interactive visualizations for functional data analyses, allowing researchers to create common graphics for standard analyses with just a few lines of code. Currently, `refund.shiny` builds plots for functional principal component analysis (FPCA), multilevel FPCA (MFPCA), time-varying FPCA (TV-FPCA), and function-on-scalar regression (FoSR). The workflow separates analysis and visualization steps: analyses are performed by functions in the `refund` package (Crainiceanu et al., 2015) and interactive visualizations are generated by the `plot_shiny()` function in the `refund.shiny` package. Changes to the analysis – increasing the number of retained principal components, for example, or augmenting a regression model with new predictors – are easily incorporated into the graphical interface. User interaction with the displayed graphics facilitates comparisons and streamlines navigation between visualizations.

We illustrate the tools in `refund.shiny` using a single dataset, which we describe briefly here. The diffusion tensor imaging (DTI) dataset available in the `refund` package includes cerebral white matter tracts for multiple sclerosis patients and healthy controls. White matter tracts are collections of axons, the projections of neurons that transmit electrical signals that are coated by a fatty substance called myelin (Greven et al., 2010; Goldsmith et al., 2011; Staicu et al., 2012). DTI is a magnetic resonance imaging modality that measures diffusion of water in the brain; because water movement is restricted in white matter fibers, DTI allows the quantification of white matter tract integrity. The DTI dataset contains tract profiles – continuous summaries of tract properties along their major axis – for 142 subjects across multiple visits, with a median of 4 scans per subject. The dataset includes tract profiles for several tracts, the PASAT score (a continuous variable that indicates brain reactivity and attention span), subject sex, subject ID, visit number, and time of visit (Strauss et al., 2006). Because we observe tract profiles for each subject over time, the DTI dataset is a functional dataset with longitudinal structure; in order to use the same dataset across examples we sometimes neglect this structure or subset the data. The following code can be used to install `refund` and `refund.shiny` and load the DTI data:

```
> install.packages("refund.shiny")

> library(refund.shiny)

> library(refund)

> data(DTI)
```

Sections 2, 3, 4, and 5 each provide a brief methodological overview of an analysis technique for FDA and describe the corresponding interactive visualization tools in the `refund.shiny` package. Section 6 details the structure of the `refund.shiny` package. We close in section 7 with a discussion.

2 Functional Principal Component Analysis

We start with FPCA, one of the most common exploratory tools for functional datasets.

2.1 FPCA Model

FPCA characterizes modes of variability by decomposing functional observations into population level basis functions and subject-specific scores (Ramsay and Silverman, 2005). The basis functions have a clear interpretation, analogous to that of PCA: the first basis function explains the largest direction of variation, and each subsequent basis function describes less. The FPCA model is typically written

$$Y_i(t) = \mu(t) + \sum_{k=1}^K c_{ik} \psi_k(t) + \varepsilon_i(t) \quad (1)$$

where $\mu(t)$ is the population mean, $\psi_k(t)$ are a set of orthonormal population-level basis functions, c_{ik} are subject-specific scores with mean zero and variance λ_k , and $\varepsilon_i(t)$ are residual curves. Estimated basis functions $\hat{\psi}_1(t), \hat{\psi}_2(t), \dots, \hat{\psi}_K(t)$ and corresponding variances $\hat{\lambda}_1 > \hat{\lambda}_2 \dots \hat{\lambda}_K$ are obtained from a truncated Karhunen-Loève decomposition of the sample covariance $\hat{\Sigma}(s, t) = \widehat{\text{Cov}}(Y_i(s), Y_i(t))$. In practice, the covariance $\hat{\Sigma}(s, t)$ is often smoothed using a bivariate smoother that omits entries on the main diagonal to avoid a “nugget effect” attributable to measurement error, and scores are estimated in a mixed model framework (Yao et al., 2005; Goldsmith et al., 2013). The truncation lag K is often chosen so that the resulting approximation accounts for at least 95% of observed variance.

