

# Functional Data Analysis – “What to do when your data are a curve or spectra”

**Ryan Parker, PhD**

Research Statistician Developer

[ryan.parker@jmp.com](mailto:ryan.parker@jmp.com)

**Clay Barker, PhD**

Research Statistician Developer

[clay.barker@jmp.com](mailto:clay.barker@jmp.com)

**Tom Donnelly, PhD, CAP**

Systems Engineer, Defense & Aerospace Team

[tom.donnelly@jmp.com](mailto:tom.donnelly@jmp.com)

**DATAWorks 2025  
Short Course  
April 22  
IDA**



# Schedule

|           |   |
|-----------|---|
| 0900-0930 | Intro – <i>Tom Donnelly</i>   |
| 0930-1130 | Basic Functional Data Analysis (FDA) –<br><i>Ryan Parker</i>                    |
| 1130-1200 | Advanced FDA Direct Methods –<br><i>Clay Barker</i>                             |
| 1200-1330 | LUNCH   |
| 1330-1530 | Advanced FDA Direct Methods continued –<br><i>Clay Barker &amp; Ryan Parker</i> |
| 1530-1600 | Case studies – <i>Tom Donnelly</i>  |
| 1600-1630 | Wrap-up, Q & A, ( <b>PLEASE ask at any time!</b> )                              |

\*NOTE: We'll take 10-min. breaks every hour  
≈ 1000, 1100, 1430, & 1530

# Modeling Streamed Sensor Data with Functional Data Analysis

- 1) Using the Sensor Stream as an Input to a Machine Learning Model, and
- 2) Predicting the Shape of the Sensor Stream using Design of Experiments
- 3) Functional Data as both Input & Output for a Glycemic Response Model

**Tom Donnelly**, PhD, CAP

JMP Defense & Aerospace Team

*Principal System Engineer*

[tom.donnelly@jmp.com](mailto:tom.donnelly@jmp.com)

302-489-9291

DATAWorks 2025

FDA Short Course

April 22

IDA



# Who in DoD is Looking at or Already Using FDA?

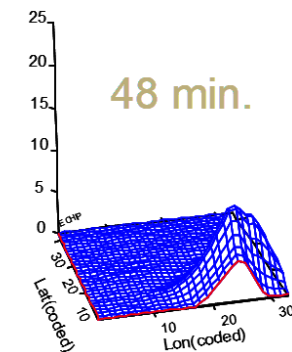
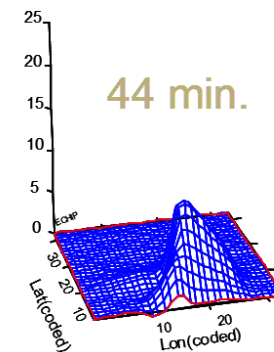
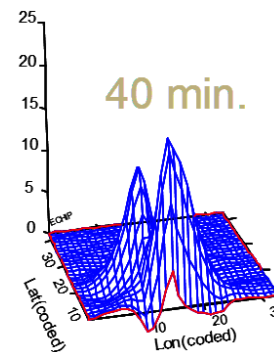
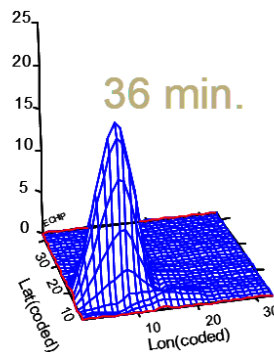
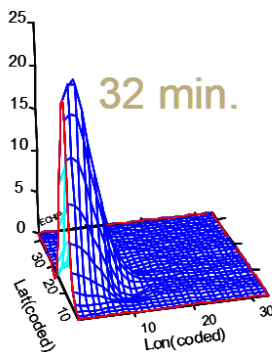
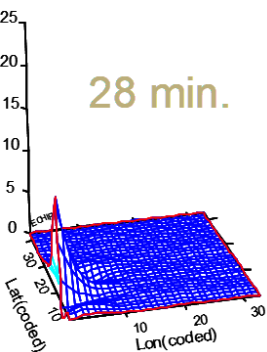
- IDA – [see DATAWorks 2024 video](#) by Curtis Miller
- DEVCOM – Chemical Biological Center – [see 2024 video](#) by Matthew Lux, John (Jay) Davies, David Garcia
- DEVCOM – Armaments Center
- Army Evaluation Center
- Eglin AFB
- Edwards AFB
- NAWCWD
- COTF
- NUWC
- MCOTEA
- JHU/APL
- LMCO
- NGC

