Explaining Neural Network Predictions for Functional Data Using Principal Component Analysis and Feature Importance

Talk for ISU Graphics Group

September 24, 2020

Katherine Goode

Intern with the Department of Statistical Sciences Sandia National Laboratories

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

SAND no SAND2020-10056 PE

Talk Overview

1. Internship with Sandia National Labs
2. Project Objectives
3. Our Approach
4. Background on FPCA and PFI
5. Application and Visualizations
6. Going Forward

Sandia National Labs

A brief overview and summary of my internship experience

Sandia National Labs

National security is our business. We apply science to help detect, repel, defeat, or mitigate threats.

For more than 70 years, Sandia has delivered essential science and technology to resolve the nation's most challenging security issues.

Sandia National Laboratories is operated and managed by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc. National Technology and Engineering Solutions of Sandia operates Sandia National Laboratories as a contractor for the U.S. Department of Energy's National Nuclear Security Administration (NNSA) and supports numerous federal, state, and local government agencies, companies, and organizations.

As a Federally Funded Research and Development Center (FFRDC), Sandia may perform work for industry responding to certain types of federal government solicitations. The solicitation must allow FFRDC participation and meet the requirements of Sandia's management and operating contract with DOE/NNSA.

National security programs

We work with other government agencies, industry, and academic institutions to accomplish our missions in the following strategic areas:

- Nuclear Weapons
- National Security Programs
- Energy
- Global Security

Computer modeling to understand, track and prepare for COVID-19

National models of COVID-19's spread may be of limited use to state and local planners because the models generally average population data instead of considering the distinctive character of smaller regions. Sandia Labs researchers are using regional data provided by the State of New Mexico to determine the most effective regional distribution of medical supplies and resources and where pop-up testing centers would be best located. Sandia researchers have applied these techniques to other states and localities across the United States.

Continue Reading

Testing for COVID-19

To protect mission-essential functions, keep our workface safe and help reduce the spread of COVID-19 at Sandia facilities, Sandia became the first Department of Energy lab on April 7 to start testing its workforce for COVID-19 infection. The Labs continues to offer on-site swabbing to collect samples for COVID-19 diagnosis of employees and on-site subcontractors who are symptomatic or have come into contact with a presumed or COVID-19 symptomatic case. Swabbing is also provided to employees to meet travel requirements and personnel needing to be...

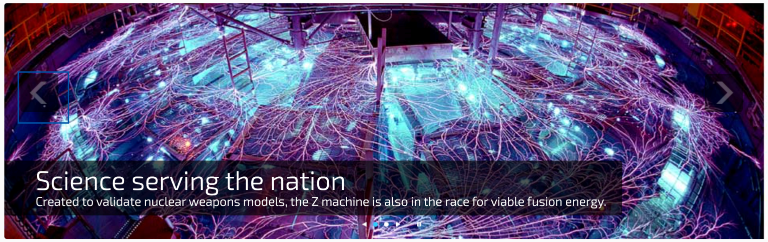
Continue Reading

Manufacturing R&D for personal protective equipment, supplies

Sandia Labs has joined a coalition of national labs aiming to help ensure the nation has access to COVID-19 medical supplies. The coordinated effort is using innovative engineering and technology, such as 3D printing to bridge supply-chain gaps in masks, ventilators and one-time use supplies, such as testing swabs. Examples of this effort are Sandia's method of converting hospital-grade BiPAP machines into ventilators and the labs' help in assessing new decontamination processes to enable the reuse of N95 masks. With an eye to the future, the coalition is...

Continue Reading

Z Pulsed Power Facility



About Z

Sandia's Z machine is the world's most powerful and efficient laboratory radiation source. It uses high magnetic fields associated with high electrical currents to produce high temperatures, high pressures, and powerful X-rays for research in high energy density science. The **Z machine creates conditions found nowhere else on Earth**. Z is part of Sandia's Pulsed Power program, which began in the 1960s.

News

- Four Sandia researchers win Presidential Early Career Award
  August 1, 2019
- Three Sandia researchers elected fellows of

All statements and images from the Sandia website: https://www.sandia.gov/

National Solar Thermal Test Facility

https://twitter.com/sandialabs/status/1123279003853312001?lang=en

https://energy.sandia.gov/programs/renewable-energy/solar-energy/csp-2/nsttf/ - :~:text=Solar Tower,, high-flux materials samples.

Internship (expectations)

Internship (reality)

In all seriousness…

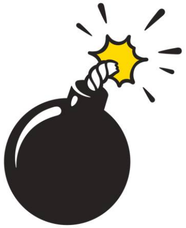
- A really good experience
- Working on research with Danny Ries (PhD in statistics from ISU) and other statisticians, computer scientists, and scientists
- In a group of summer group of mathematics, statistics, and computer science interns
- Able to attend seminars from employees at Sandia

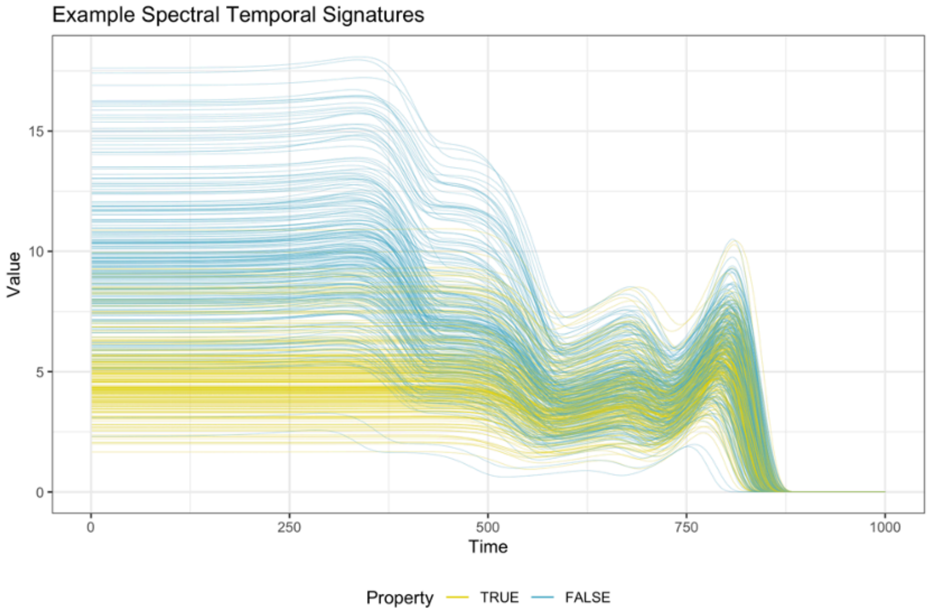
Project Objectives

Predicting explosive device characteristics with machine learning

Main Objective

Use machine learning to predict characteristics of explosive devices based on optical spectral-temporal signatures of explosions



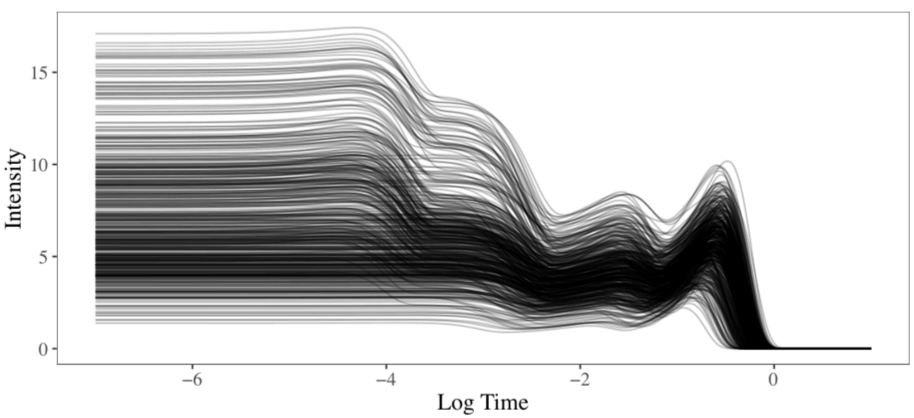


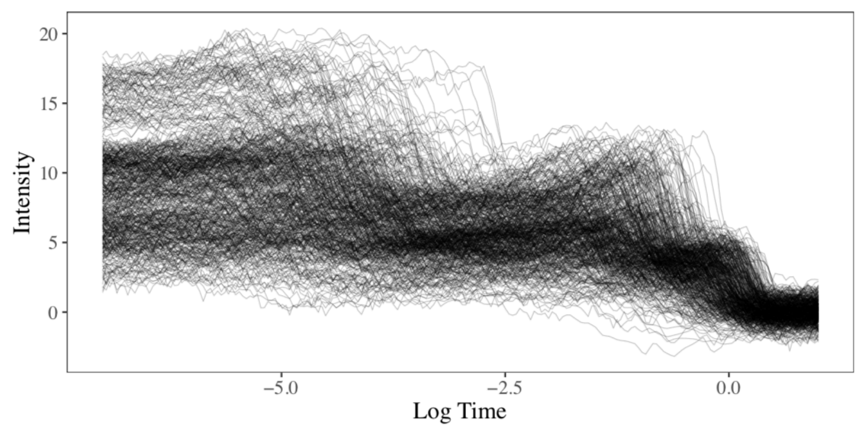
Sub-Objectives

- Return uncertainty quantification
  - Provide an estimate of uncertainty (such as a confidence interval)
  - Use methods such as Bayesian neural networks (BNNs)
- Provide explanations for model predictions
  - Important for providing trust in the model
  - Motivate use of machine learning in this application to decision makers

Data

- 10,000 simulated signatures
- Generated based on scientific understanding of relationship between shape of optical signature and three explosive device characteristics
- Simple and degraded versions of the data
- Example signatures:





Response Variables

Explosive Device Characteristics

Y1:

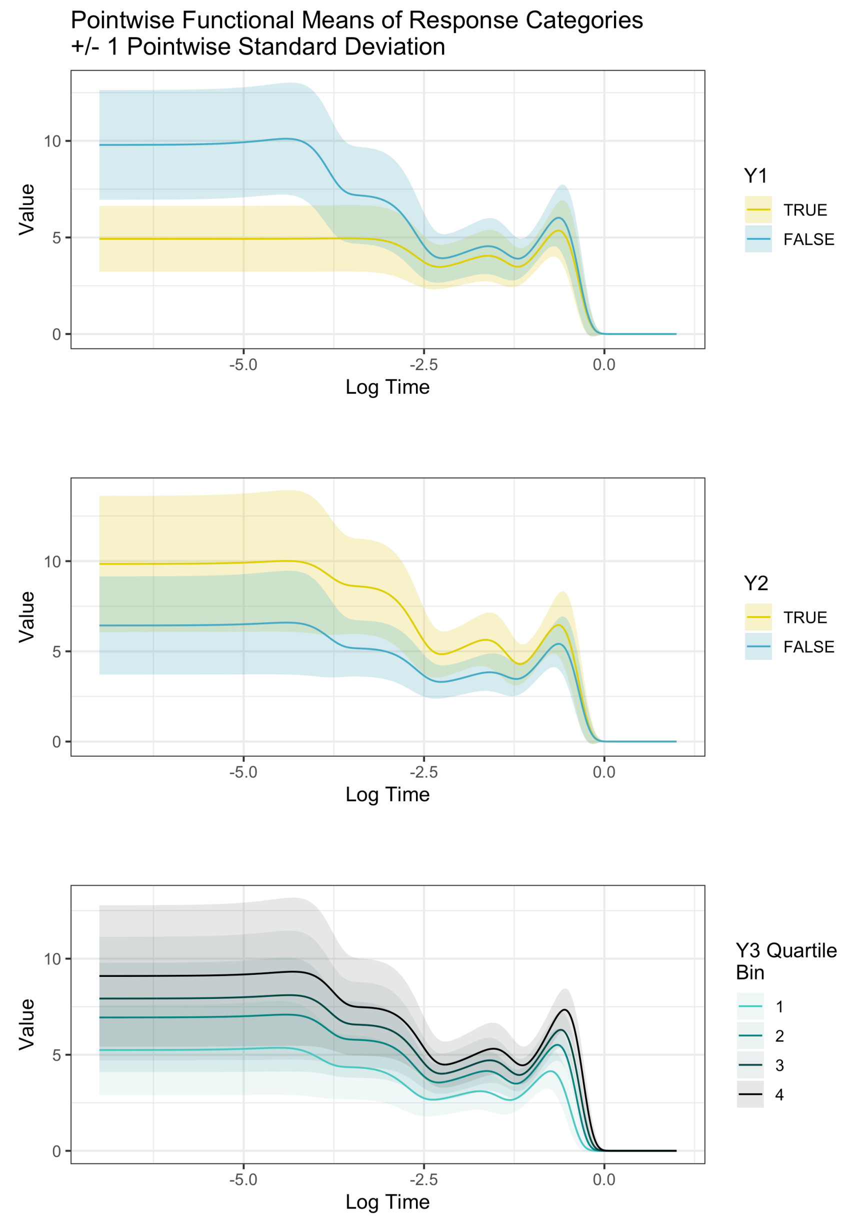
- Binary variable
- Affects intensity of signature early in time and timing of first peak

