

# Brief overview of MLAB

## Innovative Approaches to Complex Data Structures in Bioinformatics and Functional Data

Mehdi Maadooliat

mehdi.maadooliat@mu.edu

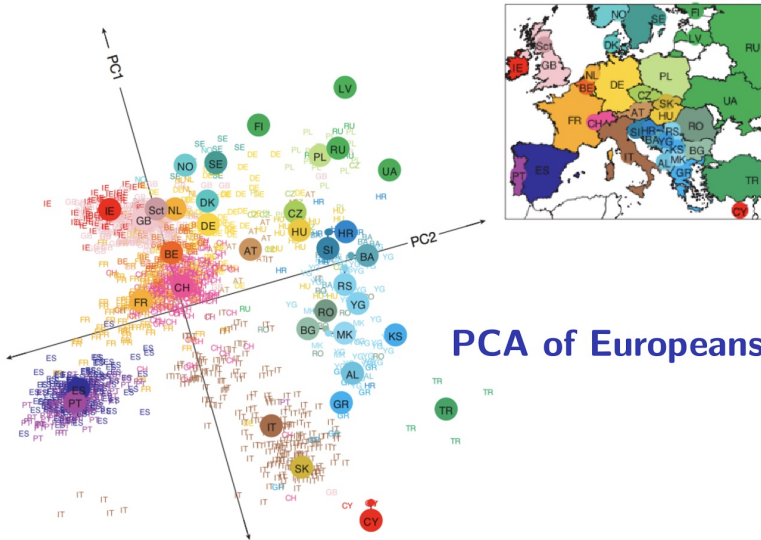
Department of Mathematical and Statistical Sciences  
Marquette University



- 1 Motivating Application in Bioinformatics (GWID)
- 2 Application II – Clustering Brain Signals
- 3 Application III – Remote Sensing Data
- 4 Application IV – Regularization
- 5 Research Outlook

# Complex Data

# Complex Data – Bioinformatics



**PCA of Europeans**



# Genome-Wide Association Studies (GWAS)

Select a Data  
 I.RA

Select Chromosome  
 3

Show Chromosome

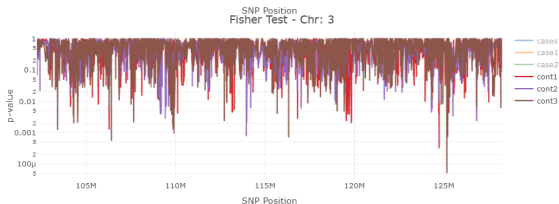
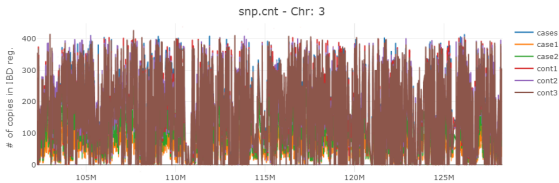
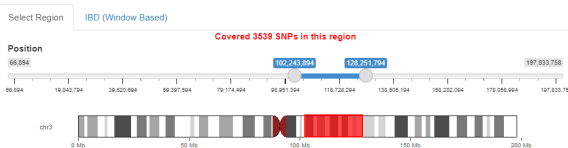
Select a Plot  
 Number of SNPs

Downl. res. (Chr 3 : 894 - 794)

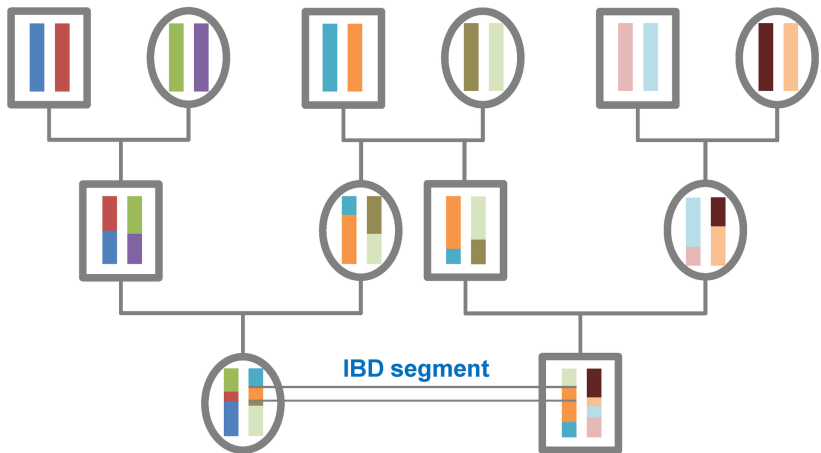
Fisher Exact (G-) Test

Test    Alt.  
 CASES (388)    >

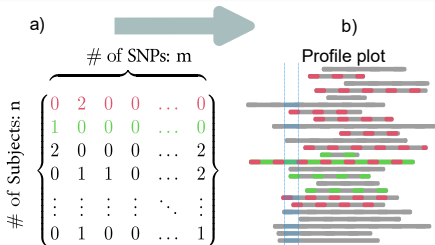
Test if (cases-3) greater 894 - 794



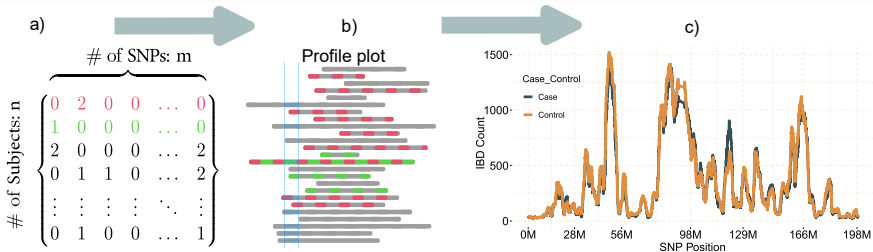
# Identity By Descent (IBD)