First ran into Functional Data 18 Years ago at the  
Army's *Edgewood Chemical Biological Center*

Now DEVCOM CBC



U.S. Army Combat Capabilities Development Command  
Chemical Biological Center



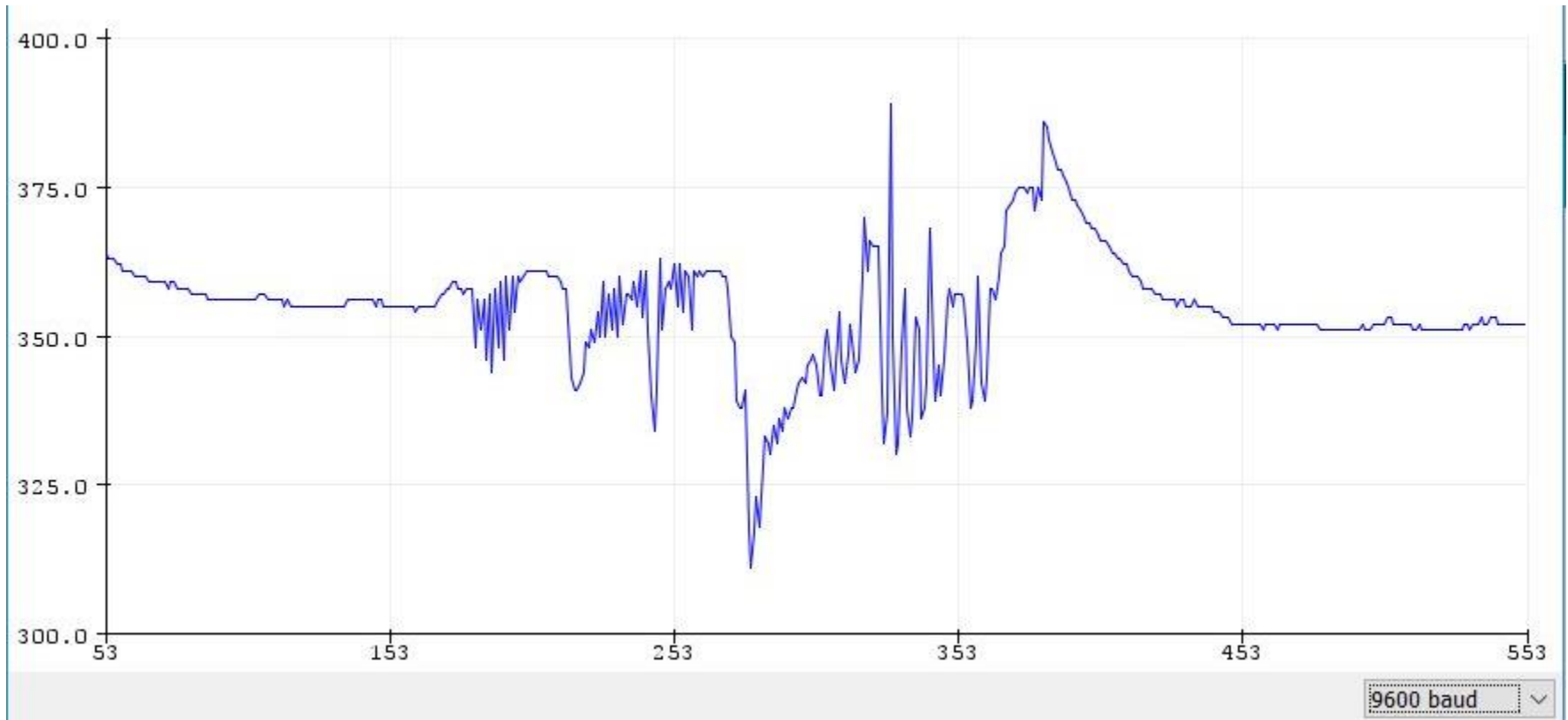
## 10-factor Agent Transport & Dispersion Simulation