2.2 Graphics for FPCA

Our example uses the `fpca.sc()` function from the `refund` package. Several other implementations of FPCA are available in `refund`, including `fpca.face()`, `fpca.ssvd()`, and `fpca2s()`, all of which are compatible with `refund.shiny`. The number of functional principal components (FPCs) is chosen by percent variance explained, with the default set to 0.99. See `?plot_shiny` for examples. Graphics for FPCA are implemented by the code below:

```
> fit.fpca = fpca.sc(Y = DTI$caca)

> plot_shiny(obj = fit.fpca)
```

Executing this code produces a user interface with five tabs. The first tab shows

$\hat{\mu}(t) \pm \sqrt{\hat{\lambda}_k} \hat{\psi}_k(t)$, and includes a drop-down menu through which the user can select k (an example for a similar tab, based on multilevel data, is shown in Section 3). The second tab presents static scree plots of the eigenvalues $\hat{\lambda}_k$ and the percent variance explained by each eigenvalue. The third tab shows $\hat{\mu}(t) + \sum_{k=1}^K c_k \hat{\psi}_k(t)$, and includes slider bars through which the values of c_k can be set; adjusting the sliders allows the user to see a fitted curve for a

hypothetical subject with the selected combination of scores. The fourth tab allows users to assess quality-of-fit by plotting fitted and observed values for any subject in the dataset.

The fifth tab for the interactive graphic produced by the code above is shown as a static plot in Figure 1. A scatterplot of estimated FPC loadings \hat{c}_{ik} against $\hat{c}_{ik'}$ is shown in the upper plot, and k and k' are selected using drop-down menus at the left. The lower plot shows fitted curves for all subjects. In the scatterplot, a subset of FPC loadings can be selected by clicking-and-dragging to create a blue box; blue curves in the plot of fitted values correspond to selected subjects in upper plot. In Figure 1 the first and second FPCs are selected for the x and y axes of the score plot, respectively, and several subjects that have negative values for FPC 1 are highlighted. Fitted values for these subjects are clustered at the top of the y -axis, indicating that the first FPC largely represents a vertical shift from the mean. A working example of `refund.shiny` for FPCA on a different dataset is available at <https://jeff-goldsmith.shinyapps.io/FPCA>.

3 Multilevel Functional Principal Components Analysis

Multilevel functional principal component analysis (MFPCA) extends the ideas of FPCA to functional data with a multilevel structure.

3.1 MFPCA Model

Multilevel functional data are increasingly common in practice; in the case of our DTI example, this structure arises from multiple clinical visits made by each subject. MFPCA models the within-subject correlation induced by repeated measures as well as the between-subject correlation modeled by classic FPCA. This leads to a two-level FPC decomposition, where level 1 concerns subject-specific effects and level 2 concerns visit-specific effects. Population-level basis functions and subject-specific scores are calculated for both levels (Di et al., 2009, 2014). The MFPCA model is:

$$Y_{ij}(t) = \mu(t) + \eta_j(t) + \sum_{k_1=1}^{K_1} c_{ik_1}^{(1)} \psi_{k_1}^{(1)}(t) + \sum_{k_2=1}^{K_2} c_{ijk_2}^{(2)} \psi_{k_2}^{(2)}(t) + \varepsilon_{ij}(t) \quad (2)$$

where $\mu(t)$ is the population mean, $\eta_j(t)$ is the visit-specific shift from the overall mean, $\psi_k^{(1)}(t)$ and $\psi_k^{(2)}(t)$ are the eigenfunctions for levels 1 and 2, respectively, and $c_{ij}^{(1)}$ and $c_{ijk}^{(2)}$ are the subject-specific and subject-visit-specific scores. Often, visit-specific means $\eta_j(t)$ are not of interest and can be omitted from the model. Estimation for MFPCA extends the approach for FPCA: estimated between- and within-covariances $\hat{\Sigma}^{(1)}(s, t) = \widehat{\text{Cov}}(Y_{ij}(s), Y_{ij'}(t))$ for j and $\hat{\Sigma}^{(2)}(s, t) = \widehat{\text{Cov}}(Y_{ij}(s), Y_{ij}(t))$ are derived from the observed data, smoothed, and decomposed to obtain eigenfunctions and values. Given these objects, scores are estimated in a mixed-model framework.

3.2 Graphics for MFPCA

MFPCA is implemented in the `mfpca.sc()` function from the `refund` package. By default, `mfpca.sc` does not calculate visit-means, but they can be calculated by specifying the `mfpca.sc()` argument `twoway = TRUE`.