Y2:

- Binary variable
- Affects intensity of signature over all times

Y3:

- Continuous variable
- Affects intensity and timing of all peaks



Our Approach

Functional PCA and Permutation Feature Importance

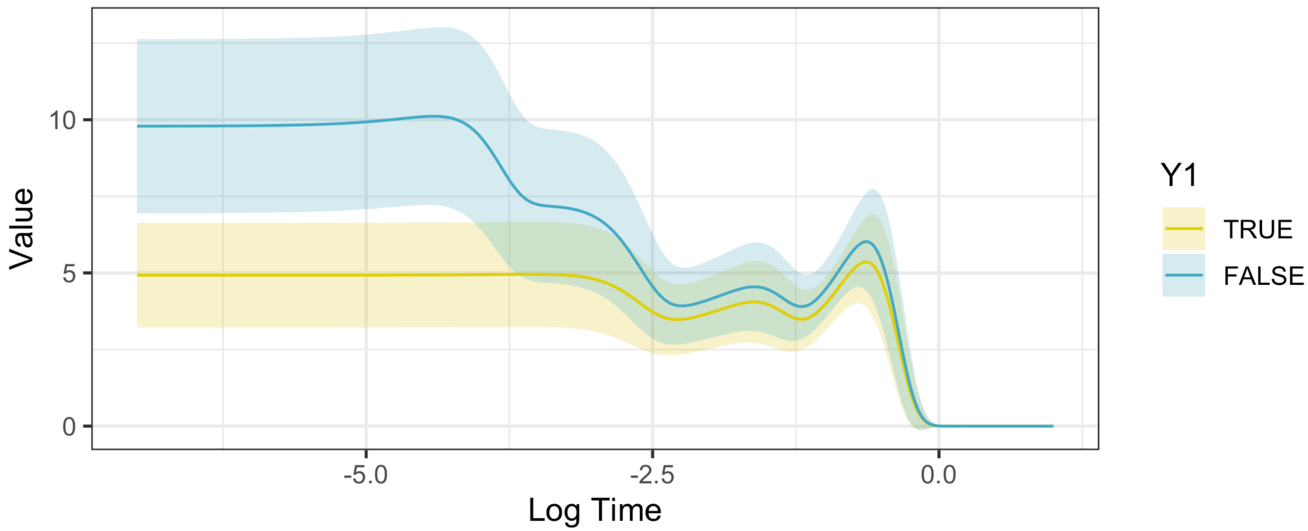
Naïve Approach

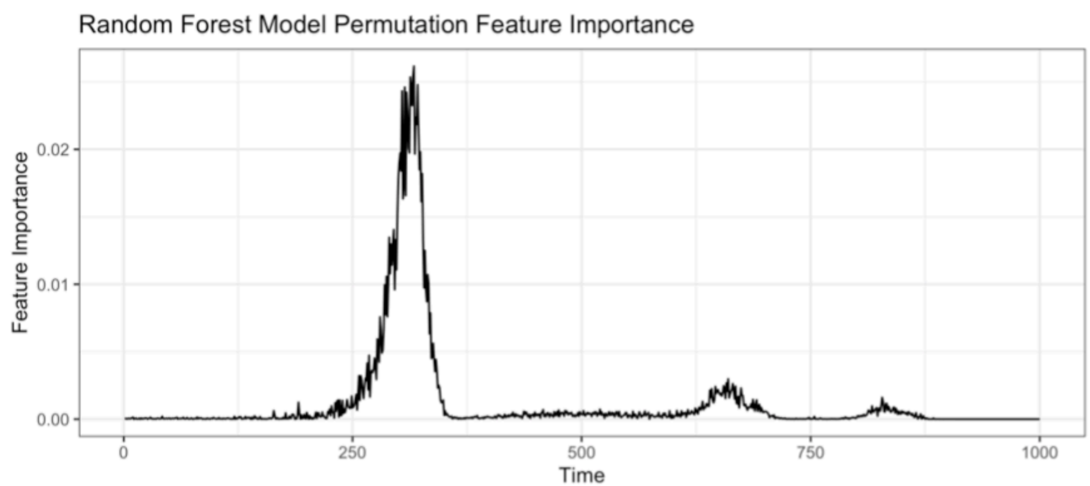
- Use each signature time point as a predictor variable in the model
- Train a random forest, neural network, etc.
- Compute variable importance

Signature	Time 1	Time 2	.....	Time 1000	Y1	Y2	Y3
1	10	9.8	…	0	T	T	82
2	15	15.1	…	0	F	T	71
…	…	…	…	…	…	…	…
10000	8	8.05	…	0	T	F	68

Hypothetical Dataset

Example Feature Importance for Y1

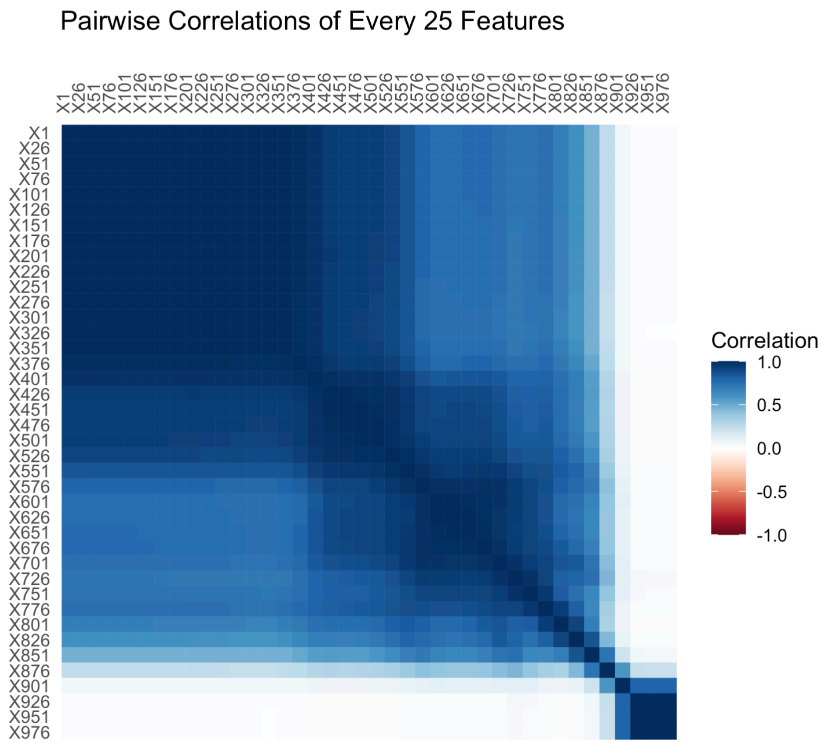




A problem….

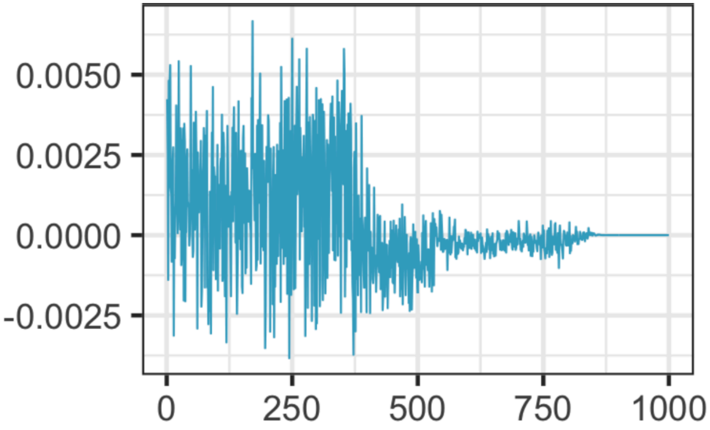
The statistician asks:

"What about the dependence in the signatures over time?"

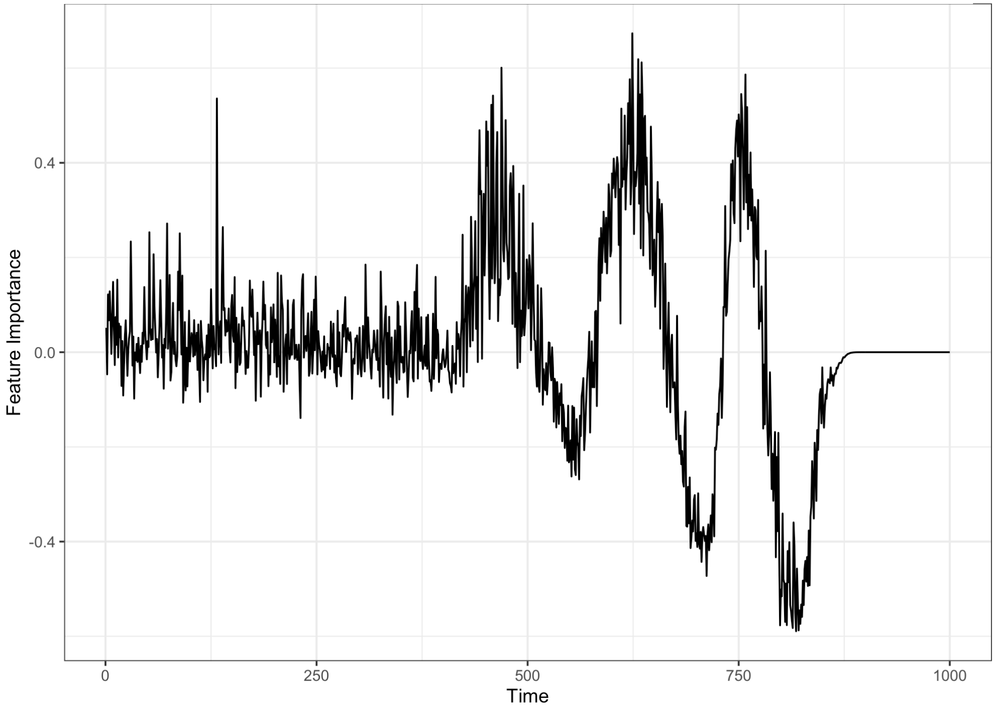


Neural network feature importance
(via the naïve approach)

Permutation Feature Importance for Y1



Permutation Feature Importance for Y3



Our Approach

1. Transform signatures using FPCA (creates uncorrelated features)

2. Fit neural network using principal components (PCs) as features

3. Use permutation feature importance to determine important PCs

4. Interpret PCs using visualizations

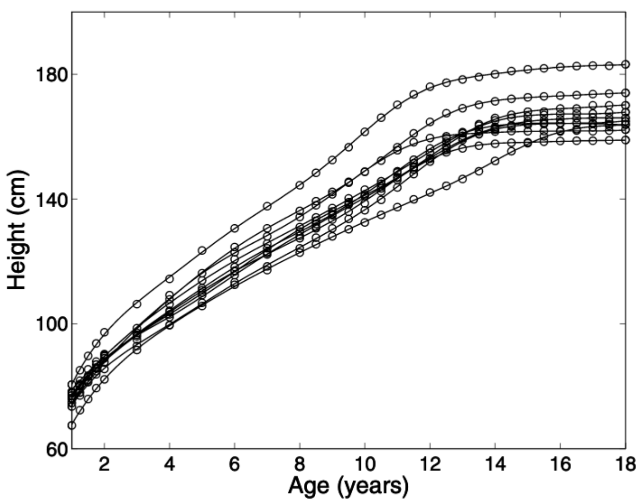
Note: Each of the statistical methods mentioned in the above steps have been developed previously. Our novel approach is using these tools in combination to provide an explanation for neural network predictions when working with functional data.