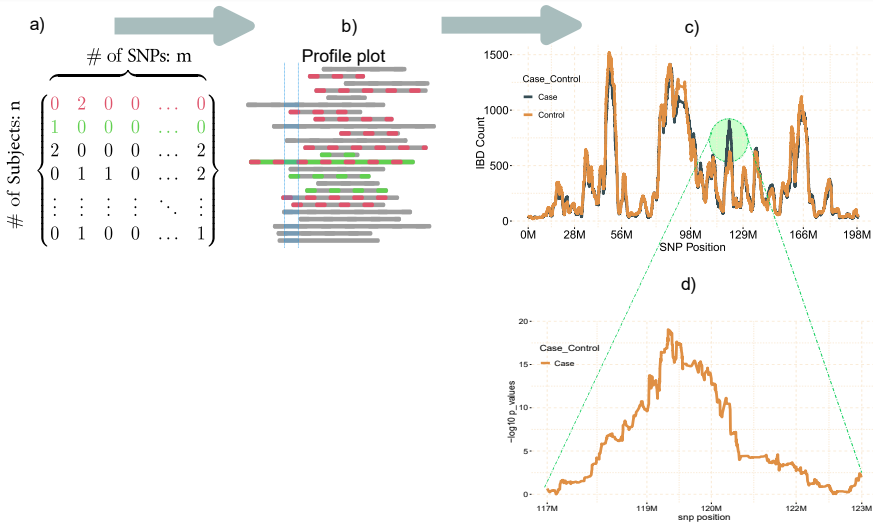
# Genome-Wide Identity-by-Descent (GWID)



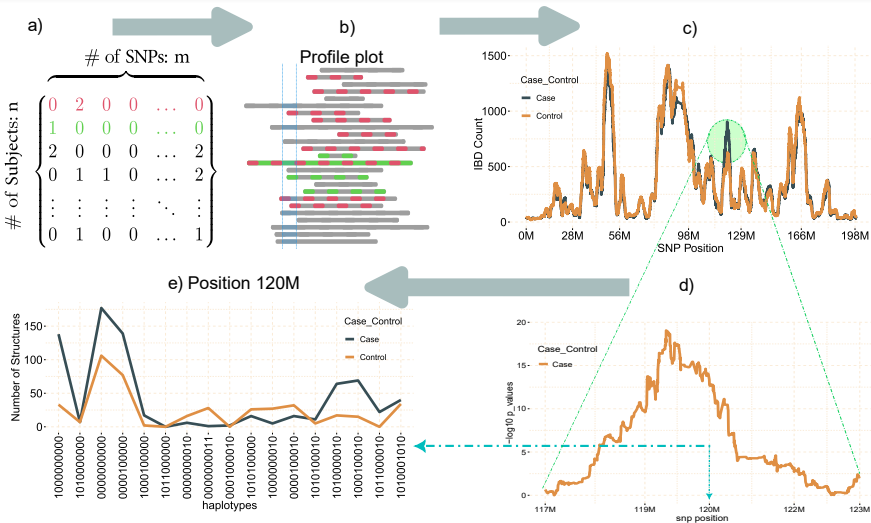
# Genome-Wide Identity-by-Descent (GWID)



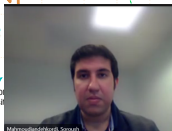
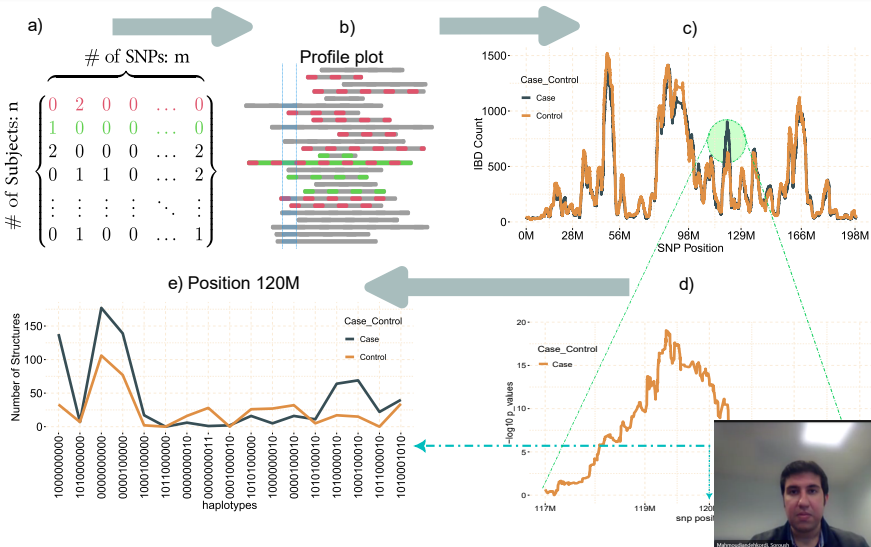
# Genome-Wide Identity-by-Descent (GWID)



# Genome-Wide Identity-by-Descent (GWID)

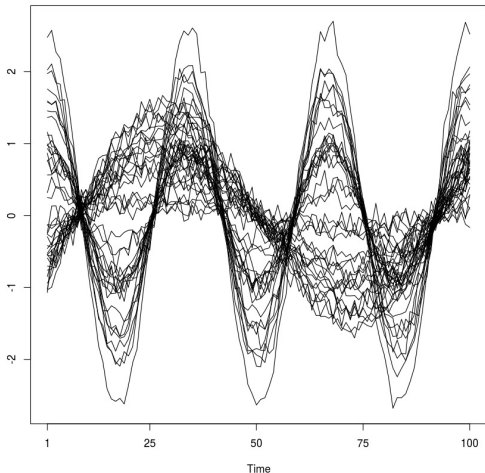


# Genome-Wide Identity-by-Descent (GWID)



# Complex Data – Functional Data

Time Series - Simulation



Select Plot:

FPCA Results ▾

Plot:

- Time Series
- True Functions
- FPCA Basis
- FPCA results

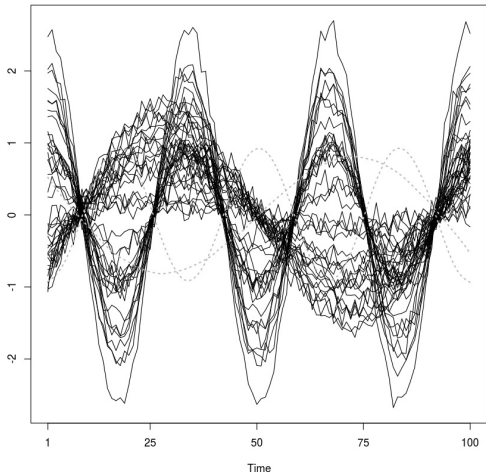
FDA Plot:

- Basis Func.
- B-Spline Smoothing
- Pen. B-Sp. Smooth.