Graphics for MFPCA are implemented by the code below:

```
> Y = DTI$cca
> id = DTI$ID
> fit.mfpca = mfpca.sc(Y = Y, id = id, twoway = FALSE)
> plot_shiny(fit.mfpca)
```

This code produces an interface with five tabs, which is similar to the interface for FPCA but includes features unique to multilevel analyses. Tabs 1, 2, 3, and 5 for MFPCA are

$\hat{\mu}(t) \pm \sqrt{\hat{\lambda}_{k_L}^{(L)} \hat{\psi}_{k_L}^{(L)}(t)}$, static scree plots of the estimated eigenvalues $\hat{\lambda}_{k_L}^{(L)}$, $\hat{\mu}(t) + \sum_{k_L=1}^{K_L} c_{k_L}^{(L)} \hat{\psi}_{k_L}^{(L)}(t)$, and scatterplots of FPC scores (similar to Figure 1), respectively.

These mirror the tabs for FPCA and include inset sub-tabs to toggle between level, L , to display results for level 1 or level 2. The fourth tab plots fitted and observed values for any user-selected subject in the dataset; the user can display all visits for the selected subject or choose a subset of visits. The first tab for the interactive visualization produced by the code above is displayed in Figure 2, and shows $\hat{\mu}(t) \pm \sqrt{\hat{\lambda}_2^{(1)} \hat{\psi}_2^{(1)}(t)}$.

4 Time-varying Functional Principal Component Analysis

Time-varying functional principal component analysis (TV-FPCA) extends the ideas of FPCA to model functional data that are observed repeatedly in a longitudinal framework. In contrast to MFPCA, TV-FPCA accounts for the actual time of visit T_{ij} at which the functional object $Y_{ij}(\cdot)$ is recorded; this allows to study the time-varying behavior of the underlying true process and make prediction of full trajectory at an unobserved visit time (Park and Staicu, 2015). Other modeling methods for longitudinal functional data that incorporate the actual visit times T_{ij} include Greven et al. (2010) and Chen and Müller (2012).

4.1 TV-FPCA Model

TV-FPCA (Park and Staicu, 2015) model for $Y_{ij}(t) = Y_i(t, T_{ij})$ is given as follows:

$$Y_{ij}(t) = \mu(t, T_{ij}) + \sum_{k=1}^K c_{ij}(T_{ij}) \psi_k(t) + \varepsilon_{ij}(t), \quad (3)$$

where $\mu(t, T_{ij})$ is the population mean that is assumed to vary smoothly over t and visit time T_{ij} ; $\psi_k(t)$ are orthogonal basis functions, $c_{ik}(T_{ij})$ are corresponding loadings that vary over T_{ij} with mean zero and variance λ_k , and $\varepsilon_{ij}(t)$ are residual curves. The time-varying scores $c_{ik}(t_{ij})$ are uncorrelated over i , but correlated over j . Estimation of the TV-FPCA model components entails: 1) estimation of the population mean by using bi-variate smoothing, 2) estimation of the marginal covariance $\Sigma(s, t) = \int \text{Cov}\{Y_\lambda(s, T), Y_\lambda(t, T)\}g(T) dT$, where $g(T)$ is the density of the T_{ij} 's using the observed data, smoothing and decomposing it to get the eigenfunctions/eigenvalues $\hat{\psi}_k(t)$ and $\hat{\lambda}_k$; 3) estimation of the k th component covariance $\hat{G}_k(T, T') = \text{Cov}\{c_{ik}(T)c_{ik}(T')\}$. The last step is carried out using either linear random effects, implying $c_{ik}(T) = b_{0i}^{(k)} + b_{1i}^{(k)}T$ or FPCA implying $c_{ik}(T) = b_{ik1}\varphi_{k1}(T) + \dots + b_{ikL_k}\varphi_{kL_k}(T)$. By modeling these longitudinal dynamics, the time-varying coefficient function $c_{ik}(\cdot)$ can be used to predict scores at any longitudinal time T and, as a result, to predict the full response trajectory $Y_\lambda(\cdot, T)$.