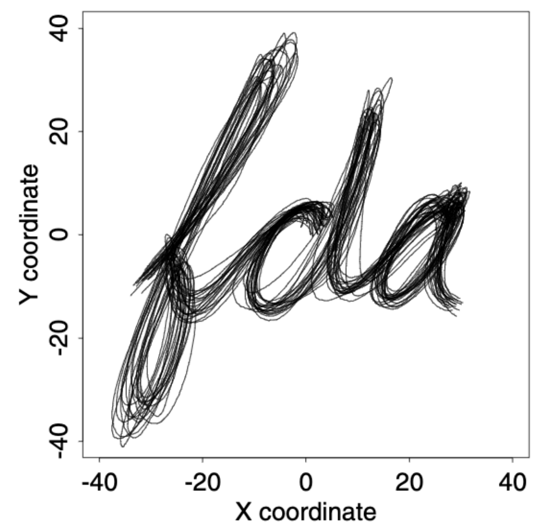
Background on Statistical Methods

Functional Principal Components and Permutation Feature Importance

Functional Data Analysis (FDA)

- Wang, Chiou, and Müller (2015) define:
  - "Functional data analysis (FDA) deals with the analysis and theory of data that are in the form of functions, images and shapes, or more general objects"





Images from Ramsay and Silverman (2005)

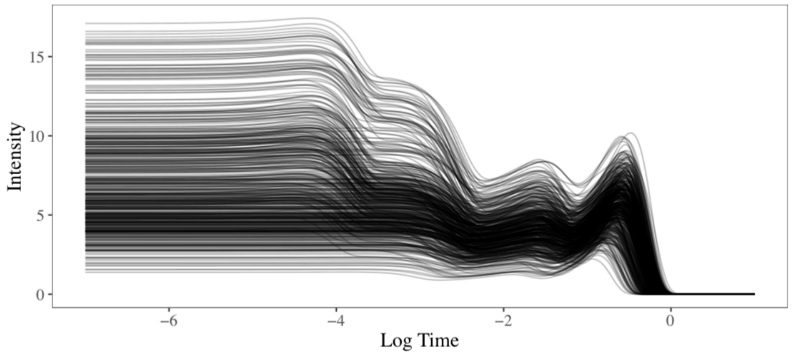
Functional Principal Components Analysis (FPCA)

- FPCA is essentially PCA with "functional data"
- Transforms original functions in a way that provides nice properties:
  - Independent features
  - First few features capture majority of the variation in original data
- Use first few transformed features to fit model

Ways to implement FPCA

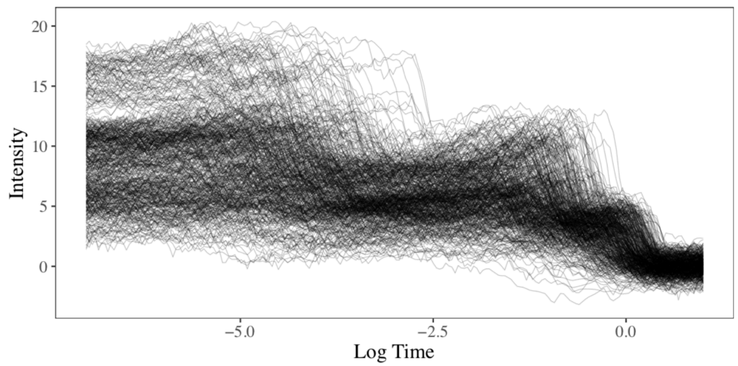
Basic Method:

- Treat each time as a variable and apply basic PCA
- Accounts for vertical variability



More Sophisticated Method:

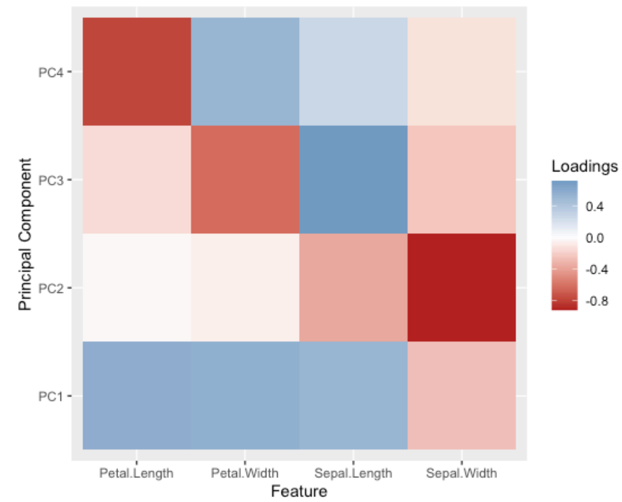
- Joint FPCA
  - [Tucker, Wu, and Srivastava (2012)]
  - [Sungwon Lee and Sungkyu Jung (2017)]
- Accounts for vertical and horizontal variability

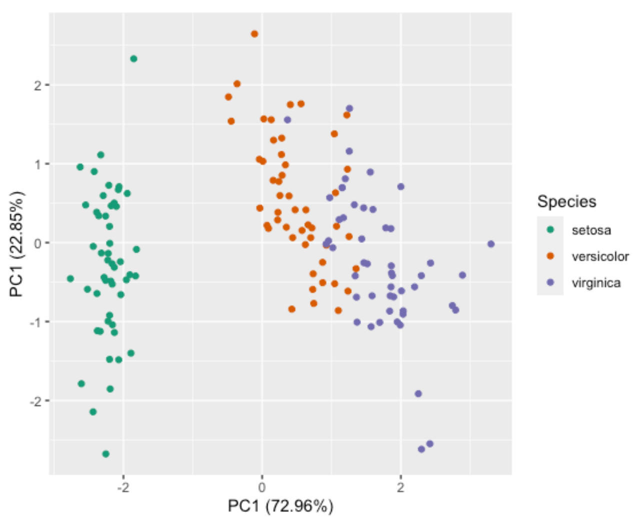


Interpreting and Visualizing PCA

Applying PCA to iris data:

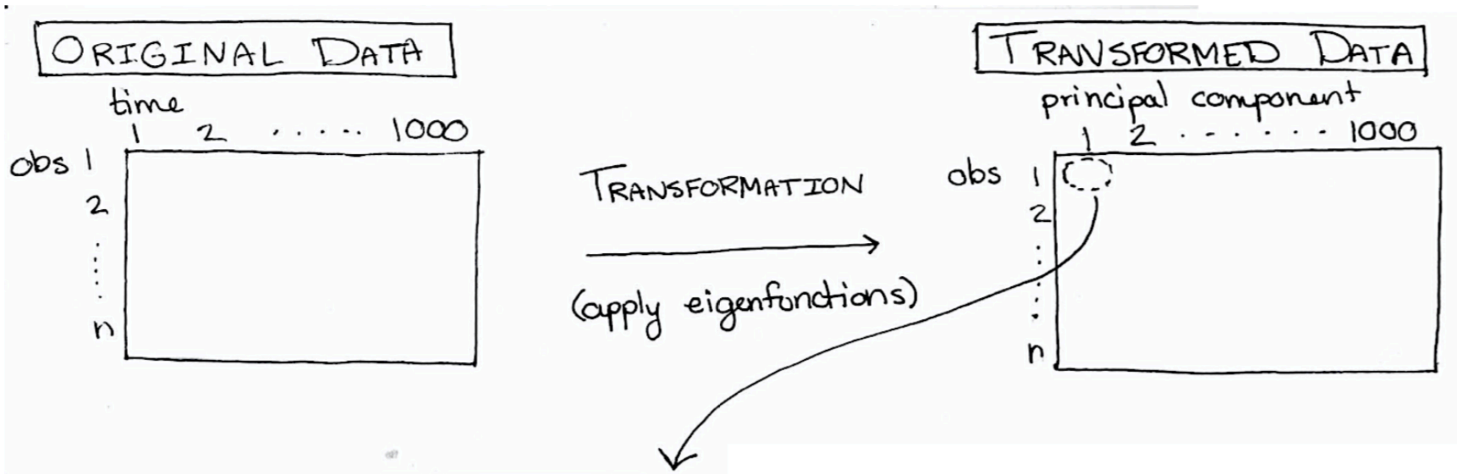
```
> head(iris)
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1         3.5          1.4         0.2  setosa
2          4.9         3.0          1.4         0.2  setosa
3          4.7         3.2          1.3         0.2  setosa
4          4.6         3.1          1.5         0.2  setosa
5          5.0         3.6          1.4         0.2  setosa
6          5.4         3.9          1.7         0.4  setosa
> pca = prcomp(iris[,-5], scale = TRUE)
> head(pca$x)
           PC1        PC2         PC3          PC4
[1,] -2.257141 -0.4784238  0.12727962  0.024087508
[2,] -2.074013  0.6718827  0.23382552  0.102662845
[3,] -2.356335  0.3407664 -0.04405390  0.028282305
[4,] -2.291707  0.5953999 -0.09098530 -0.065735340
[5,] -2.381863 -0.6446757 -0.01568565 -0.035802870
[6,] -2.068701 -1.4842053 -0.02687825  0.006586116
> pca$rotation
                    PC1         PC2        PC3        PC4
Sepal.Length  0.5210659 -0.37741762  0.7195664  0.2612863
Sepal.Width  -0.2693474 -0.92329566 -0.2443818 -0.1235096
Petal.Length  0.5804131 -0.02449161 -0.1421264 -0.8014492
Petal.Width   0.5648565 -0.06694199 -0.6342727  0.5235971
```



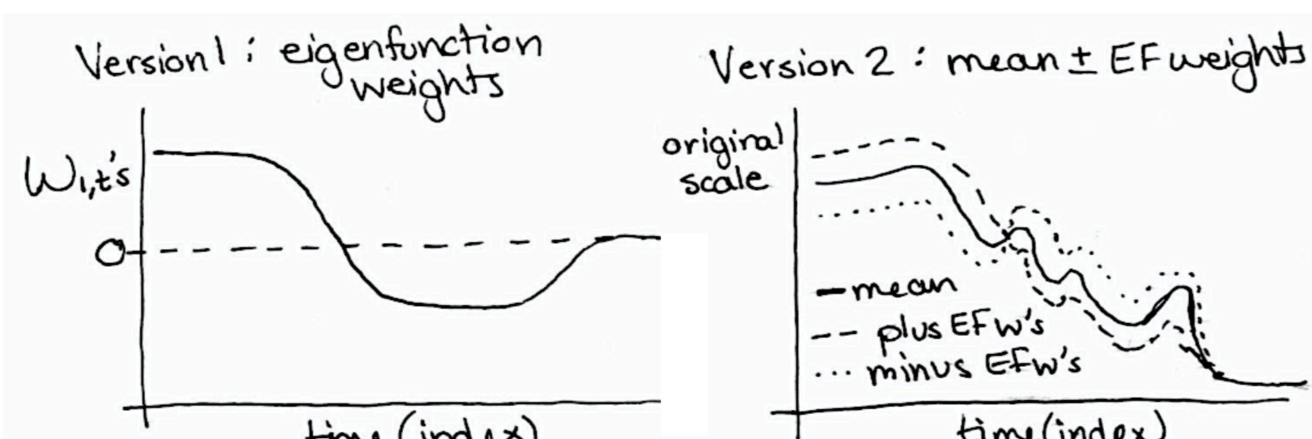


Visualizing FPCA

Eigenfunctions: Functions that transform the original data in FPCA



Eigenfunction 1: $\mathrm{EF1}(\mathrm{obs}\ 1) = (\omega_{1,t1})(\mathrm{obs}_{1,t1}) + (\omega_{1,2})(\mathrm{obs}_{1,t2}) + \cdots + (\omega_{1,1000})(\mathrm{obs}_{1,t1000})$



Visualizing FPCA

Example: Canadian weather temperatures measured over time

Visualizations of eigenfunctions

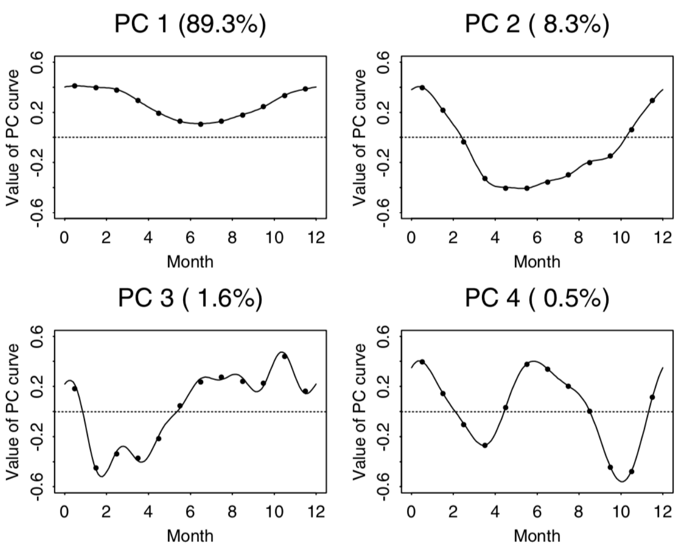


Figure 8.1. The first four principal component curves of the Canadian temperature data estimated by two techniques. The points are the estimates from the discretization approach, and the curves are the estimates from the expansion of the data in terms of a 12-term Fourier series. The percentages indicate the amount of total variation accounted for by each principal component.