- Able to model Concentration *at a particular time*,
- or Dosage *at end of time*,
- but **NOT** Concentration *shape over time*
- Profs. Jeff Wu & Roshan Joseph from Georgia Tech ISyE suggested using Functional Data Analysis

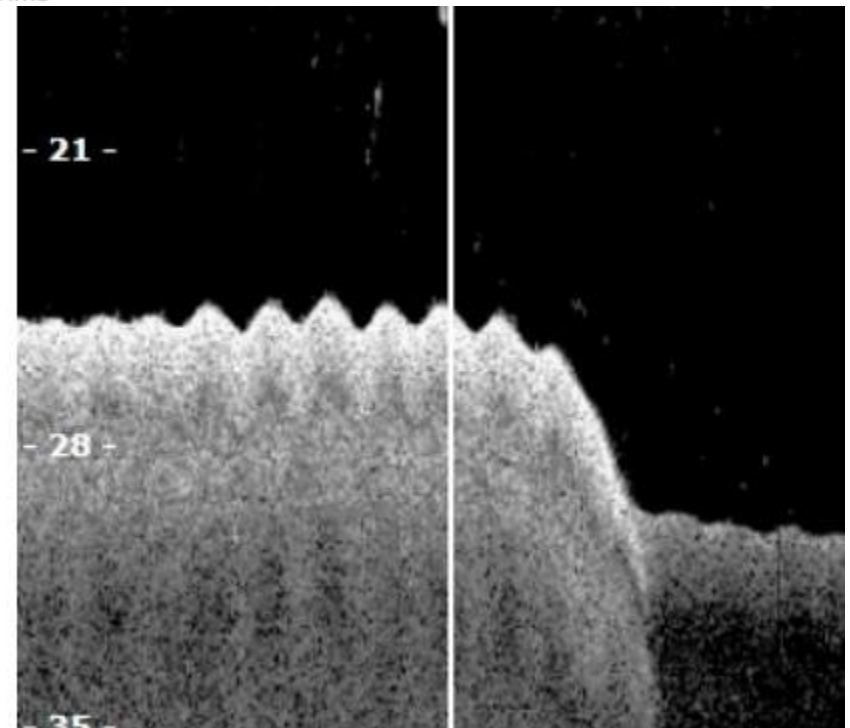
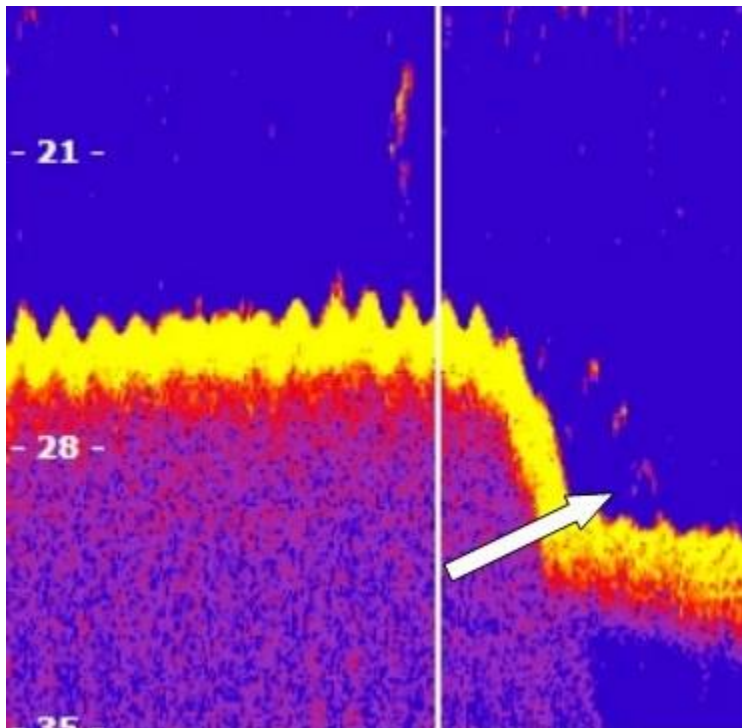
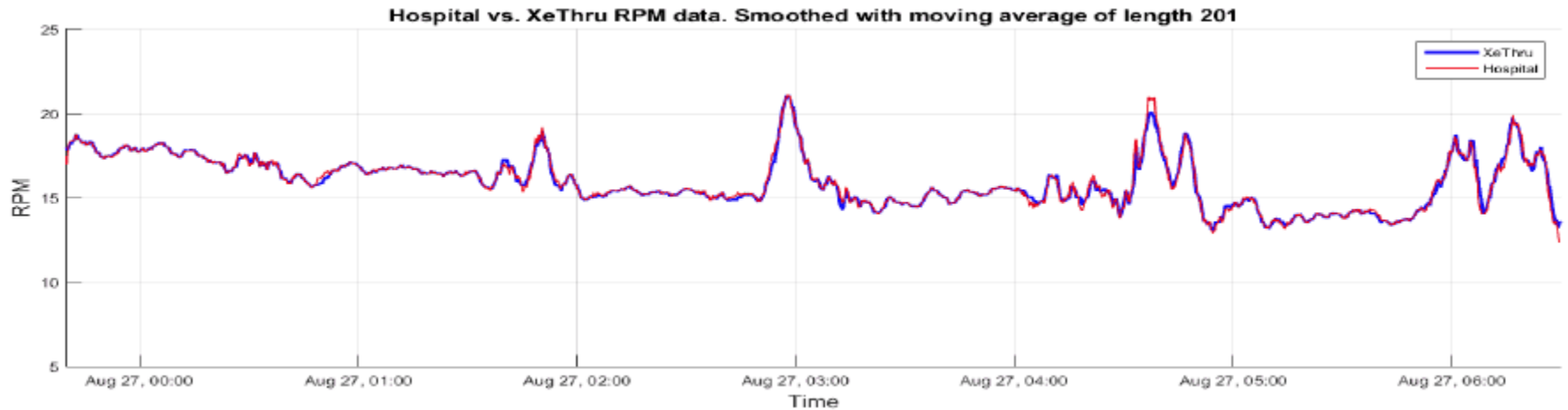
# Examples of Functional Data