4.2 Graphics for TV-FPCA

TV-FPCA is implemented in the `fpca.lfda()` function in the `refund` package. In Section 4.1, we have used t to denote the functional argument for consistency with the rest of the paper; however to maintain consistency with the notations used in Park and Staicu (2015), `plot_shiny()` function for TV-FPCA uses s to denote the functional argument and T to denote the longitudinal time.

Graphics for TV-FPCA are implemented by the code below:

```
> MS <- subset(DTI, case == 1)

> index.na <- which(is.na(MS$cca)); Y <- MS$cca; Y[index.na] <- fpca.sc(Y)
$Yhat[index.na]

> id <- MS$ID

> visit.index <- MS$visit

> visit.time <- MS$visit.time/max(MS$visit.time)

> fit.tfpc <- fpca.lfda(Y = Y, subject.index = id,
+ visit.index = visit.index, obsT = visit.time,
+ LongiModel.Method = 'lme')

> plot_shiny(fit.tfpc)
```

The code produces an interface with two tabs. Tab 1 shows exploratory plots and includes three inset sub-tabs. The first sub-tab, shown in Figure 3, plots the observed curves for any user-selected subject, and includes options to display the observed curves of all subjects in the background and to display the estimated pointwise mean curve, denoted by $m(t)$. The

second sub-tab allows the user to see the longitudinal changes of the observed curves for a user-selected subject i ; a slider bar animates the subject's visit times and highlights the corresponding observed curve in the plot. The last sub-tab shows two plots of the actual visit times T_{ij} : the bottom plot presents static histogram of visit times of all subjects, while the top plot presents all of observed visit times on a horizontal line to help visualize the sparsity of the longitudinal sampling.

Tab 2 shows estimated model components and predictions, and includes 8 inset sub-tabs. Sub-tabs 1 and 2 present static images of the estimated mean surface $\mu(\hat{t}, T)$ and estimated marginal covariance $\Sigma(\hat{s}, t)$. Sub-tabs 3, 4, and 5 illustrate the first step of estimation, and plot estimates of eigenfunctions $\psi_k(\hat{t}), m(t) \pm 2\sqrt{\hat{\lambda}_k}\hat{\psi}_k(t)$, and static scree plots of the estimated eigenvalues $\hat{\lambda}_k$, respectively. Sub-tab 6 shows the estimated covariance of the time-varying loadings $c_{ik}(\cdot)$ for user-specified k . Sub-tab 7 shows the prediction of $c_{ik}(T)$ for any user-selected subject i and component k ; it also has an option of displaying predicted values of $c_{ik}(T)$ for all subjects in the background. Lastly, sub-tab 8 shows the prediction of a full response trajectory $Y_i(\cdot, T)$ for user-selected subject i in animation with change of values across 21 equi-spaced grid of points of T in the range of observed visit times of all subjects.

5 Function-on-Scalar Regression

In many cases, a length p vector of scalar covariates $\mathbf{x}_i = [x_{i1}, \dots, x_{ip}]$ is observed in addition to the function $Y_i(t)$. In these situations, it is often of interest to model the conditional expectation of the functional response as it depends on the scalar predictors; indeed, this problem has been the focus of a large literature (Brumback and Rice, 1998; Guo, 2002; Morris et al., 2003; Morris and Carroll, 2006; Reiss et al., 2010; Scheipl et al., 2015; Goldsmith and Kitago, 2015; Goldsmith et al., 2015).

5.1 FoSR Model

The most common function-on-scalar regression model is

$$Y_i(t) = \beta_0(t) + \sum_{k=1}^p x_{ik} \beta_k(t) + \varepsilon_i(t) \quad (4)$$

where the $\beta_k(t)$ are fixed effects associated with scalar covariates and the $\varepsilon_i(t)$ are residual curves. The coefficients $\beta_k(t)$ are interpreted analogously to coefficients in a (non-functional) multiple linear regression – as the expected change in response for each one unit change in the predictor – with the exception that they, like the outcome, are defined over t . Many estimation and inferential strategies are available for model (4); a popular approach is to expand coefficients $\beta_k(t)$ using a spline basis, which allows one to recast (4) as a traditional linear regression model and focus estimation on a vector of unknown spline coefficients. Our example uses the `bayes_fosr()` function in the `refund` package, which uses a rich cubic B-spline basis and estimates spline coefficients in a Bayesian framework with priors

specified to enforce smoothness in the resulting coefficient functions. Both a Gibbs sampler and a computationally efficient variational approximation are available in `refund`.