Functional mean +/- eigenfunction weights

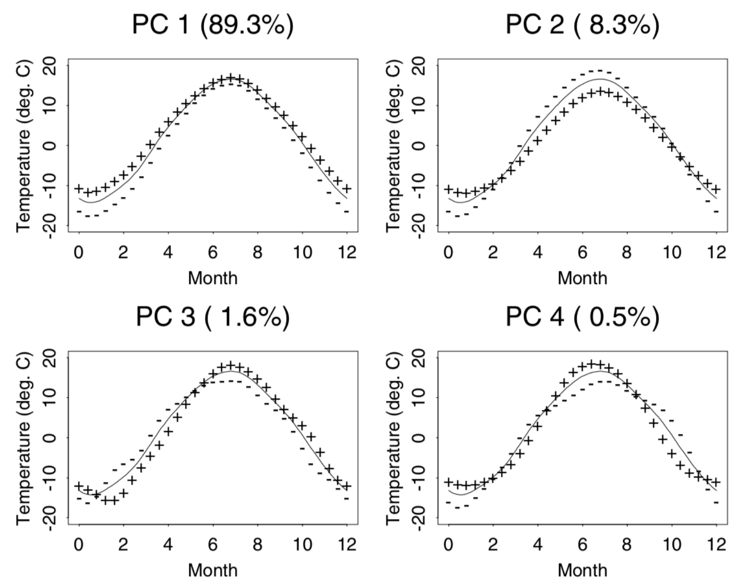


Figure 8.2. The mean temperature curves and the effects of adding (+) and subtracting (−) a suitable multiple of each PC curve.

Images from Ramsay and Silverman (2005)

Visualizing FPCA (multivariate case)

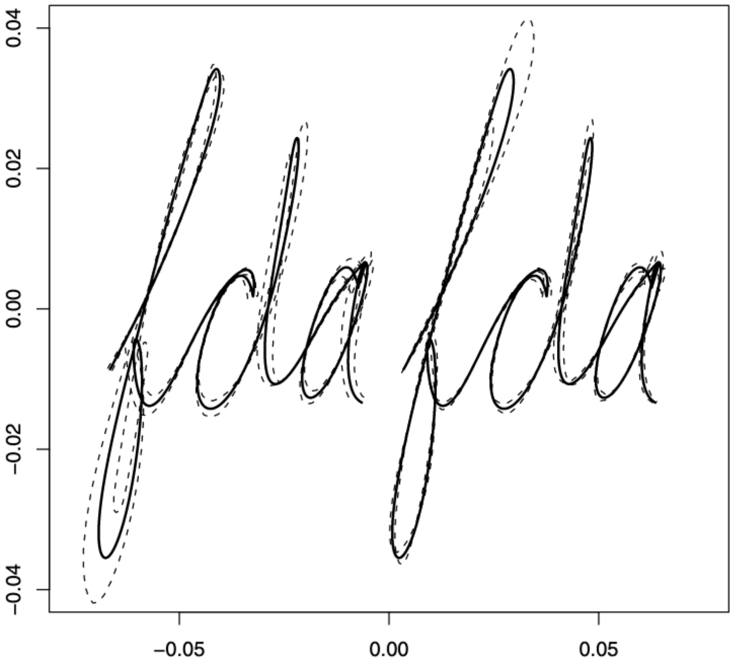


Fig. 7.7 Two of the rotated harmonics are plotted as a perturbations of the mean "fda" script, shown as a heavy solid line.

Image from Ramsay, Hooker, and Graves (2009)

Permutation Feature Importance (PFI)

- Originally developed by Breiman (2001; paper) for random forests

- Breiman developed two types of importance:
  - Node impurity based method
  - Permutation based method

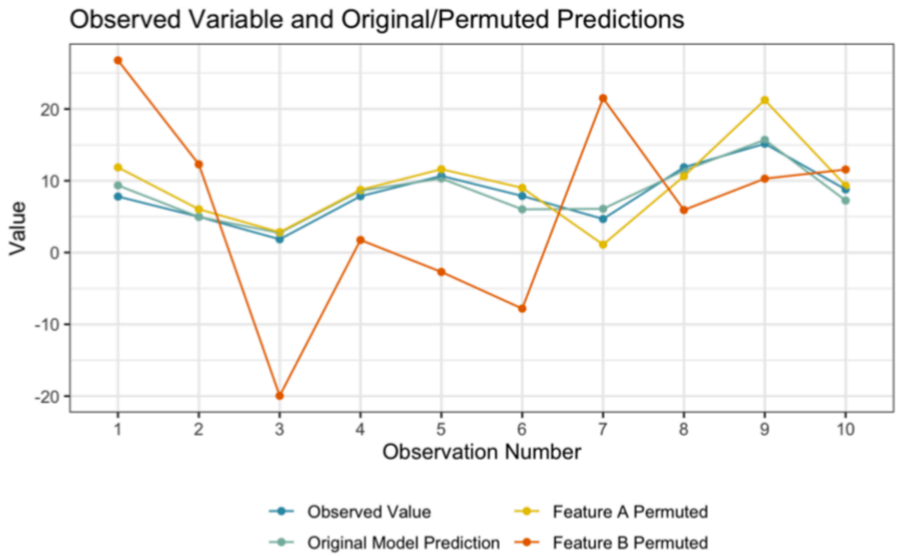
```
> library(randomForest)
> set.seed(71)
> iris.rf <- randomForest(Species ~ ., data=iris, importance=TRUE,
+                          proximity=TRUE)
> round(importance(iris.rf), 2)
             setosa versicolor virginica MeanDecreaseAccuracy MeanDecreaseGini
Sepal.Length   5.88       5.87      9.21                10.62             9.37
Sepal.Width    5.23       0.31      4.71                 4.94             2.45
Petal.Length  21.60      31.41     27.71                32.39            42.13
Petal.Width   22.96      33.74     32.07                33.85            45.28
```

- Generalized random forest version for all by Fisher, Rudin, and Dominici (2019; paper)

PFI: General Idea

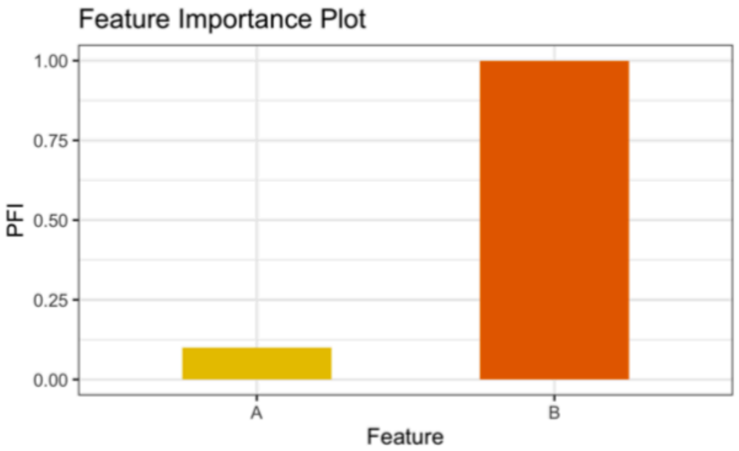
Procedure

- Permute a feature
- Determine how model predictions are affected
- Repeat for all other features
- Repeat to account for random variation



Interpretation

- Positive PFI: Prediction affected
- PFI near 0: Prediction NOT affected
- Negative PFI
  - Prediction improved
  - Feature is counterproductive



PFI: Formula

Define:

- f predictive model
- y response vector
- X training matrix of feature values (rows x_i)
- $X^{\pi,j}$ matrix with feature j of X permuted (rows $x_i^{\pi,j}$)
- $L(y_i, f(x_i))$ loss for predicting y_i from $f(x_i)$

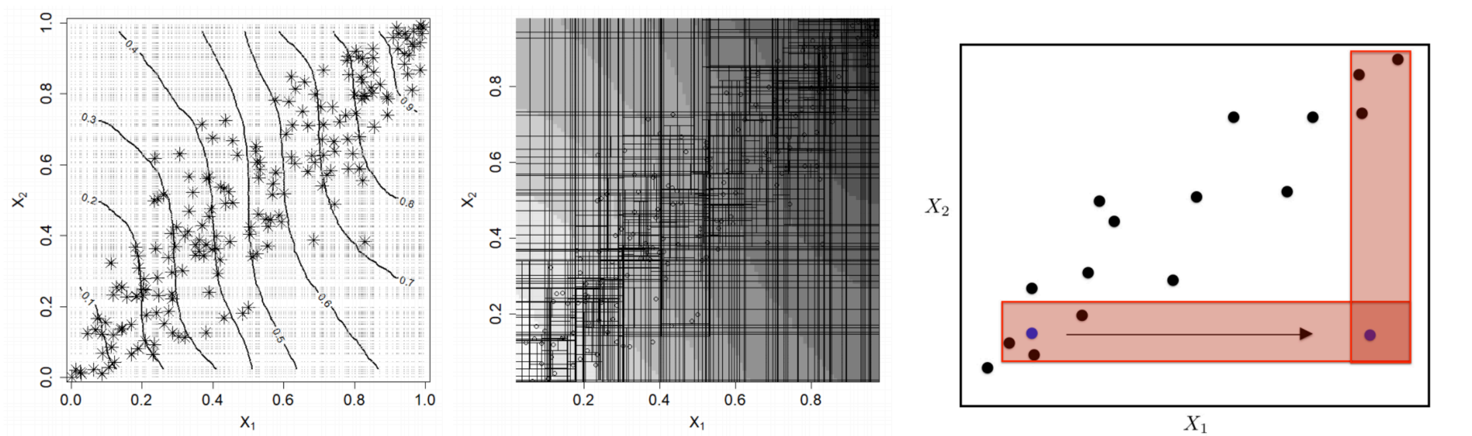
Permutation Feature Importance:

$$VI_j^{\pi} = \sum_{i=1}^{N} L(y_i, f(x_i^{\pi,j})) - L(y_i, f(x_i))$$

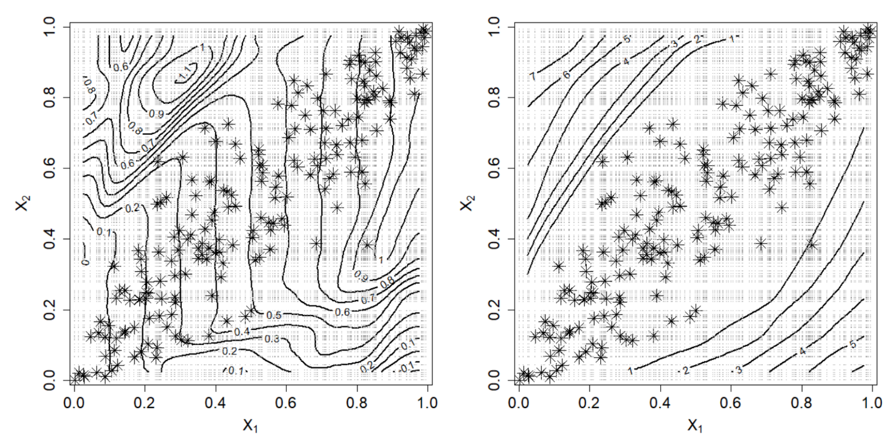
PFI and Correlation

- Research has shown that PFI is biased when correlated variables are involved (Strobl et al. (2007), Archer and Kimes (2008), Nicodemus et al. (2010))
- Hooker and Mentch (2019) attempt to provide an explanation

Random forest



Neural network



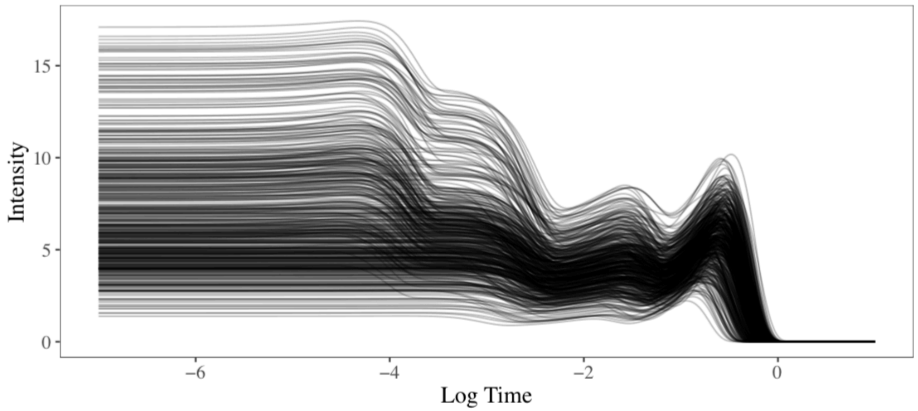
Images from Hooker and Mentch (2019)