# Complex Data – Functional Data

Time Series - Simulation



Select Plot:

FPCA Results ▾

Plot:

- Time Series
- True Functions
- FPCA Basis
- FPCA results

FPC Basis Cont.



Tun. Par. ( $2^i$ ):

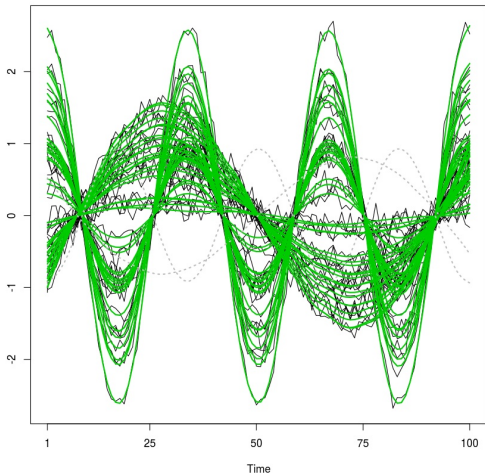


FDA Plot:

- Basis Func.
- B-Spline Smoothing
- Pen. B-Sp. Smooth.

# Complex Data – Functional Data

Time Series - Simulation



Select Plot:

FPCA Results ▾

Plot:

- Time Series
- True Functions
- FPCA Basis
- FPCA results

FPC Basis Cont.



Tun. Par. ( $2^i$ ):

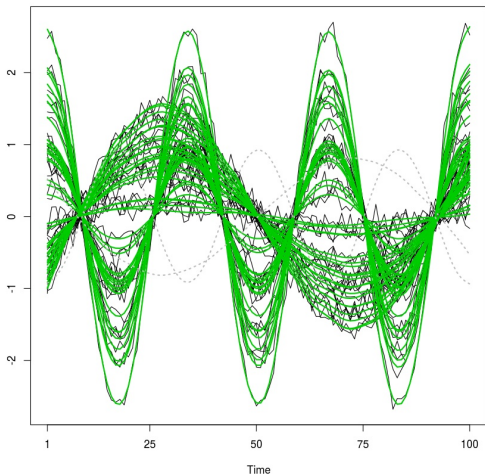


FDA Plot:

- Basis Func.
- B-Spline Smoothing
- Pen. B-Sp. Smooth.

# Complex Data – Functional Data

Time Series - Simulation



Select Plot:

FPCA Results ▾

Plot:

- Time Series
- True Functions
- FPCA Basis
- FPCA results

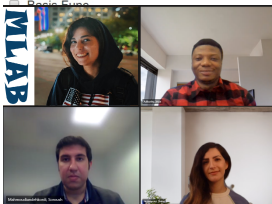
FPC Basis Cont.



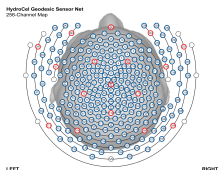
Tun. Par. ( $2^{\Delta i}$ ):



FDA Plot:

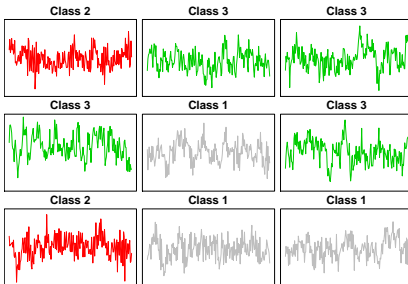


# Nonparametric Collective Spectral Density Estimation (NCSDE)

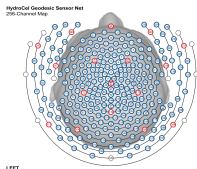


Time Series True Spectral Densities(SD) Periodogram SD: Separate Estimates SD: Collective Estimates

AR(3) Models with :  $\theta = (0.1, 0.5, 0.1)$ ,  $\theta = (0.1, 0.1, 0.5)$  and  $\theta = (0.5, 0.1, 0.1)$ .

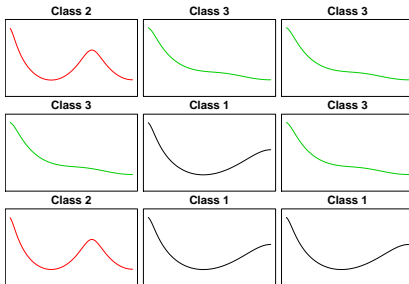


# Nonparametric Collective Spectral Density Estimation (NCSDE)

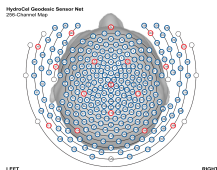


Time Series    **True Spectral Densities(SD)**    Periodogram    SD: Separate Estimates    SD: Collective Estimates

AR(3) Models with :     $\theta = (0.1, 0.5, 0.1)$ ,     $\theta = (0.1, 0.1, 0.5)$     and     $\theta = (0.5, 0.1, 0.1)$ .

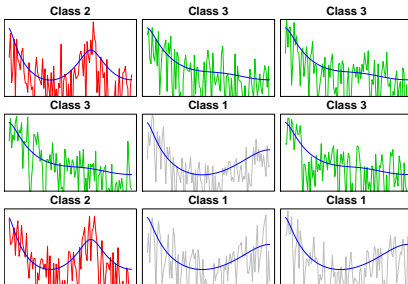


# Nonparametric Collective Spectral Density Estimation (NCSDE)

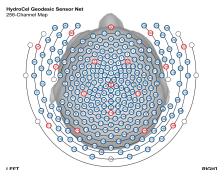


Time Series    True Spectral Densities(SD)    **Periodogram**    SD: Separate Estimates    SD: Collective Estimates

AR(3) Models with :     $\theta = (0.1, 0.5, 0.1)$ ,     $\theta = (0.1, 0.1, 0.5)$     and     $\theta = (0.5, 0.1, 0.1)$ .