- Sensor streams (*temperature, pressure, humidity, proximity, vibration, flow, force, intensity, concentration, etc.*)
- Measurements taken over a range (*time, frequency, wavelength, distance, energy, etc.*)
- Tool wear/gun barrel degradation
- Radar/sonar readings
- Trajectories of flights between cities, landing patterns, etc.
- Tracking of surgeon hand movement
- Electrocardiograms (EKGs)
  
- Using wavelets as basis functions enables SPECTRAL FDA (*Raman, NMR, NIR, mass spec, chromatography, PIXE, etc.*)
  
- ***Almost any response in a longitudinal order is functional data***

# Vibration Sensor

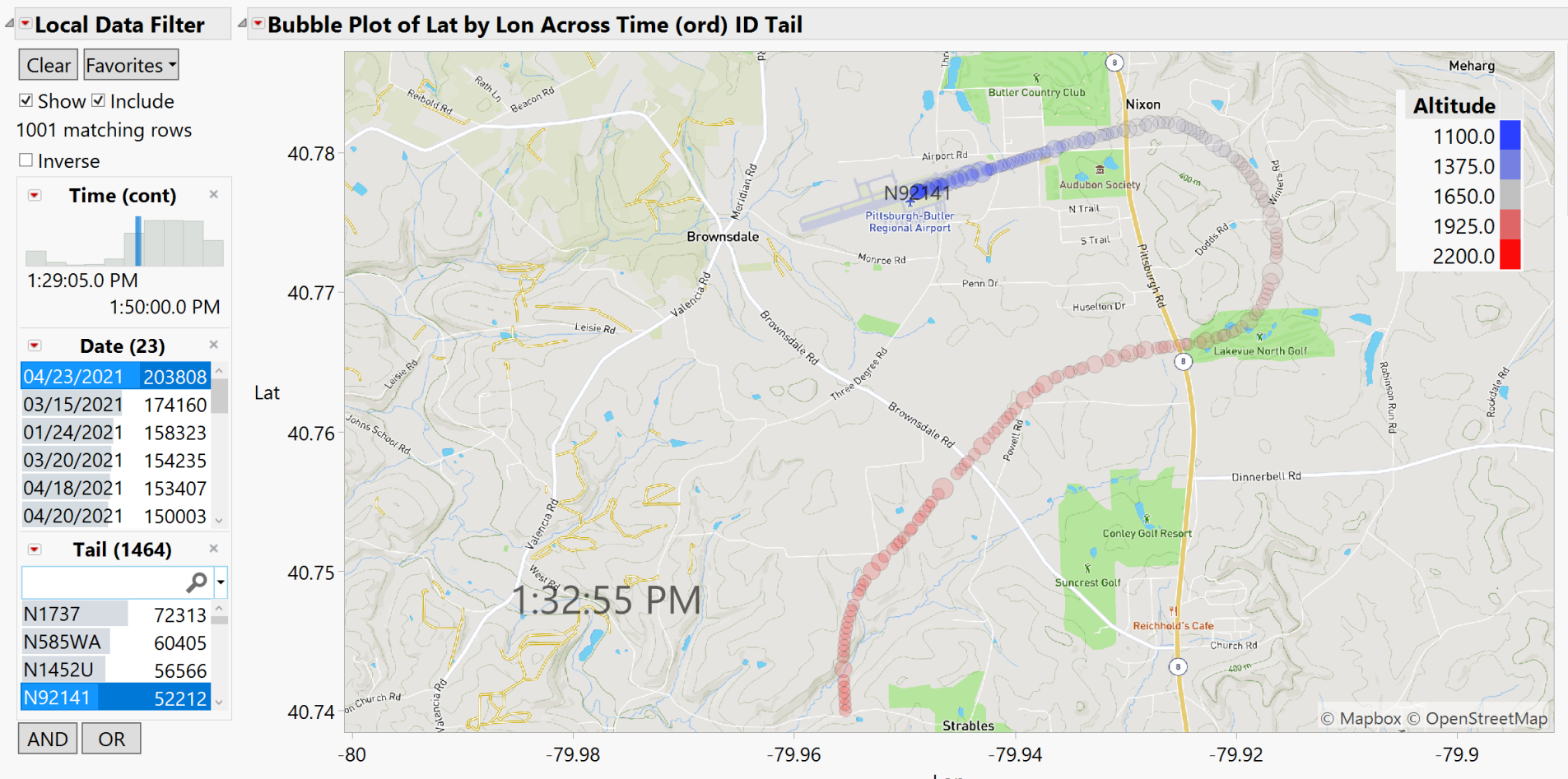


# Radar and Sonar Data

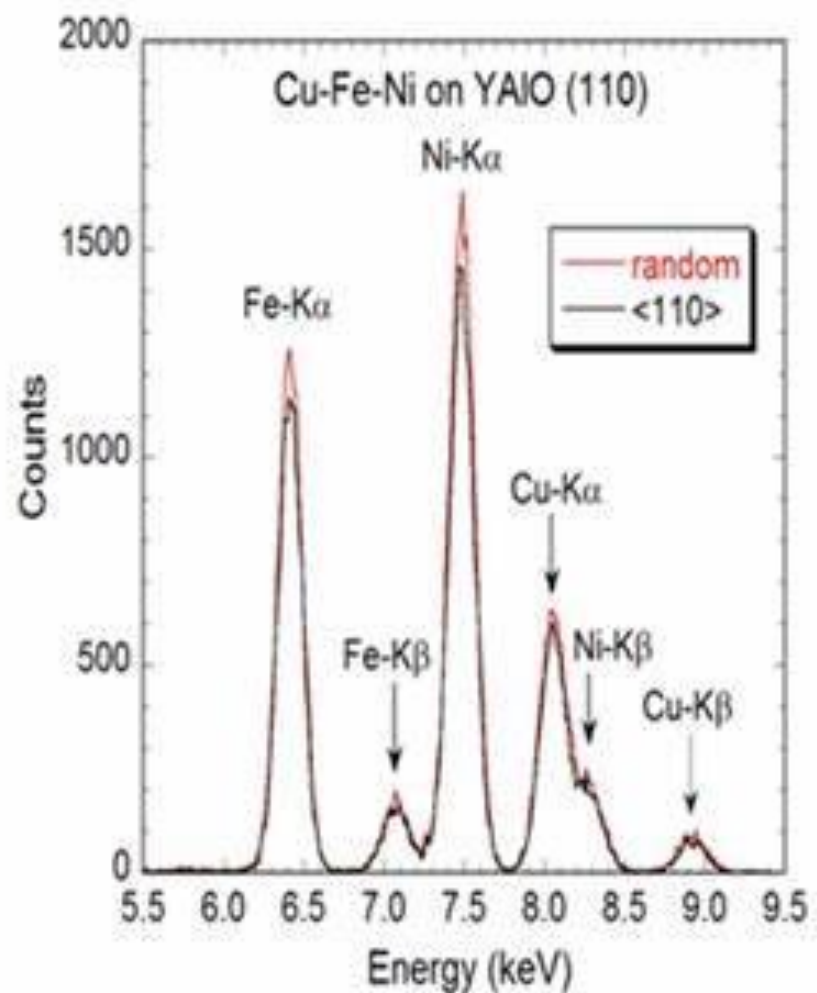
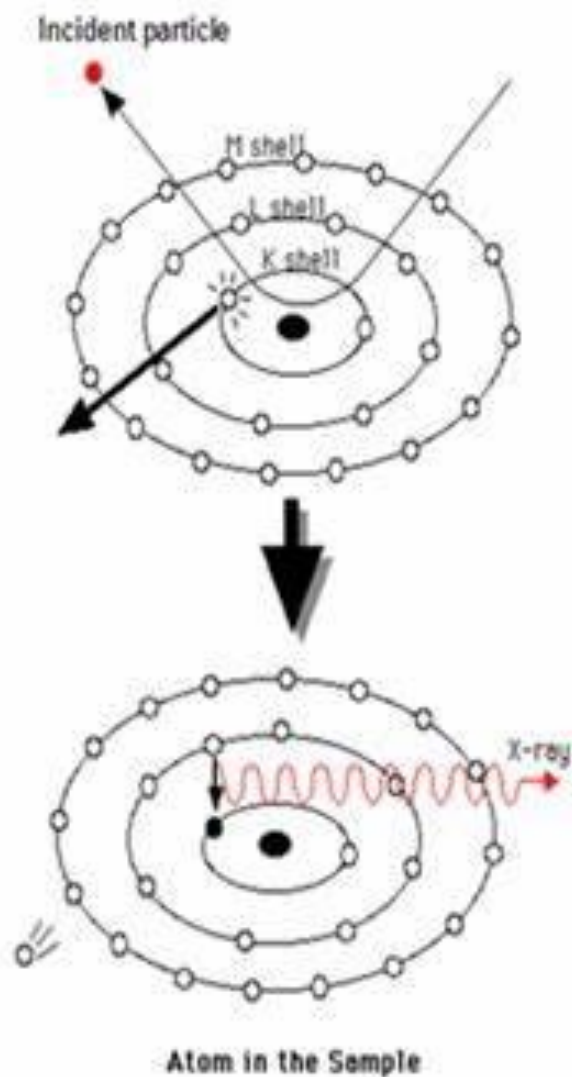