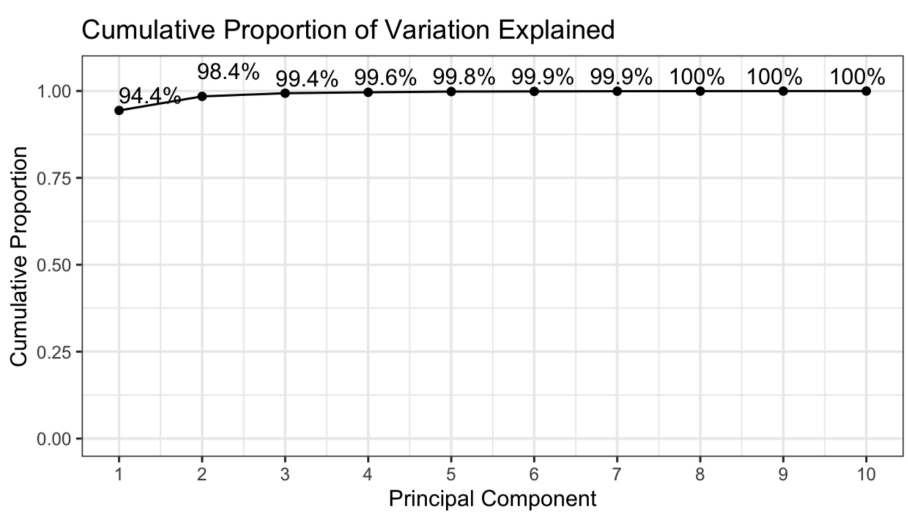
Application and Visualizations

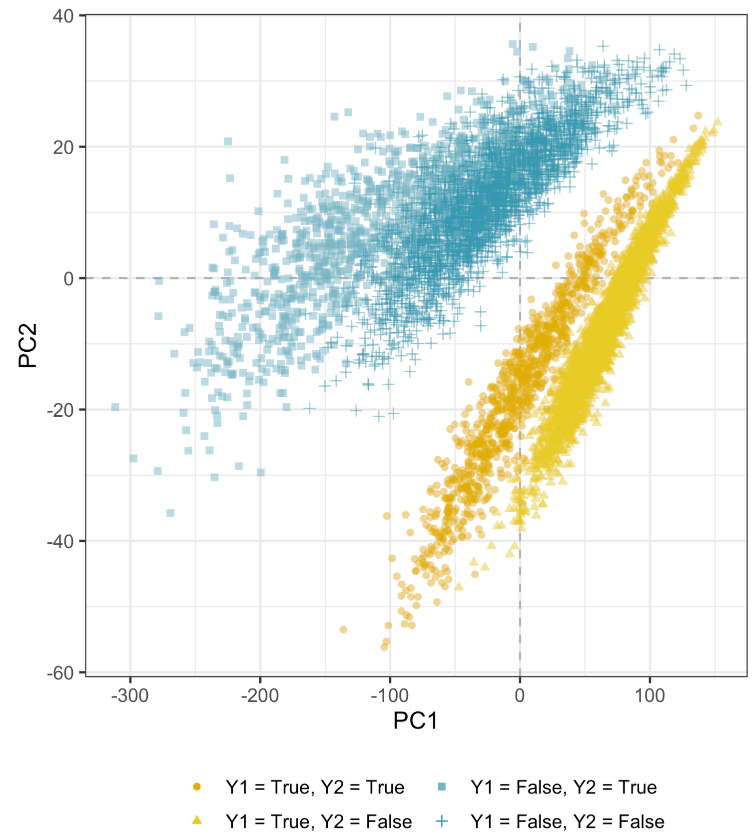
Explaining neural network predictions

Step 1: Transform signatures using FPCA

- Started with basic FPCA approach on the simple data
- Resulted in 1000 FPCs

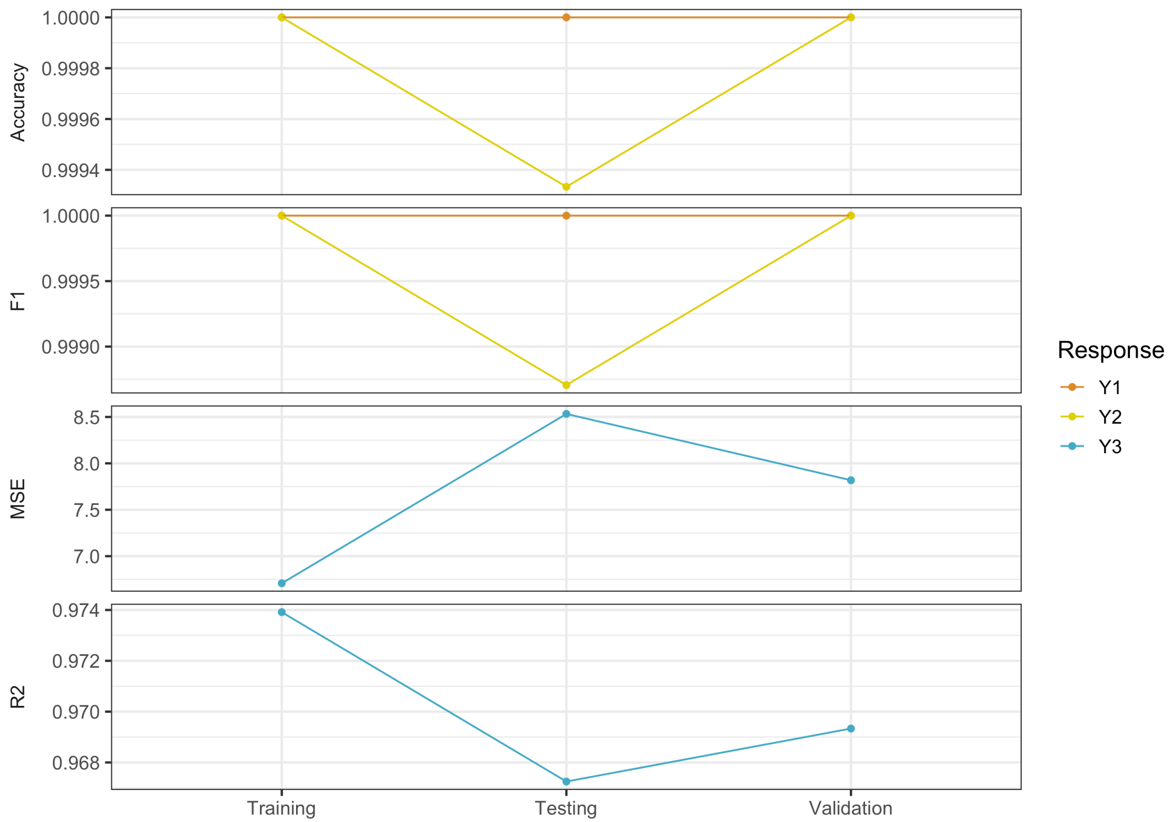




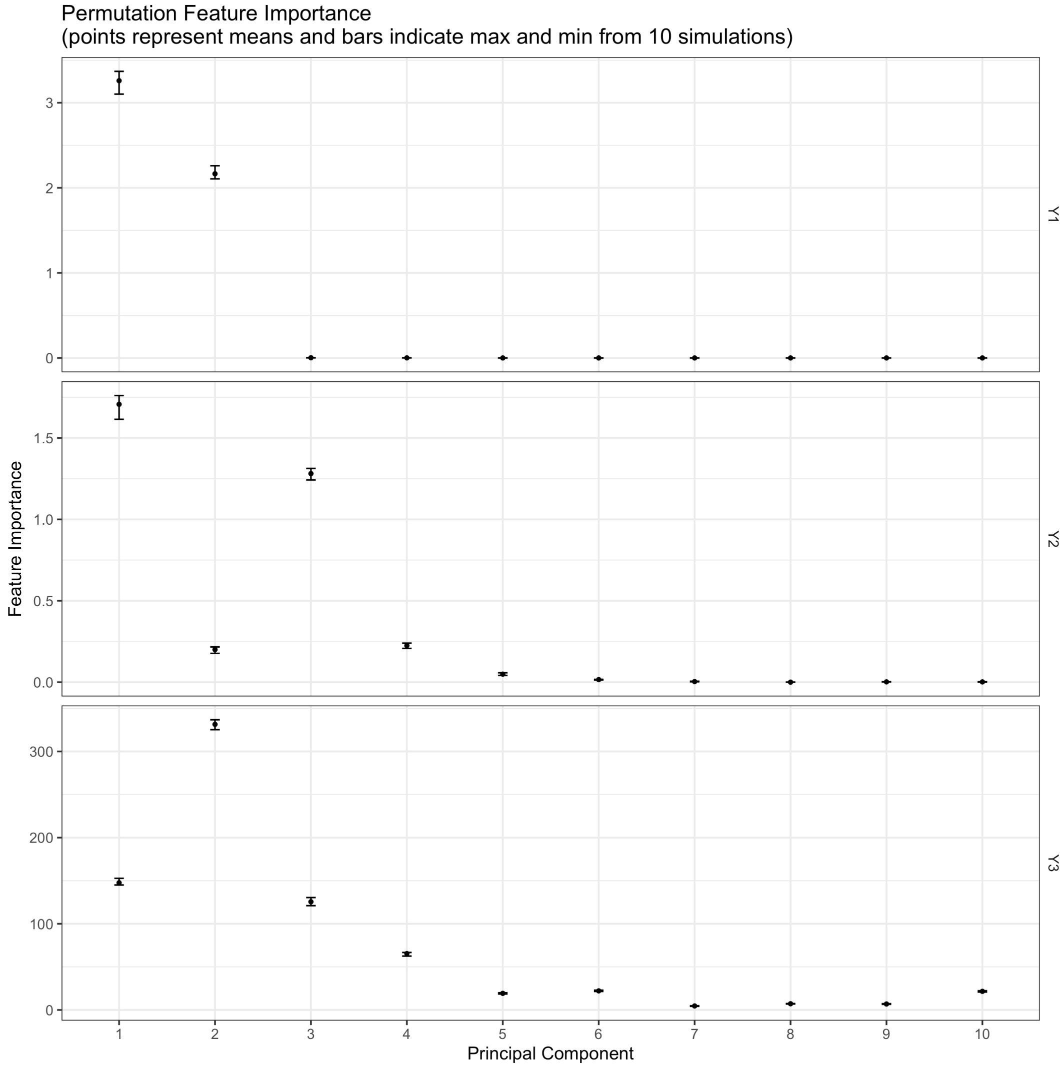


Step 2: Fit neural network using PCs

- Three models (one per characteristic)
- 3 layers (50, 40, and 30 nodes)
- All 1000 PCs as feature