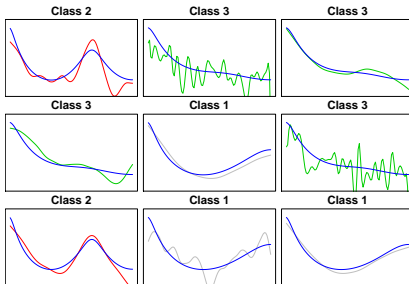
# Nonparametric Collective Spectral Density Estimation (NCSDE)



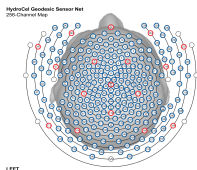
LEFT RIGHT

Time Series    True Spectral Densities(SD)    Periodogram    **SD: Separate Estimates**    SD: Collective Estimates

AR(3) Models with :     $\theta = (0.1, 0.5, 0.1)$ ,     $\theta = (0.1, 0.1, 0.5)$     and     $\theta = (0.5, 0.1, 0.1)$ .

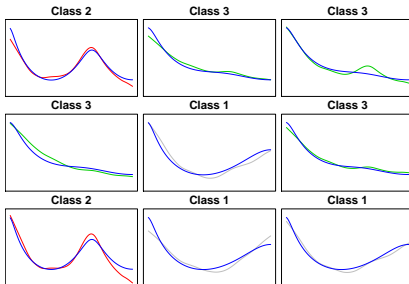


# Nonparametric Collective Spectral Density Estimation (NCSDE)



Time Series    True Spectral Densities(SD)    Periodogram    SD: Separate Estimates    **SD: Collective Estimates**

AR(3) Models with :     $\theta = (0.1, 0.5, 0.1)$ ,     $\theta = (0.1, 0.1, 0.5)$     and     $\theta = (0.5, 0.1, 0.1)$ .





# Singular Spectrum Analysis ( SSA)

The time series of length  $N$  represented by  $Y = \{X_1, X_2, X_3 \dots \dots X_N\}$

1. Select an appropriate window length ( $L$ ) following the constraint  $2 < L < N/2$ . Here  $N$  is the total number of data points and  $K = N-L+1$ .

Form the trajectory matrix with  $K$  lagged vectors of  $Y$  as given by  $\mathbf{T}$ .

$$\mathbf{T} = \begin{matrix} \text{Trajectory matrix} \\ \begin{bmatrix} X_1 & X_2 & X_3 & \dots & X_K \\ X_2 & X_3 & X_4 & \dots & X_{K+1} \\ X_3 & X_4 & X_5 & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ X_L & \cdot & \cdot & \dots & X_{K+L-1} \end{bmatrix} \\ L \times K \end{matrix}$$

# Functional Singular Spectrum Analysis (FSSA)

The **functional** time series of length  $N$  represented by  $Y = \{X_1, X_2, X_3, \dots, X_N\}$



1. Select an appropriate window length ( $L$ ) following the constraint  $2 < L < N/2$ . Here  $N$  is the total number of data points and  $K = N-L+1$ .

Form the trajectory matrix with  $K$  lagged vectors of  $Y$  as given by  $\mathbf{T}$ .

$$\mathbf{T} = \begin{matrix} \text{Trajectory matrix} \\ \begin{bmatrix} X_1 & X_2 & X_3 & \dots & X_K \\ X_2 & X_3 & X_4 & \dots & X_{K+1} \\ X_3 & X_4 & X_5 & \dots & \cdot \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_L & \cdot & \cdot & \cdot & X_{K+L-1} \end{bmatrix} \\ L \times K \end{matrix}$$

# Functional Singular Spectrum Analysis (FSSA)

The **functional** time series of length  $N$  represented by  $Y = \{X_1, X_2, X_3 \dots \dots X_N\}$



1. Select an appropriate window length ( $L$ ) following the constraint  $2 < L < N/2$ . Here  $N$  is the total number of **functions** and  $K = N-L+1$ .

Form the trajectory **operator** with  $K$  lagged vectors of  $Y$  as given by  $T$ .

$$T = \begin{bmatrix} \text{Trajectory operator} \\ \text{X}_1 & \text{X}_2 & \dots & \text{X}_L & \dots & \text{X}_K \\ \text{X}_2 & \text{X}_3 & \dots & \text{X}_{L+1} & \dots & \text{X}_{K+1} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \text{X}_L & \dots & \dots & \dots & \dots & \text{X}_K \end{bmatrix}_{L \times K}$$

# Functional Singular Spectrum Analysis (FSSA)

The **functional** time series of length  $N$  represented by  $Y = \{X_1, X_2, X_3, \dots, X_N\}$



1. Select an appropriate window length ( $L$ ) following the constraint  $2 < L < N/2$ . Here  $N$  is the total number of **functions** and  $K = N-L+1$ .

Form the trajectory **operator** with  $K$  lagged vectors of  $Y$  as given by  $\mathbf{T}$ .

$$\mathbf{T} = \begin{bmatrix} \text{Trajectory operator} \\ X_1 & X_2 & \dots & X_N \\ X_2 & X_3 & \dots & X_{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ X_L & \dots & \dots & X_{N-L+1} \end{bmatrix}_{L \times K}$$

2. Decompose the trajectory matrix by applying Singular Value Decomposition (SVD)

$$\mathbf{T} = \begin{bmatrix} U_1 & U_r & U_{r+1} & U_L \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}_{L \times L} \times \begin{bmatrix} \sqrt{\lambda_1} & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_2} & 0 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \sqrt{\lambda_r} & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \end{bmatrix}_{L \times K} \times \begin{bmatrix} \text{---} \\ \vdots \\ \text{---} \\ \vdots \\ \text{---} \\ \vdots \\ \text{---} \end{bmatrix} \begin{matrix} V_1 \\ V_r \\ V_{r+1} \\ V_K \end{matrix}_{K \times K}$$

# Functional Singular Spectrum Analysis (FSSA)

The **functional** time series of length  $N$  represented by  $Y = \{X_1, X_2, X_3, \dots, X_N\}$



1. Select an appropriate window length ( $L$ ) following the constraint  $2 < L < N/2$ . Here  $N$  is the total number of **functions** and  $K = N-L+1$ .