# Aircraft Trajectory Modeling



# Proton-Induced X-ray Emission (PIXE)



# Remaining Useful Life Estimation Using Functional Data Analysis

Qiyao Wang, Shuai Zheng, Ahmed Farahat, Susumu Serita, Chetan Gupta  
 Industrial AI Laboratory, Hitachi America, Ltd. R&D  
 Santa Clara, CA, USA  
 firstname.lastname@hal.hitachi.com

**Abstract**—Remaining Useful Life (RUL) of an equipment or one of its components is defined as the time left until the equipment or component reaches its end of useful life. Accurate RUL estimation is exceptionally beneficial to Predictive Maintenance, and Prognostics and Health Management (PHM). Data driven

contrary, when the end of the equipment’s life is approaching, accurate RUL estimation provides early enough warning to the maintenance departments such that they can plan their actions in advance.

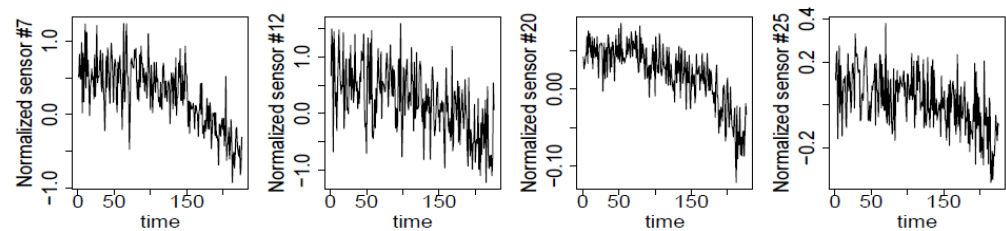


Fig. 3: Removing the effect of operating conditions on sensor data

TABLE III: Score comparison on C-MAPSS data and improvement (‘IMP’) of functional MLP over LSTM [3]

| Model       | FD001             | FD002             | FD003             | FD004             |
|-------------|-------------------|-------------------|-------------------|-------------------|
| MLP [10]    | $1.8 \times 10^4$ | $7.8 \times 10^5$ | $1.7 \times 10^4$ | $5.6 \times 10^6$ |
| SVR [10]    | $1.4 \times 10^3$ | $5.9 \times 10^5$ | $1.6 \times 10^3$ | $3.7 \times 10^5$ |
| RVR [10]    | $1.5 \times 10^3$ | $1.7 \times 10^4$ | $1.4 \times 10^3$ | $2.7 \times 10^4$ |
| CNN [10]    | $1.3 \times 10^3$ | $1.4 \times 10^4$ | $1.6 \times 10^3$ | $7.9 \times 10^3$ |
| LSTMBS [11] | $4.8 \times 10^2$ | $8.0 \times 10^3$ | $4.9 \times 10^2$ | $5.2 \times 10^3$ |
| LSTM [3]    | $3.4 \times 10^2$ | $4.5 \times 10^3$ | $8.5 \times 10^2$ | $5.6 \times 10^3$ |
| FMLP        | $2.0 \times 10^2$ | $9.0 \times 10^2$ | $1.8 \times 10^2$ | $1.0 \times 10^3$ |
| IMP         | 41.18%            | 80.00%            | 78.82%            | 82.14%            |