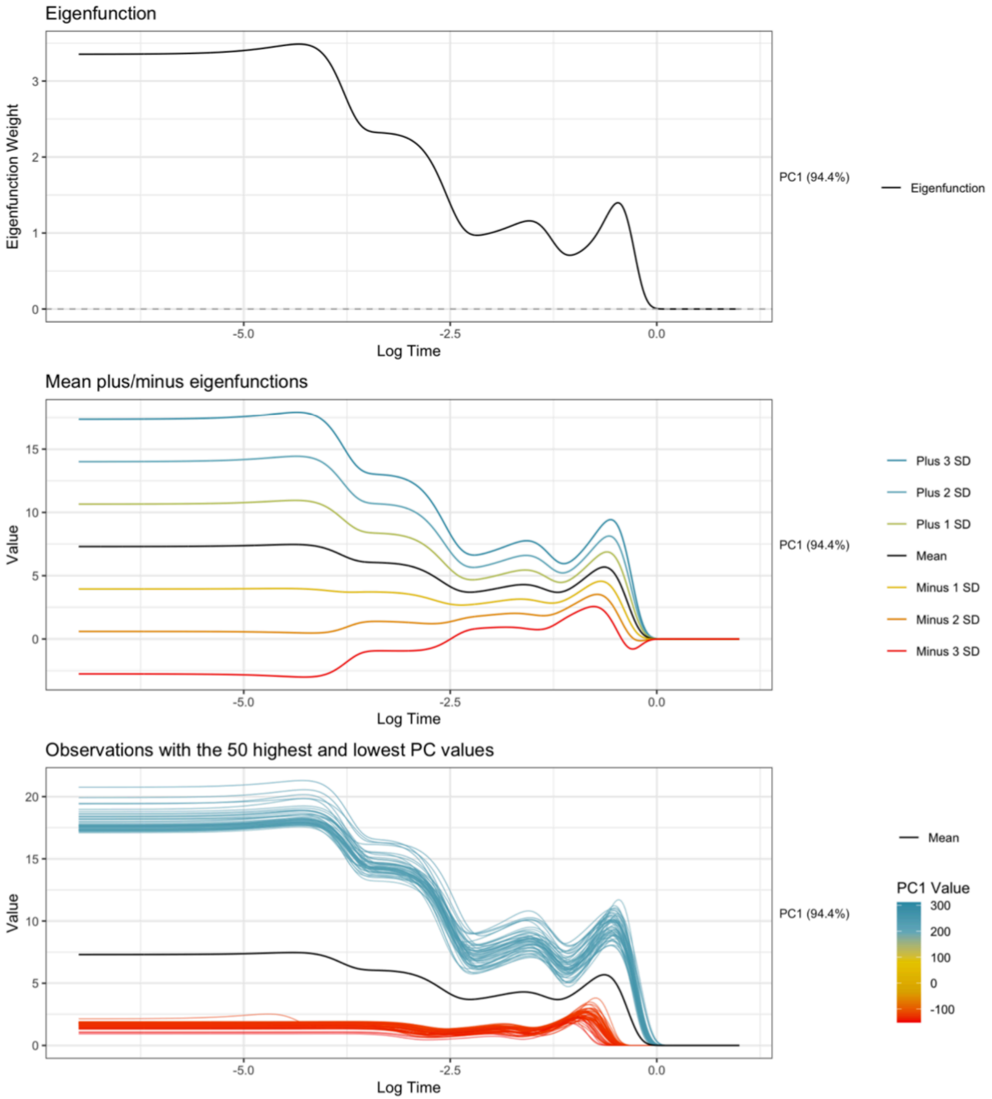


Step 3: Apply permutation feature importance to determine important PCs

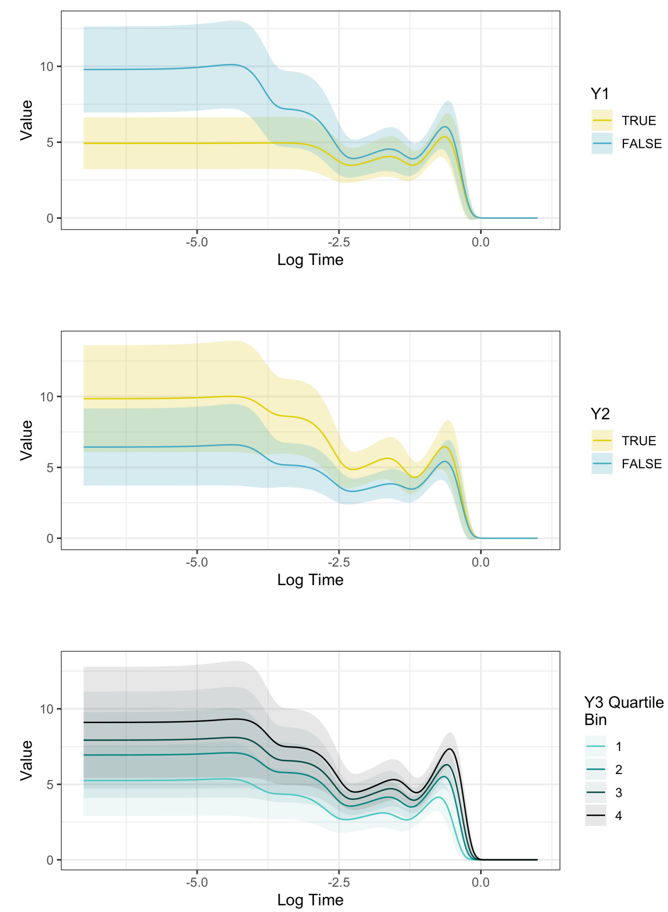


Step 4: Interpret PCs using visualizations

PC 1: Interpretation

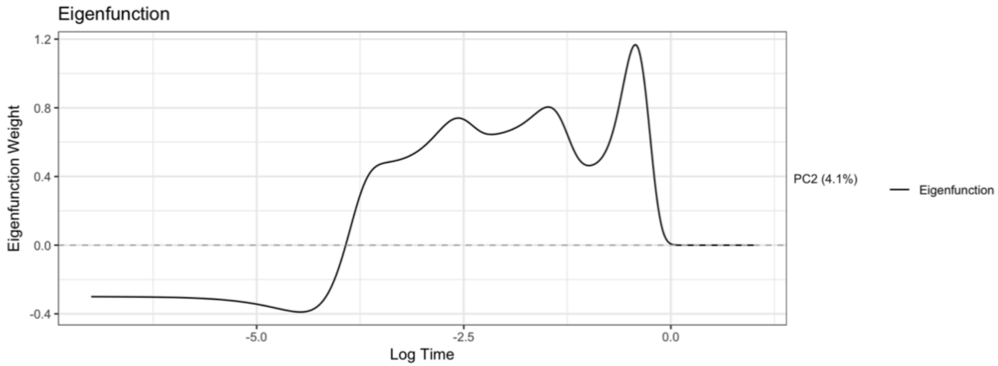


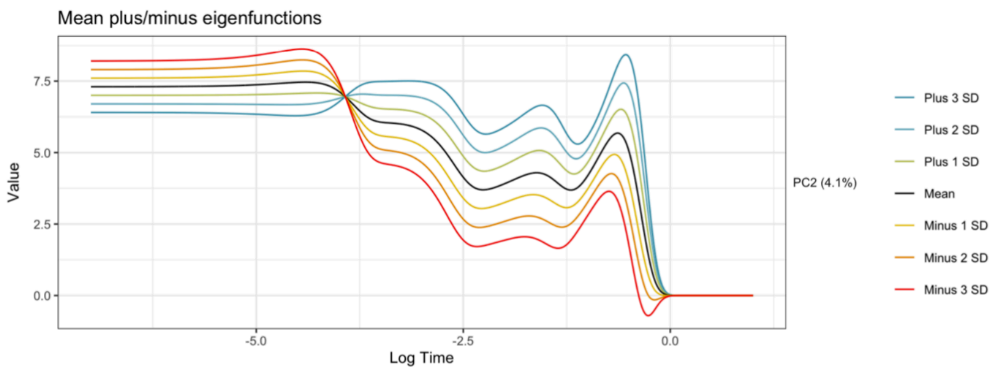
Important for Y1 and Y2
Somewhat important for Y3

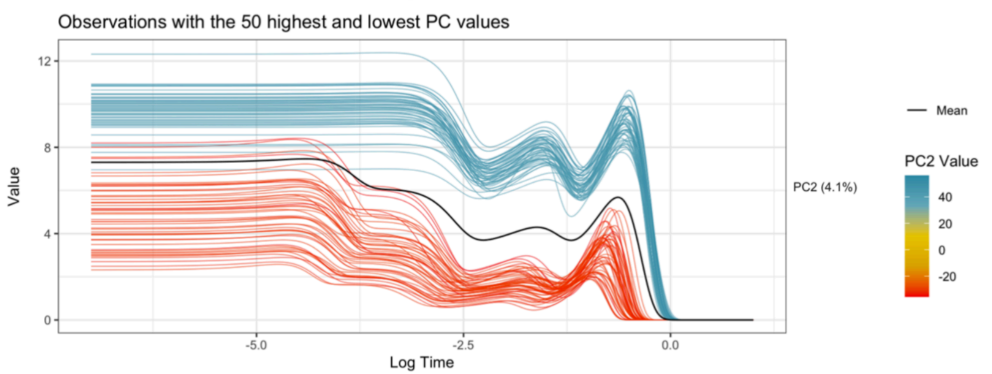


Step 4: Interpret PCs using visualizations

PC 2: Interpretation

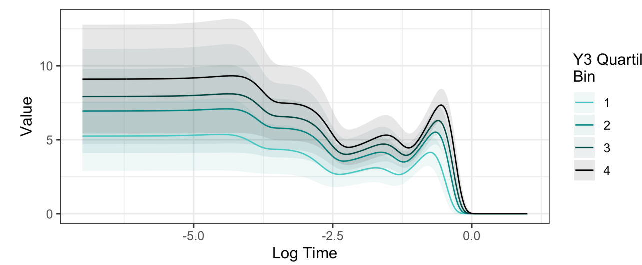


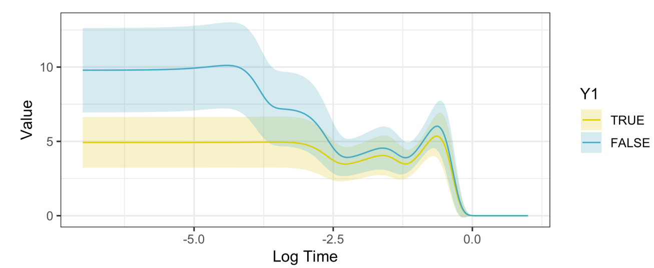




Important for Y3

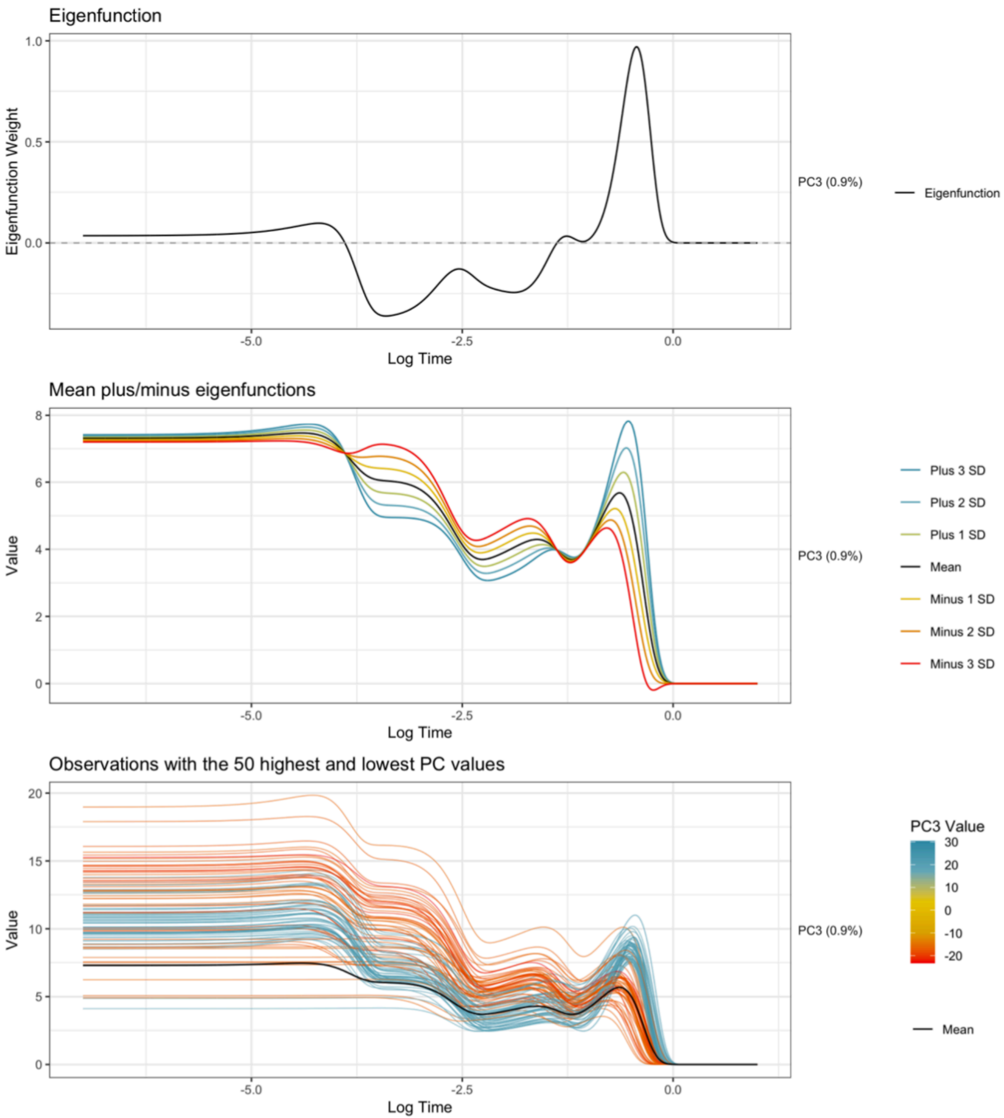
Somewhat important for Y1





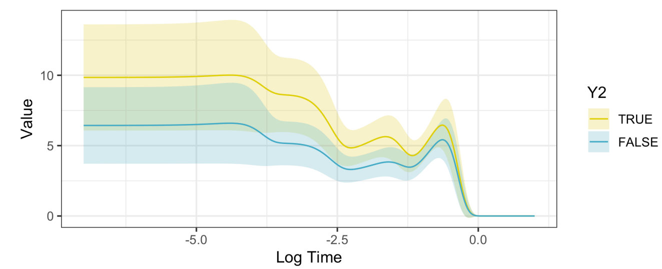
Step 4: Interpret PCs using visualizations