Form the trajectory **operator** with  $K$  lagged vectors of  $Y$  as given by  $T$ .

$$T = \begin{bmatrix} \text{Trajectory operator} \\ X_1 & X_2 & \dots & X_N \\ X_2 & X_3 & \dots & X_{N+1} \\ \vdots & \vdots & \ddots & \vdots \\ X_L & \dots & \dots & X_{N+1} \end{bmatrix}_{L \times K}$$

2. Decompose the trajectory **operator** by **functional** Singular Value Decomposition (FSVD)

$$T = \begin{bmatrix} U_1 & U_r & U_{r+1} & U_L \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}_{L \times L} \times \begin{bmatrix} \sqrt{\lambda_1} & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_2} & 0 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \sqrt{\lambda_r} & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \end{bmatrix}_{L \times K} \times \begin{bmatrix} V_1 \\ \vdots \\ V_r \\ \vdots \\ V_K \end{bmatrix}_{K \times K}$$

# Functional Singular Spectrum Analysis (FSSA)

The **functional** time series of length  $N$  represented by  $Y = \{X_1, X_2, X_3, \dots, X_N\}$



1. Select an appropriate window length ( $L$ ) following the constraint  $2 < L < N/2$ . Here  $N$  is the total number of **functions** and  $K = N-L+1$ .

Form the trajectory **operator** with  $K$  lagged vectors of  $Y$  as given by  $T$ .

$$T = \begin{bmatrix} \text{Trajectory operator} \\ X_1 & X_2 & \dots & X_K \\ X_2 & X_3 & \dots & X_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ X_L & \dots & \dots & X_{K+L-1} \end{bmatrix}_{L \times K}$$

2. Decompose the trajectory **operator** by **functional** Singular Value Decomposition (FSVD)

$$T = \begin{bmatrix} U_1 & \dots & U_r & U_{r+1} & \dots & U_L \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}_{L \times L} \times \begin{bmatrix} \sqrt{\lambda_1} & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_2} & 0 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \sqrt{\lambda_r} & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \end{bmatrix}_{L \times K} \times \begin{bmatrix} \text{---} \\ \vdots \\ \text{---} \\ \vdots \\ \text{---} \\ \vdots \\ \text{---} \end{bmatrix}_{K \times K} \begin{matrix} V_1 \\ V_r \\ V_{r+1} \\ V_K \end{matrix}$$

# Functional Singular Spectrum Analysis (FSSA)

The **functional** time series of length  $N$  represented by  $Y = \{X_1, X_2, X_3, \dots, X_N\}$



1. Select an appropriate window length ( $L$ ) following the constraint  $2 < L < N/2$ . Here  $N$  is the total number of **functions** and  $K = N-L+1$ .

Form the trajectory **operator** with  $K$  lagged vectors of  $Y$  as given by  $T$ .

$$T = \begin{bmatrix} \text{Trajectory operator} \\ X_1 & X_2 & \dots & X_K \\ X_2 & X_3 & \dots & X_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ X_L & \dots & \dots & X_{K+L-1} \end{bmatrix}_{L \times K}$$

2. Decompose the trajectory **operator** by **functional** Singular Value Decomposition (FSVD)

$$T = \begin{bmatrix} U_1 & \dots & U_r & U_{r+1} & \dots & U_L \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}_{L \times L} \times \begin{bmatrix} \sqrt{\lambda_1} & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_2} & 0 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \sqrt{\lambda_r} & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \end{bmatrix}_{L \times K} \times \begin{bmatrix} \text{---} \\ \vdots \\ \text{---} \\ \vdots \\ \text{---} \\ \vdots \\ \text{---} \end{bmatrix}_{K \times K} \begin{matrix} V_1 \\ V_r \\ V_{r+1} \\ V_K \end{matrix}$$

# Functional Singular Spectrum Analysis (FSSA)

The **functional** time series of length  $N$  represented by  $Y = \{X_1, X_2, X_3, \dots, X_N\}$



1. Select an appropriate window length ( $L$ ) following the constraint  $2 < L < N/2$ . Here  $N$  is the total number of **functions** and  $K = N-L+1$ .

Form the trajectory **operator** with  $K$  lagged vectors of  $Y$  as given by  $T$ .

$$T = \begin{bmatrix} \text{Trajectory operator} \\ X_1 & X_2 & \dots & X_N \\ X_2 & X_3 & \dots & X_{N+1} \\ \vdots & \vdots & \ddots & \vdots \\ X_{K+1} & \dots & \dots & X_N \end{bmatrix}_{L \times K}$$

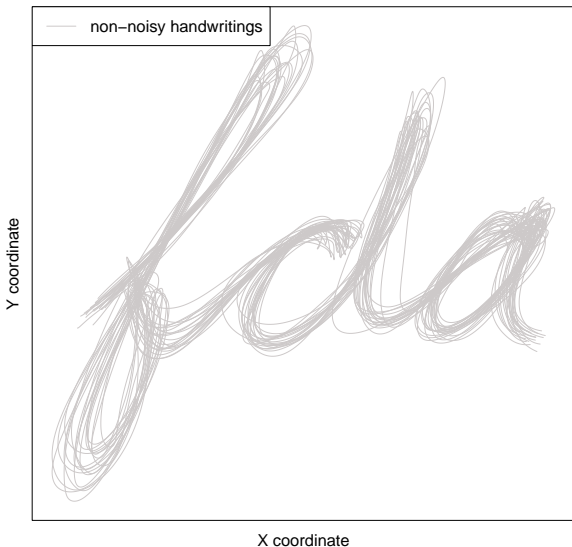
2. Decompose the trajectory **operator** by **functional** Singular Value Decomposition (FSSVD)

$$T = \begin{bmatrix} U_1 & \dots & U_r & U_{r+1} & \dots & U_L \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}_{L \times L} \times \begin{bmatrix} \sqrt{\lambda_1} & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_2} & 0 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \sqrt{\lambda_r} & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \end{bmatrix}_{L \times K} \times \begin{bmatrix} V_1 \\ \vdots \\ V_r \\ \vdots \\ V_K \end{bmatrix}_{K \times K}$$