5.2 Graphics for FoSR

Graphics for FoSR are implemented by the code below:

```
> DTI = DTI[complete.cases(DTI), ]

> fit.fosr = bayes_fosr(cca ~ pasat + sex, data = DTI)

> plot_shiny(fit.fosr)
```

This code produces a interface with four tabs, each showing plots associated with model 4. The first tab is a plot of the observed data with the option to color curves by a user-selected covariate; this builds intuition analogously to scatterplots for non-functional regression. The second tab shows $\hat{\beta}_0(t) + \sum_{k=1}^p x_k \hat{\beta}_k(t)$, where values of x_k can be set by slider bars for continuous covariates or drop-down menus for categorical covariates; adjusting the sliders or drop-down menus shows the estimated conditional expectation for a specified predictor vector. The third tab, illustrated in Figure 4, shows estimated coefficient functions $\hat{\beta}_k(t)$ with pointwise confidence intervals for the covariate x_k selected in a drop-down menu. The fourth tab is a plot of the residual curves $\varepsilon(t)$ and allows for identification of median and outlying curves by band depth (Lopez-Pintado and Romo, 2009; Sun and Genton, 2011; Sun et al., 2012); the user can also choose to ‘rainbowize by depth’, which colors the curves from the median outward based on depth.

6 Code Structure of the `refund.shiny` Package

We now briefly describe the code infrastructure used to create the `refund.shiny` package.

As indicated in the introduction, the workflow separates visualization from analysis in the following way. First, one analyzes a dataset using a function in the `refund` package. The functions in `refund` take discretely observed functional data as input, perform an analysis, and return an object whose class corresponds to the method used. For example, the `fpca.sc` function returns an object of class `fpca` and the `bayes.fosr` function returns an object of class `fosr`. The primary function in `refund.shiny`, `plot_shiny`, is a generic function whose behavior depends on the class of the object passed as an argument. Because of this structure, the user experience is uniform across a variety of analyses; this also suggests a development strategy for the addition of interactive graphics as new analysis techniques become available. Lastly, by separating the analysis and visualization steps, it is possible for analysis functions developed outside of the `refund` package to return objects of a defined class and thereby take advantage of the plotting capabilities we describe.

The interactive graphics in the `refund.shiny` are built on RStudio's R package `shiny` (RStudio Inc., 2015), which significantly reduces the barriers to producing webpage-style representations of analysis results in R. Other examples of interactive graphics that utilize the `shiny` framework are `shinymethyl` (Fortin et al., 2014) for visualization of high-

dimensional genomic data and `shiny` (Stan Development Team, 2015) for exploring Bayesian models fit using Markov Chain Monte Carlo. In `refund.shiny` the plots within tabs are produced using `ggplot2` (Wickham and Chang, 2015); it is possible to export each plot as a PDF or to save the corresponding `ggplot` object to the user's R workspace for further manipulation.

7 Concluding Remarks

Visualization has long been acknowledged as a central tool in data analysis. For functional datasets, the need for useful graphics is compounded: data are inherently complex, high-dimensional and structured. Although a robust literature for functional data exists and many methods have standard graphical representations, the creation of these graphics is often time consuming. The `refund.shiny` package was developed to ease this process by producing a visualization framework for several common functional data analyses. By leveraging new tools for interactivity, `refund.shiny` responds to user input and actions and, in so doing, can build intuition for analyses in both statisticians and practitioners. The interfaces produced by `refund.shiny` using the `shiny` framework are web applications, rendered locally by a web browser. These applications can be hosted publicly and may, in the spirit of “visuanimations” (Genton et al., 2015), be included as important parts of scientific papers and reports.

We use an analytic workflow that separates modeling from visualization. Doing so allows several methods and implementations to take advantage of the same visualization software; as an example, `fpca.sc()`, `fpca.face()`, `fpca.ssvd()`, and `fpca2s()` implement different methods for FPCA but are all compatible with `plot.shiny()`. This produces an intuitive user experience and leaves open the possibility for future approaches to FPCA or FoSR to use the `refund.shiny` package for visualization with minimal effort. Similarly, this workflow is amenable to the development of interactive visualizations for additional functional data analyses in future iterations of the package.