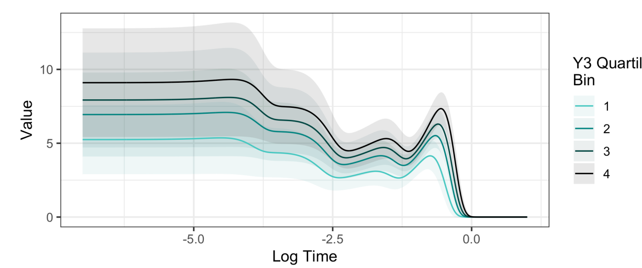
PC 3: Interpretation



Important for Y2

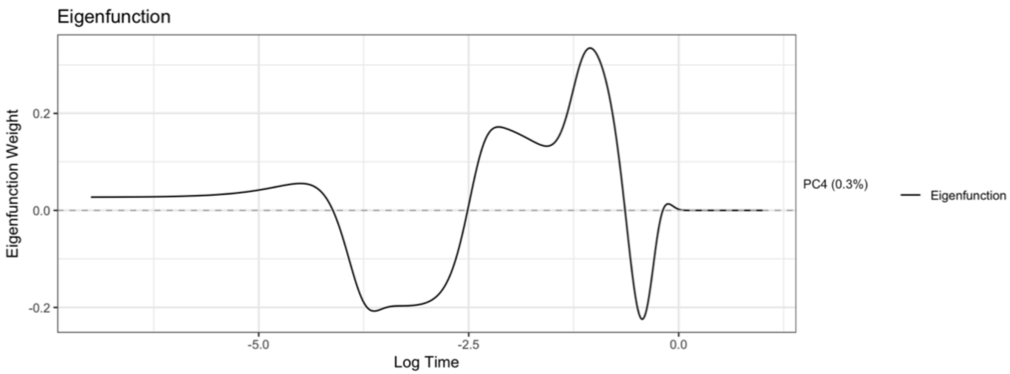
Somewhat important for Y3

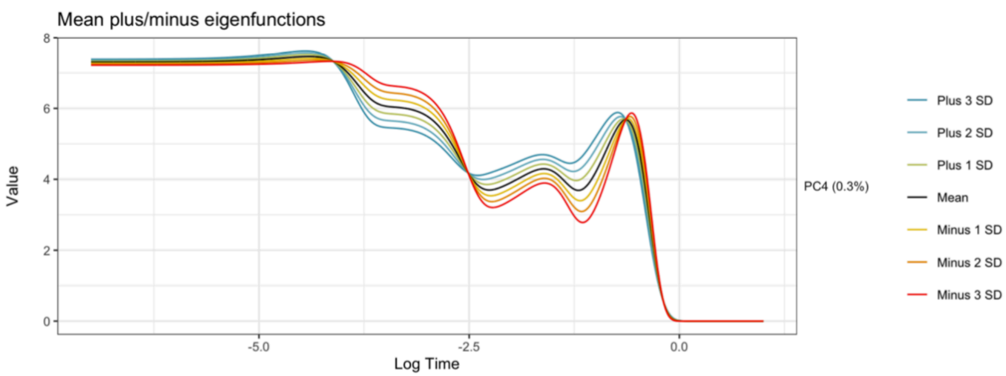


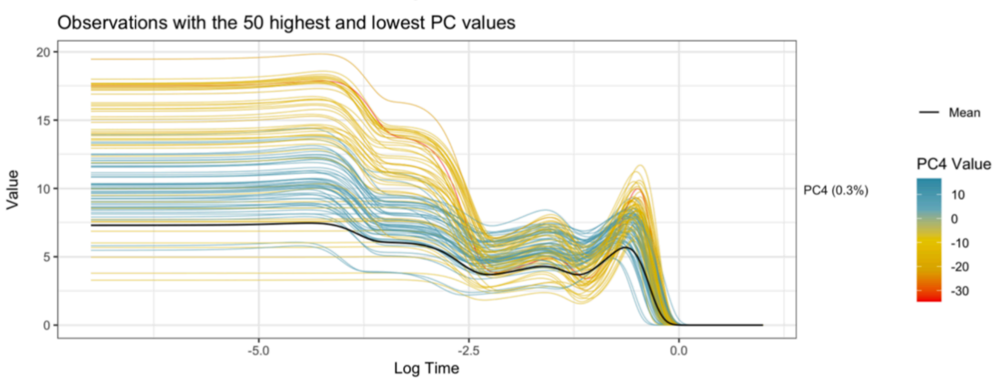


Step 4: Interpret PCs using visualizations

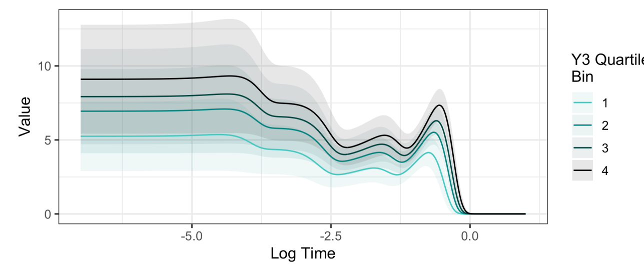
PC 3: Interpretation







Somewhat important for Y3

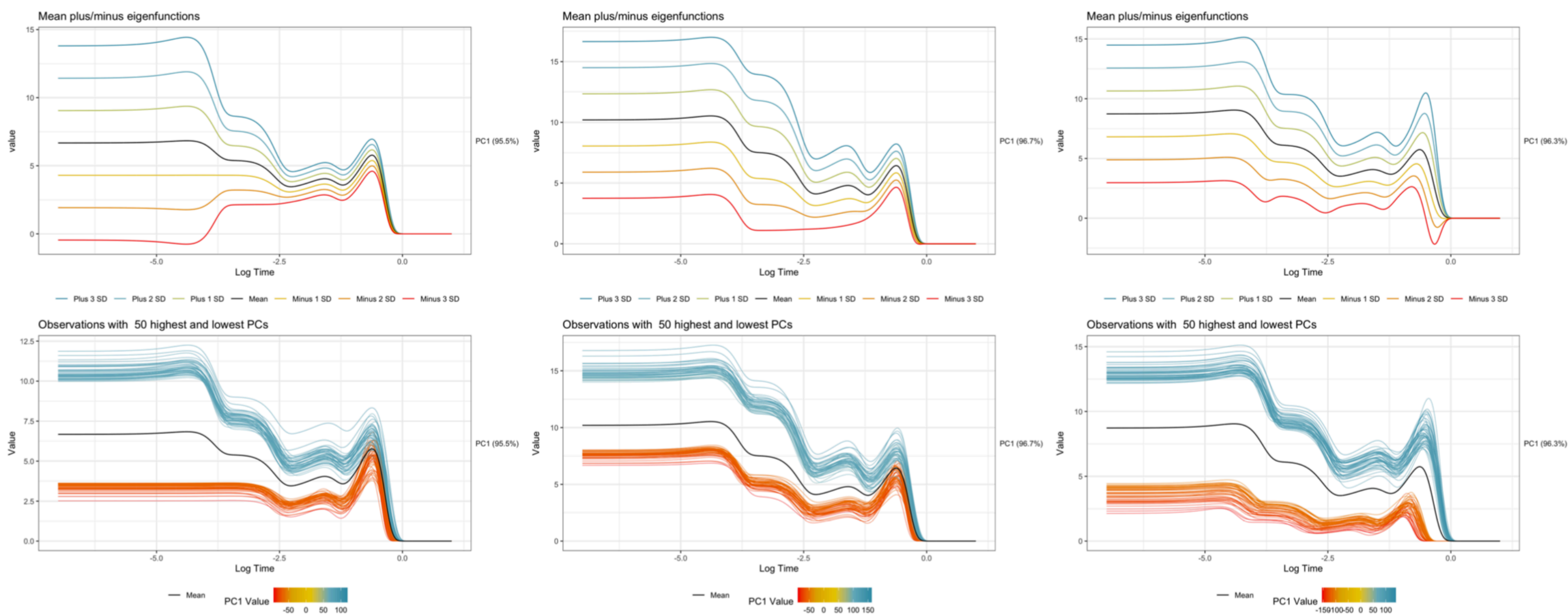


Reflection on Interpretations with Simple Data

- For PCs 1 and 2, it is possible to use PFI to connect functional variability important to neural network predictions (in ways that made sense to SME)
- More difficult with PCs 3 and 4
- Some of the difficulty comes from the interactions between characteristics

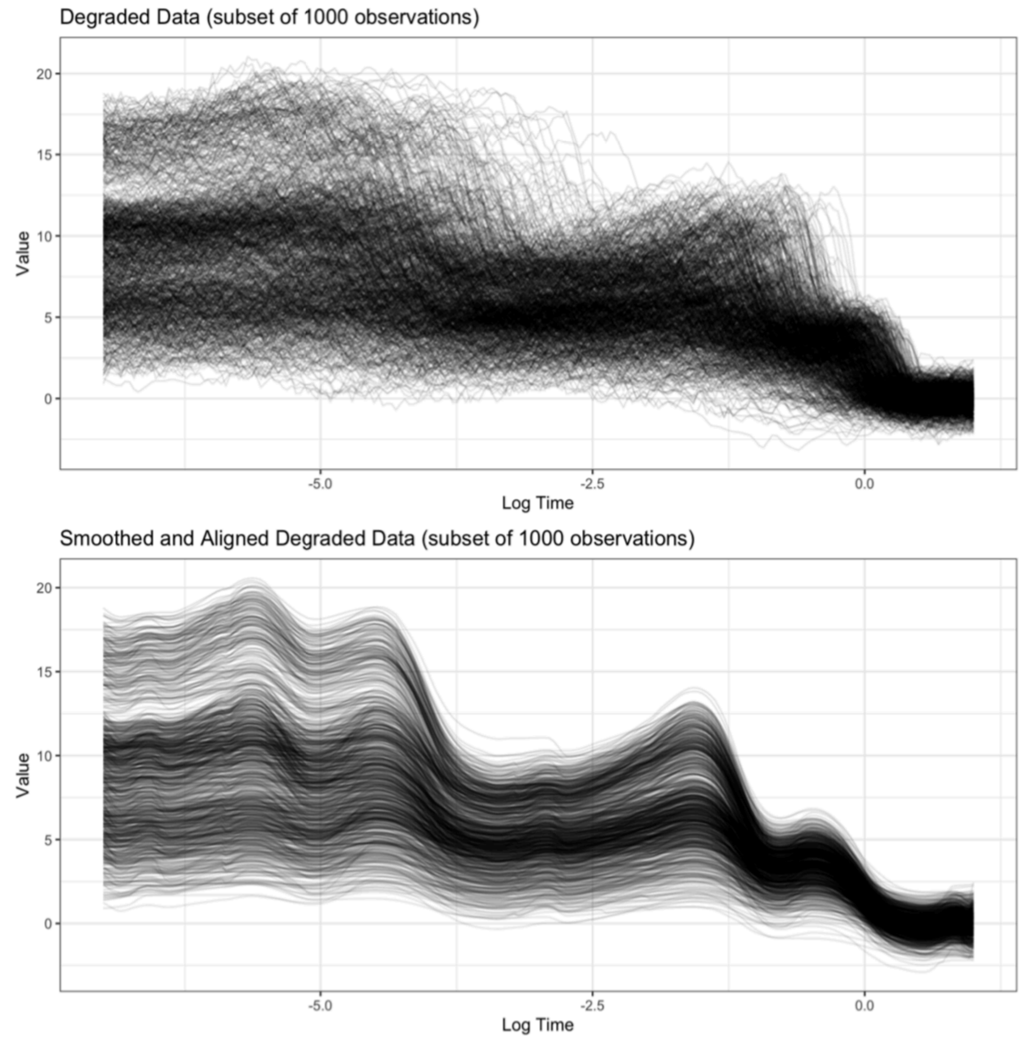
PC 1 from data subset based on characteristics