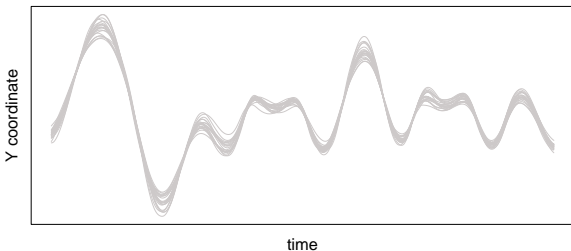
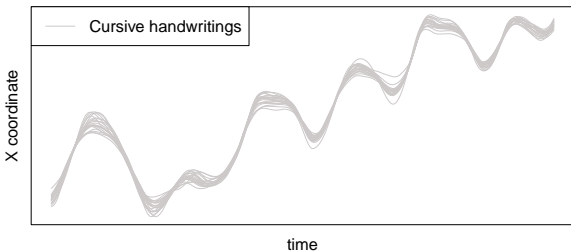
# Regularized Multivariate **FPCA**: Cursive Handwriting Example

fda – Cursive handwriting samples



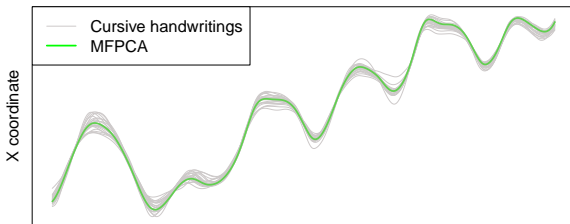
# Regularized Multivariate FPCA: Cursive Handwriting Example

## Cursive handwriting coordinates vs. time

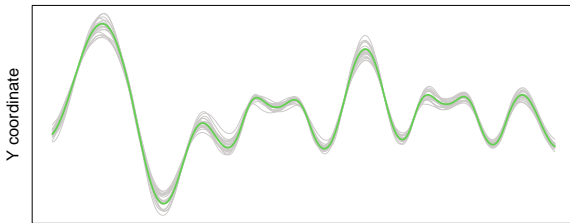


# Regularized Multivariate FPCA: Cursive Handwriting Example

Cursive handwriting coordinates vs. time



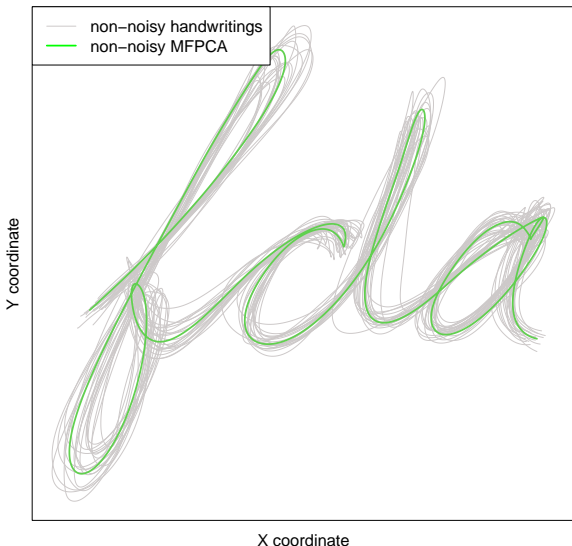
time



time

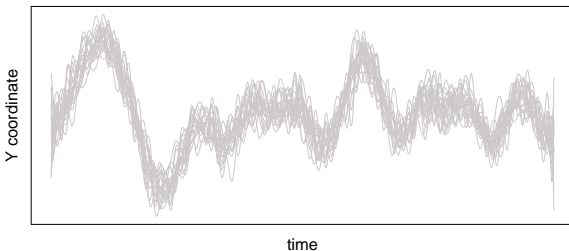
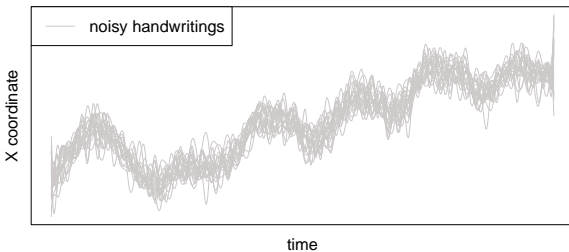
# Regularized Multivariate FPCA: Cursive Handwriting Example

MFPCA – PC 1



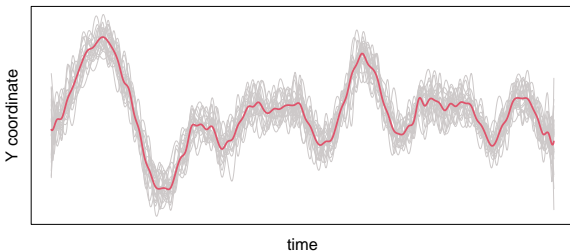
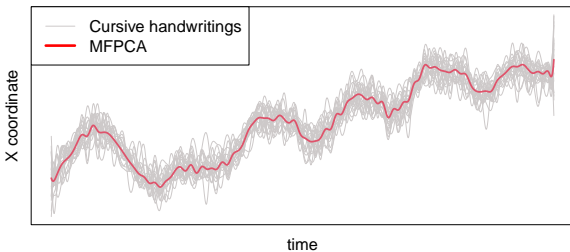
# Regularized Multivariate **FPCA**: Cursive Handwriting Example

noisy handwriting coordinates vs. time



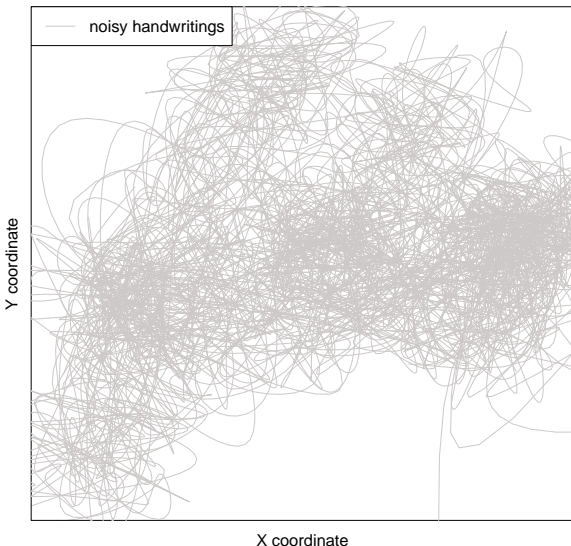
# Regularized Multivariate FPCA: Cursive Handwriting Example