Acknowledgments

The third author's research was supported partially by National Science Foundation DMS 0454942 and National Institutes of Health grants R01 NS085211 and R01 MH086633. The last author's research was supported in part by Award R01HL123407 from the National Heart, Lung, and Blood Institute and by Award R21EB018917 from the National Institute of Biomedical Imaging and Bioengineering. The MRI/DTI data were collected at Johns Hopkins University and the Kennedy-Krieger Institute.

References

- Brumback B, Rice J. Smoothing spline models for the analysis of nested and crossed samples of curves. *Journal of the American Statistical Association*. 1998; 93:961–976.
- Chen K, Müller HG. Modeling Repeated Functional Observations. *Journal of the American Statistical Association*. 2012; 107:1599–1609.
- Crainiceanu, C.; Reiss, P.; Goldsmith, J.; Gellar, J.; H, H.; McLean, MW.; Swihart, B.; Xiao, L.; Chen, Y.; Greven, S.; Kundu, MG.; Wrobel, J.; Huang, L.; Huo, L.; Scheipl, F. `refund`: Regression with Functional Data. R package version 0.1-13. 2015. URL <http://cran.r-project.org/package=refund>
- Di CZ, Crainiceanu CM, Caffo BS, Punjabi NM. Multilevel Functional Principal Component Analysis. *Annals of Applied Statistics*. 2009; 4:458–488. [PubMed: 20221415]

- Di CZ, Crainiceanu CM, Jank SJ. Multilevel Sparse Functional Principal Component Analysis. *Stat.* 2014; 3:126–143. [PubMed: 24872597]
- Fortin JP, Fertig E, Hansen K. shinyMethyl: interactive quality control of Illumina 450k DNA methylation arrays in R [version 1; referees: 2 approved]. *f1000research.* 2014; 3:175. [PubMed: 25285208]
- Genton MG, Castruccio S, Crippa P, Dutta S, Huser R, Sun Y, Vettori S. Visuanimation in statistics. *Stat.* 2015; 4:81–96.
- Goldsmith J, Bobb J, Crainiceanu CM, Caffo B, Reich D. Penalized Functional Regression. *Journal of Computational and Graphical Statistics.* 2011; 20:830–851. [PubMed: 22368438]
- Goldsmith J, Greven S, Crainiceanu CM. Corrected Confidence Bands for Functional Data using Principal Components. *Biometrics.* 2013; 69:41–51. [PubMed: 23003003]
- Goldsmith J, Kitago T. Assessing Systematic Effects of Stroke on Motor Control using Hierarchical Function-on-Scalar Regression. *Journal of the Royal Statistical Society: Series C.* 2015 To Appear.
- Goldsmith J, Wrobel J. refund.shiny: Interactive plotting for functional data analyses. R package version 0.1. 2015
- Goldsmith J, Zipunnikov V, Schrack J. Generalized multilevel function-on-scalar regression and principal component analysis. *Biometrics.* 2015; 71:344–353. [PubMed: 25620473]
- Greven S, Crainiceanu CM, Caffo B, Reich D. Longitudinal Functional Principal Component Analysis. *Electronic Journal of Statistics.* 2010; 4:1022–1054. [PubMed: 21743825]
- Guo W. Functional mixed effects models. *Biometrics.* 2002; 58:121–128. [PubMed: 11890306]
- Lopez-Pintado S, Romo J. On The Concept of Depth for Functional Data. *Journal of the American Statistical Association.* 2009; 104:486–503.
- Morris JS. Functional Regression Analysis. *Annual Review of Statistics and Its Application.* 2015; 2(1)
- Morris JS, Carroll RJ. Wavelet-based functional mixed models. *Journal of the Royal Statistical Society: Series B.* 2006; 68:179–199.
- Morris JS, Vannucci M, Brown PJ, Carroll RJ. Wavelet-Based Nonparametric Modeling of Hierarchical Functions in Colon Carcinogenesis. *Journal of the American Statistical Association.* 2003; 98:573–583.
- Park S, Staicu AM. Longitudinal functional data analysis. *Stat.* 2015; 4:212–226. [PubMed: 26594358]
- Ramsay, JO.; Silverman, BW. *Functional Data Analysis.* New York: Springer; 2005.
- Reiss PT, Huang L, Mennes M. Fast Function-on-Scalar Regression with Penalized Basis Expansions. *International Journal of Biostatistics.* 2010; 6:Article 28. [PubMed: 21969982]
- RStudio Inc. shiny: Web Application Framework for R. R package version 0.12.2. 2015. URL <http://cran.r-project.org/package=shiny>
- Scheipl F, Staicu AM, Greven S. Functional additive mixed models. *Journal of Computational and Graphical Statistics.* 2015 To Appear.
- Sørensen H, Goldsmith J, Sangalli L. An Introduction with Medical Applications to Functional Data Analysis. *Statistics in Medicine.* 2013; 32:5222–5240. [PubMed: 24114808]
- Staicu AM, Crainiceanu CM, Reich DS, Ruppert D. Modeling functional data with spatially heterogeneous shape characteristics. *Biometrics.* 2012; 68(2):331–343. [PubMed: 22050118]
- Stan Development Team. shinystan: Interactive Visual and Numerical Diagnostics and Posterior Analysis for Bayesian Models. R package version 2.0.1. 2015. URL <http://cran.r-project.org/package=shinystan>
- Strauss, E.; Sherman, E.; Spreen, O. *Compendium of neuropsychological tests: Administration, norms, and commentary.* New York: Oxford University Press; 2006.
- Sun Y, Genton MG. Functional boxplots. *Journal of Computational and Graphical Statistics.* 2011; 20(2)
- Sun Y, Genton MG, Nychka DW. Exact fast computation of band depth for large functional datasets: How quickly can one million curves be ranked? *Stat.* 2012; 1(1):68–74.
- Wickham, H.; Chang, W. ggplot2: An Implementation of the Grammar of Graphics. R package version 1.0.1. 2015. URL <http://CRAN.R-project.org/package=ggplot2>