Focus on Y1 | Focus on Y2 | Focus on Y3

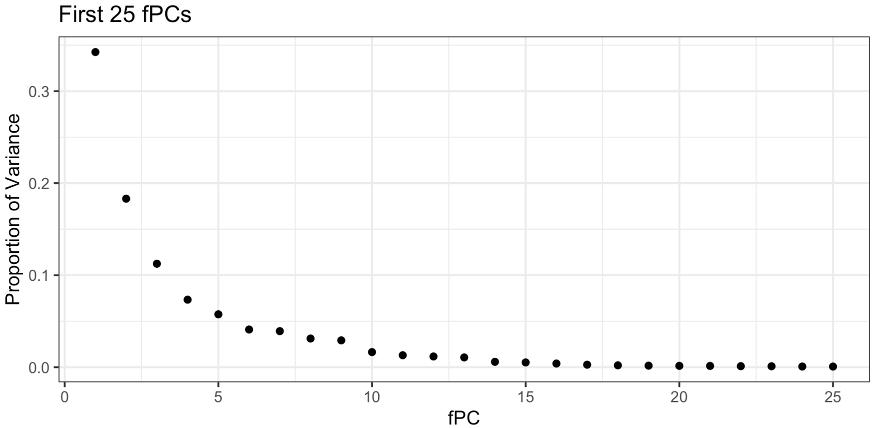


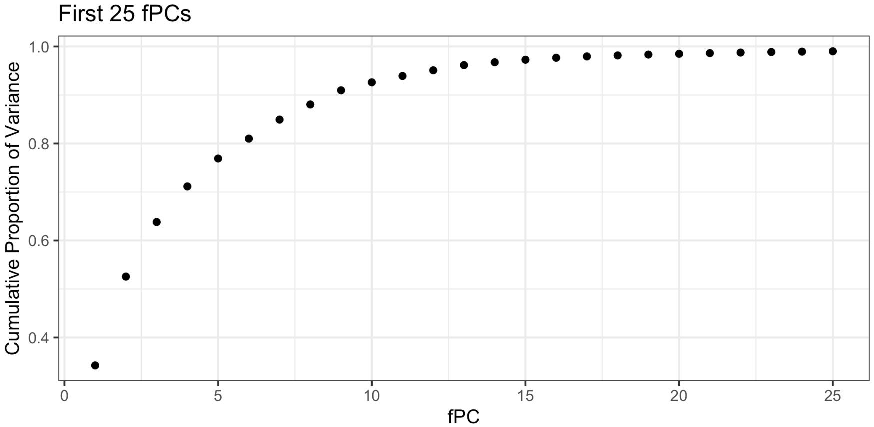
Degraded Data and Joint fPCA

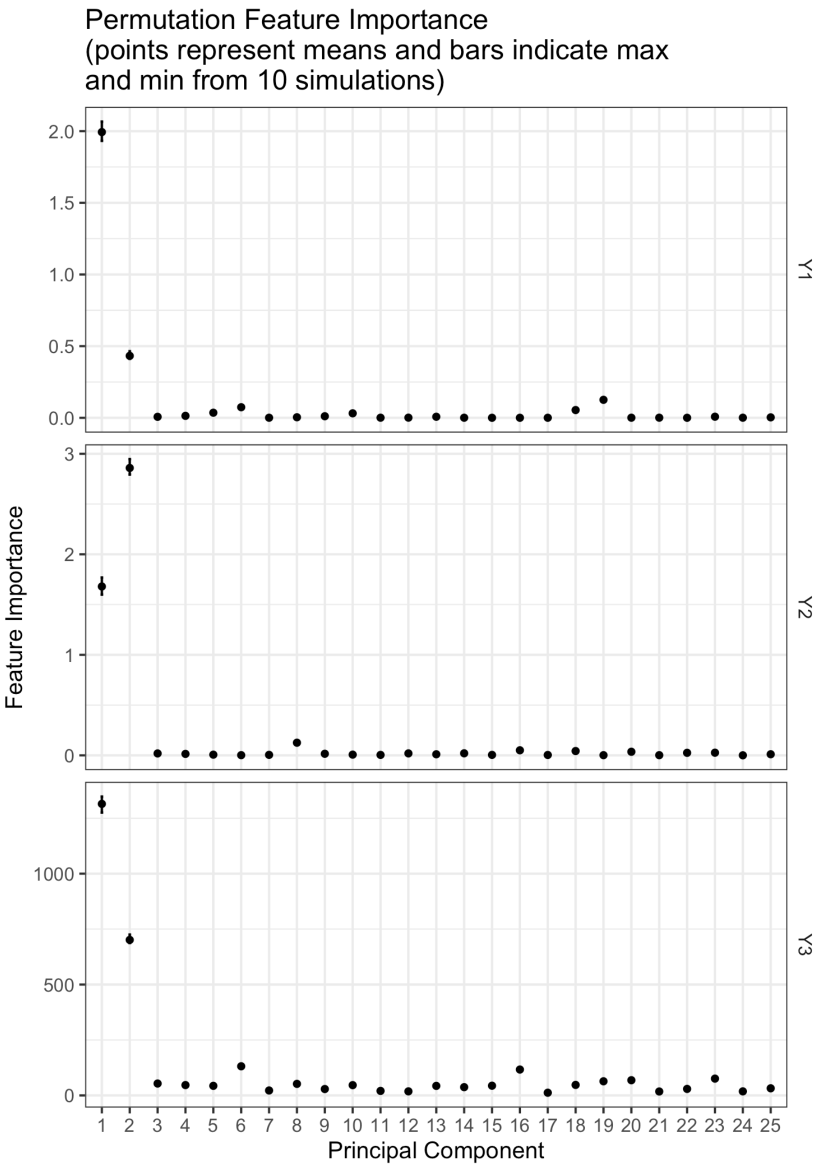
- Alignment done using time warping (fdasrvf package)
- Joint fPCA applied to the function to account for horizontal and vertical variability (fdasrvf package)
- Neural networks fit using the joint fPCs (one for each characteristic)
- PFI applied to identify important fPCs



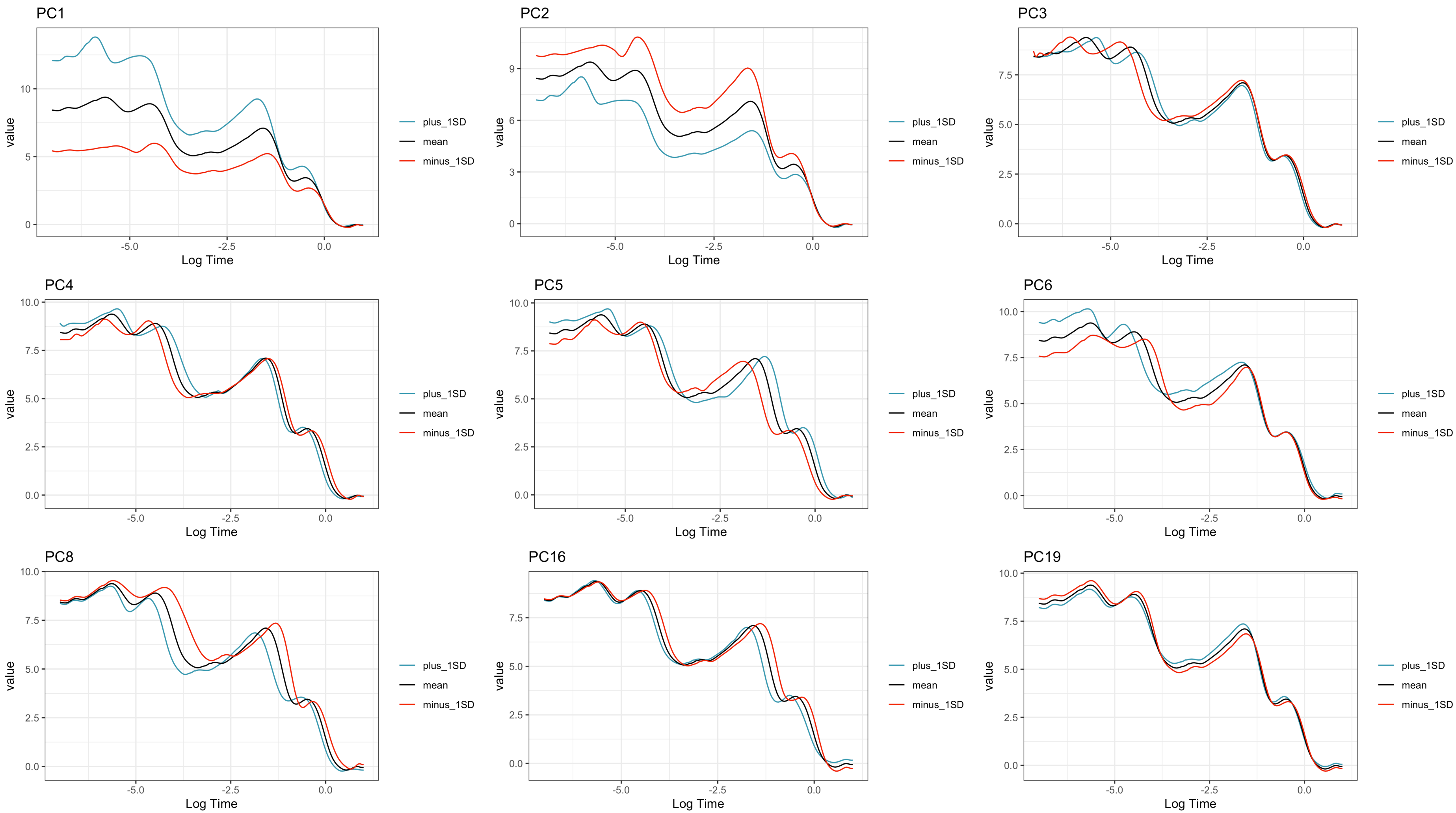
Joint FPCA Proportion of Variance and PFI



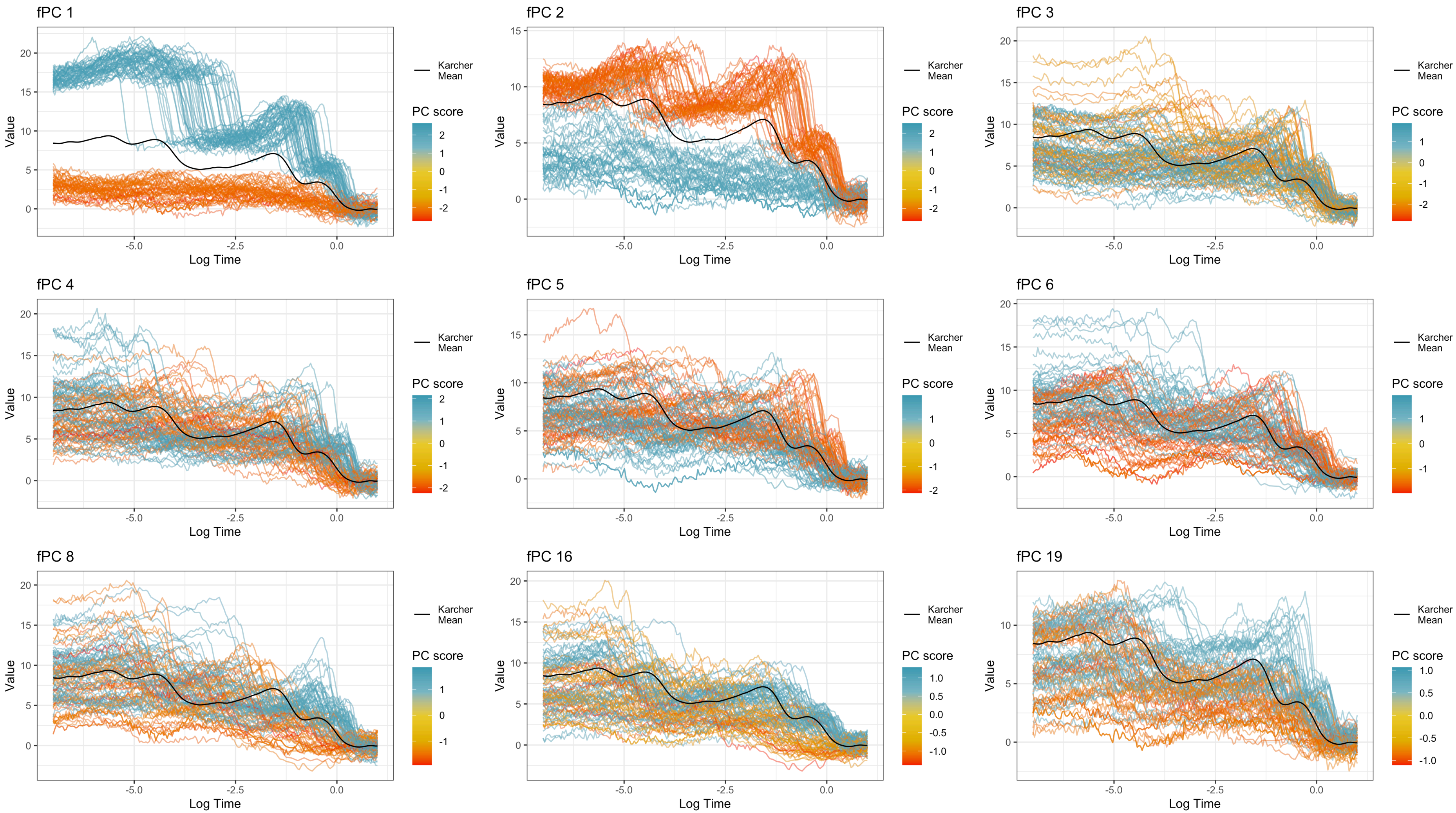




PC Visualizations with Degraded Data



PC Visualizations with Degraded Data



Going Forward

Plans for future work

Future Work

- Implementation on other datasets
- Application of varimax rotation (or other methods) to help with interpretation
- How to adjust PFI in the Bayesian setting?
- Simulation studies to better understand how correlation in functional data affects PFI if not accounted for

Questions?