noisy handwriting coordinates vs. time



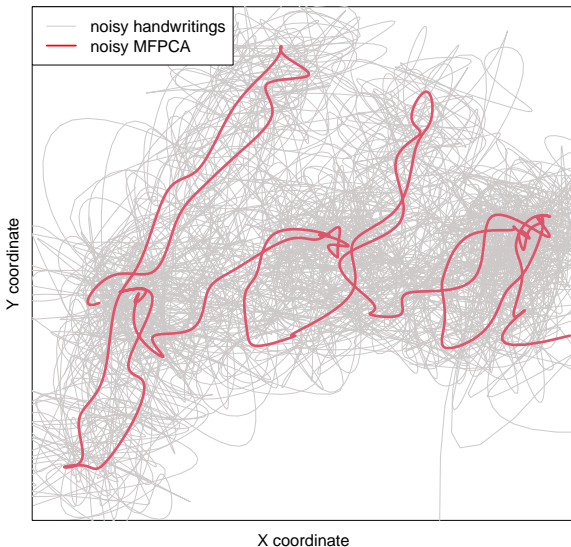
# Regularized Multivariate **FPCA**: Cursive Handwriting Example

fda – noisy handwriting samples



# Regularized Multivariate **FPCA**: Cursive Handwriting Example

**MFPCA – PC 1**

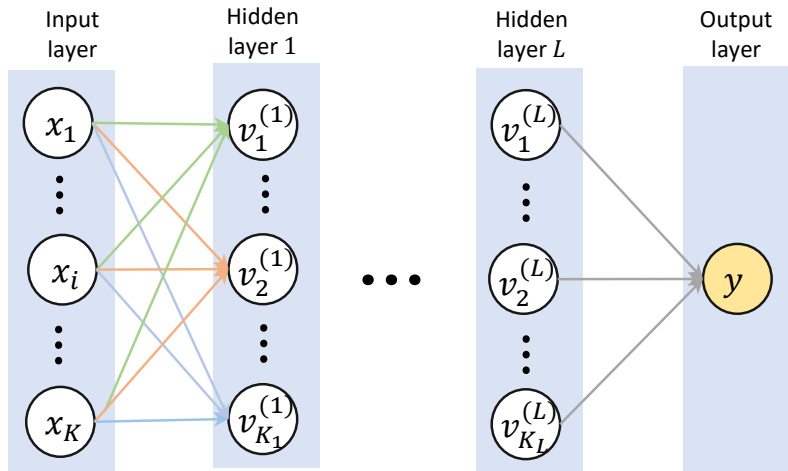




# Regularized Multivariate **FPCA**: Cursive Handwriting Example

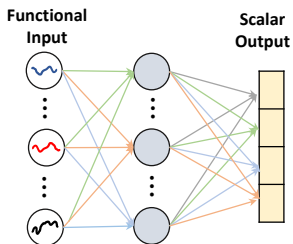
# Regularized Multivariate **FPCA**: Cursive Handwriting Example

# Deep Neural Networks

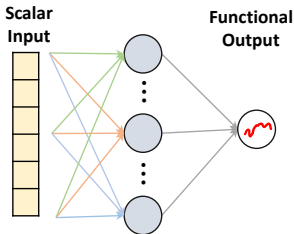


# FDA in Deep Neural Networks

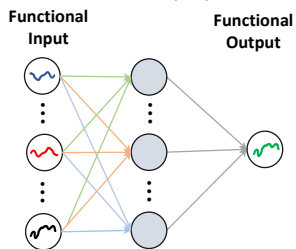
Type (I)



Type (II)

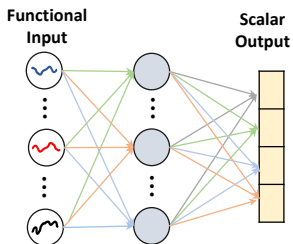


Type (III)

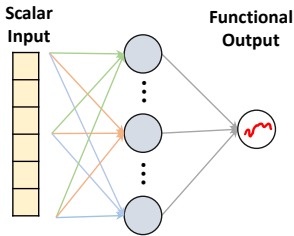


# FDA in Deep Neural Networks

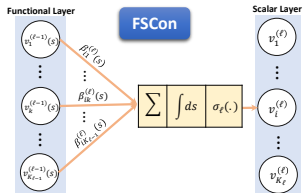
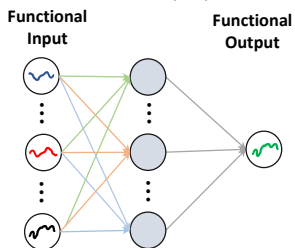
Type (I)



Type (II)

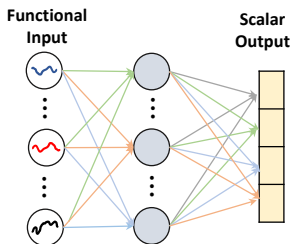


Type (III)

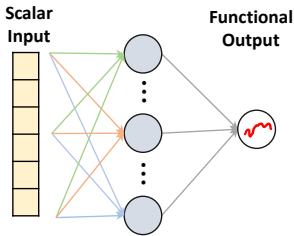


# FDA in Deep Neural Networks

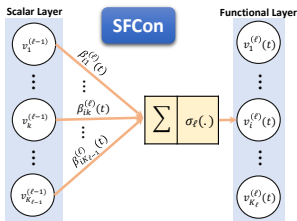
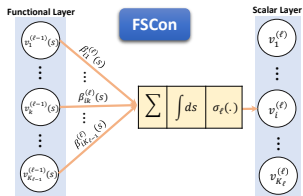
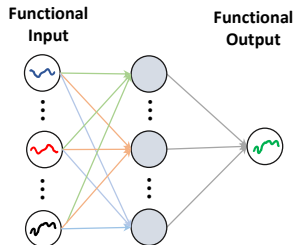
Type (I)



Type (II)

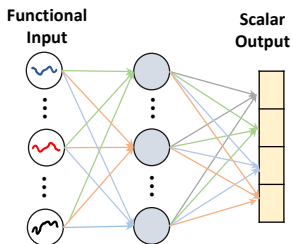


Type (III)