Yao F, Müller H, Wang J. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*. 2005; 100(470):577–590.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



Figure 1. Screenshot showing tab 5 of the interactive graphics for FPCA. A scatterplot of FPC loadings \hat{c}_{ik} against $\hat{c}_{i k'}$ is shown in the upper plot, and k and k' are selected using drop-down menus at the left. The lower plot shows fitted curves for all subjects. In the scatterplot, a subset of estimated loadings can be selected by clicking-and-dragging to create a blue box; blue curves in the plot of fitted values correspond to selected points in upper plot.

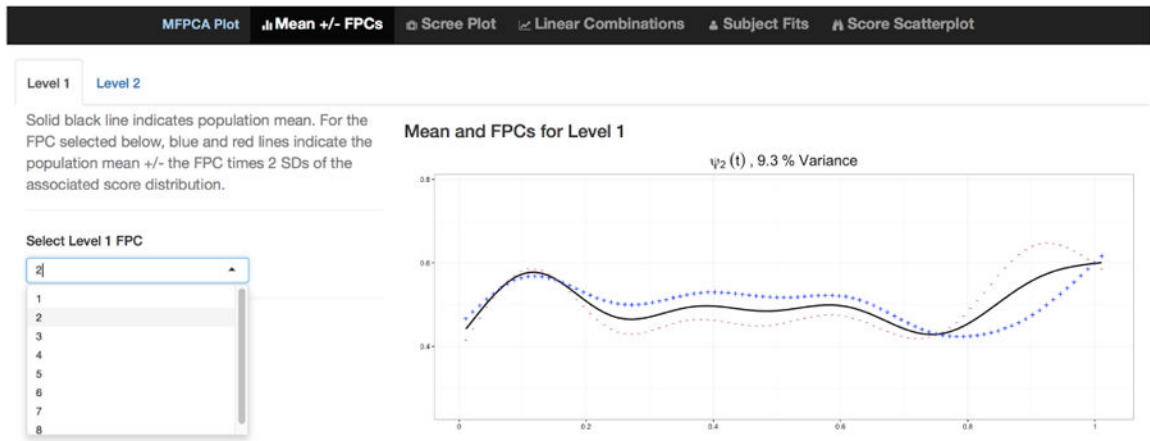


Figure 2. Screenshot showing tab 1 of the interactive graphic for MFPCA. The plot at right shows $\hat{\mu}(t) \pm \sqrt{\hat{\lambda}_{k_L}^{(L)}} \hat{\psi}_{k_L}^{(L)}(t)$; k_L is chosen by the drop-down menu in at left, and the user can switch between level L by clicking *Level 1* or *Level 2* inset tabs at the top left.