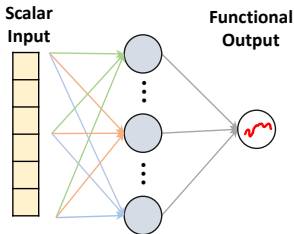


# FDA in Deep Neural Networks

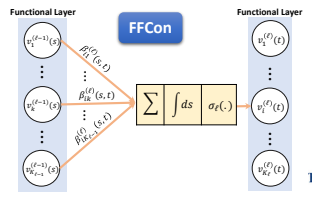
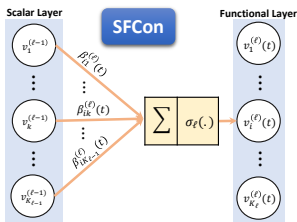
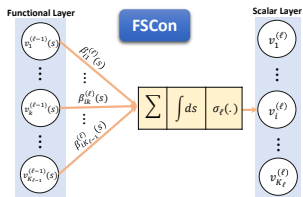
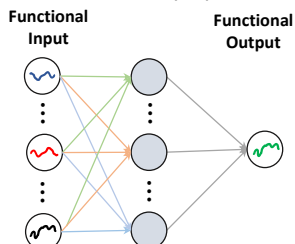
Type (I)



Type (II)

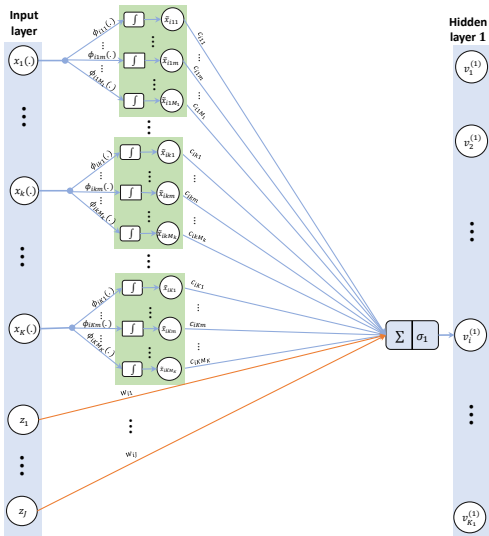


Type (III)



BE THE DIFFERENCE.

# FDA in Deep Neural Networks





# MLAB Team

- Collaborators:

- Steven Schrodi at UW-Madison
- Hossein Haghbin at Persian Gulf University



- Ph.D. Students:

- Soroush Dehkordi (GWID & HPCA)
- Jesse Adikorley (ReHSSA)
- Mobina Pourmoshir (ReHPCA)
- Shirin Nezampour (Multivariate NCSDE)

- Alumni:

- Yue Zhao (ReFPCA)  
Postdoctoral Associate @ University of Minnesota School of Public Health
- Jordan Trinka (FSSA)  
AI Engineer @ Booz Allen Hamilton
- Ana Nodehi (TPPCA)  
Research Asso. @ University of Bristol
- Morteza Najibi (PSCDE)  
Data Scientist @ Lund University

# MLAB Team

- Collaborators:

- Steven Schrodi at UW-Madison
- Hossein Haghbin at Persian Gulf University



- Ph.D. Students:

- Soroush Dehkordi (**GWID & HPCA**)
- Jesse Adikorley (**ReHSSA**)
- Mobina Pourmoshir (**ReHPCA**)
- Shirin Nezampour (Multivariate **NCSDE**)



- Alums:

- Yue Zhao (**ReFPCA**)

Postdoctoral Associate @ University of Minnesota School of Public Health

- Jordan Trinko (**FSSA**)

AI Engineer @ Booz Allen Hamilton

- Ana Nodehi (**TPPCA**)

Research Asso. @ University of Bristol

- Morteza Najibi (**PSCDE**)

Data Scientist @ Lund University

# MLAB Team

- Collaborators:

- Steven Schrodi at UW-Madison
- Hossein Haghbin at Persian Gulf University



- Ph.D. Students:

- Soroush Dehkordi (**GWID & HPCA**)
- Jesse Adikorley (**ReHSSA**)
- Mobina Pourmoshir (**ReHPCA**)
- Shirin Nezampour (Multivariate **NCSDE**)



- Alumns:

- Yue Zhao (**ReFPCA**)

Postdoctoral Associate @ University of Minnesota School of Public Health

- Jordan Trinka (**FSSA**)

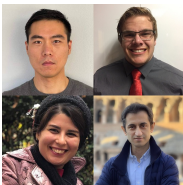
AI Engineer @ Booz Allen Hamilton

- Ana Nodehi (**TPPCA**)

Research Asso. @ University of Bristol

- Morteza Najibi (**PSCDE**)

Data Scientist @ Lund University



## Reference

My Personal Homepage.

**Maadooliat, M.** Homepage:

<http://www.mssc.mu.edu/~mehdi>.

Dehkordi, S., Schrodi, S. J. and **Maadooliat, M.** (2024). GWID: A new Genome-Wide identity-by-descent visualization and analysis tool for disease gene mapping. *Bioinformatics Advances* **Accepted**.

**Maadooliat, M.**, Sun, Y. and Chen, T. (2018). Nonparametric Collective Spectral Density Estimation with an Application to Clustering the Brain Signals. *Statistics in Medicine* **37**, 4789—4806.

Hagbin, H., Najibi, SM., Mahmoudvand, R., Trink, J., and **Maadooliat, M.** (2020). Functional Singular Spectrum Analysis. *STAT (Special Issue in Statistics and Data Science)*.

# Useful Courses and Skills

- Courses:
  - MSSC 6000 – Scientific Computing
  - MSSC 6010 – Computational Probability
  - MSSC 6020 – Statistical Simulation
  - MSSC 6040 – Applied Linear Algebra
  - MSSC 6250 – Statistical Machine Learning
  - BIIN 6000 – Introduction to Bioinformatics
- More in MLAB:
  - Advanced R, Second Edition by Wickham, 2019
  - Functional Data Analysis by Ramsay and Silverman, 2005
- Research Direction in MLAB
  - Dimension Reduction for Complex Data-Structures with Applications in Bioinformatics and Functional Data