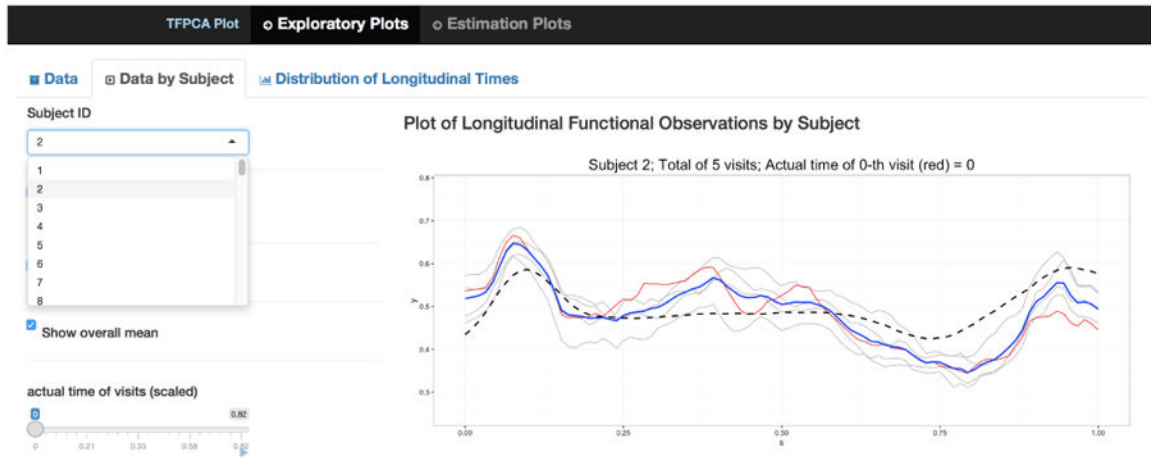


Figure 3. Screenshot showing Tab 1 of the interactive graphic for TV-FPCA. The plot shows observed data of the selected subject.

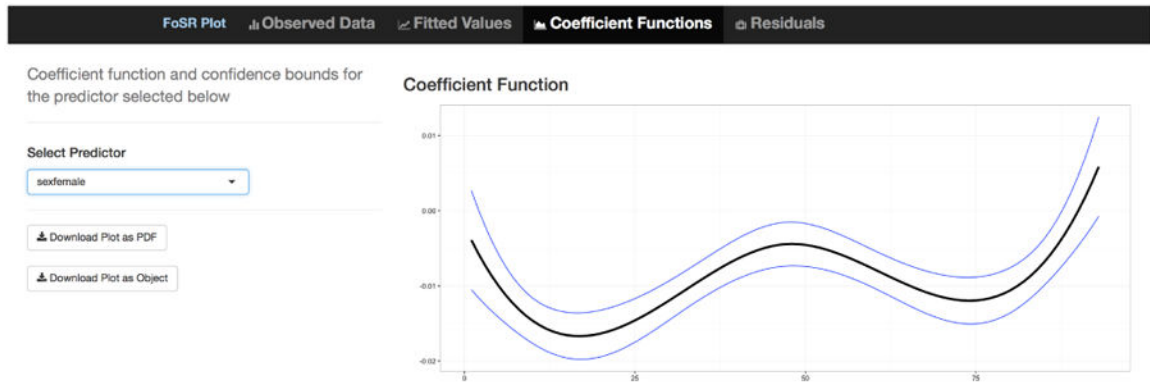


Figure 4. Screenshot showing tab 3 of the interactive graphic for FoSR. The plot shows the estimated coefficient function $\hat{\beta}_k(t)$ for the selected covariate x_k with pointwise confidence intervals.