# Electrocardiograms

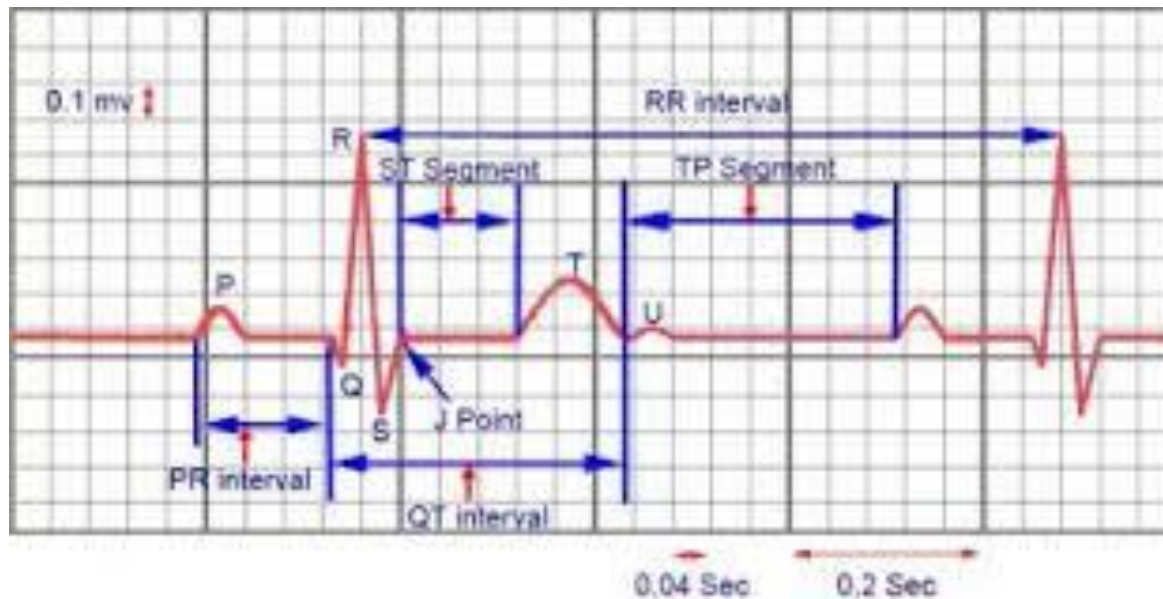
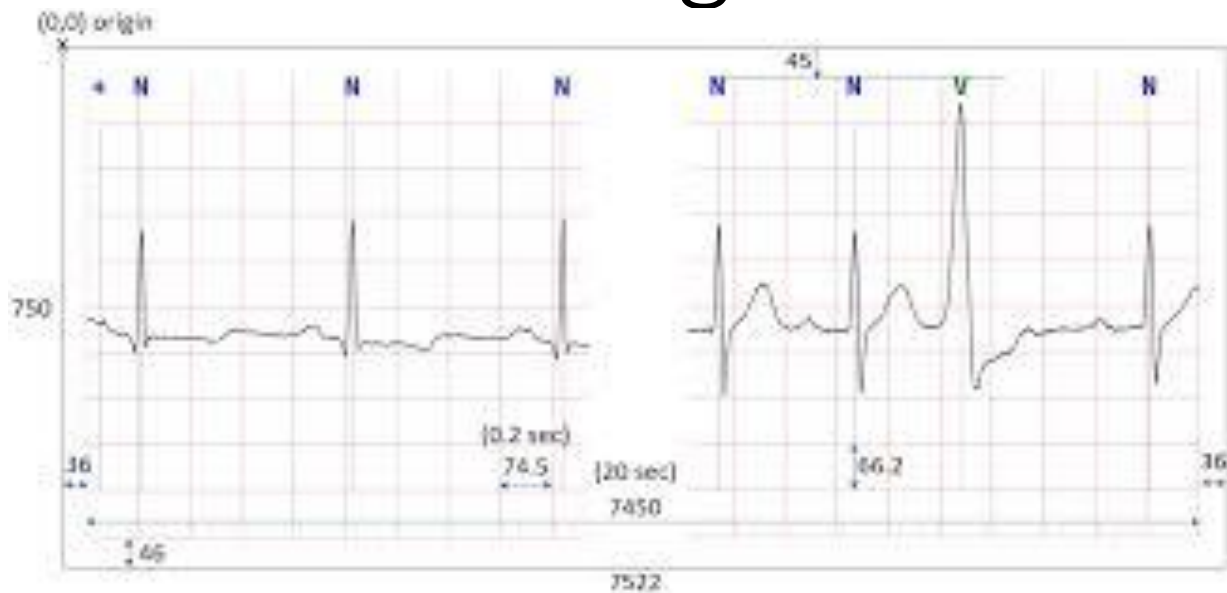
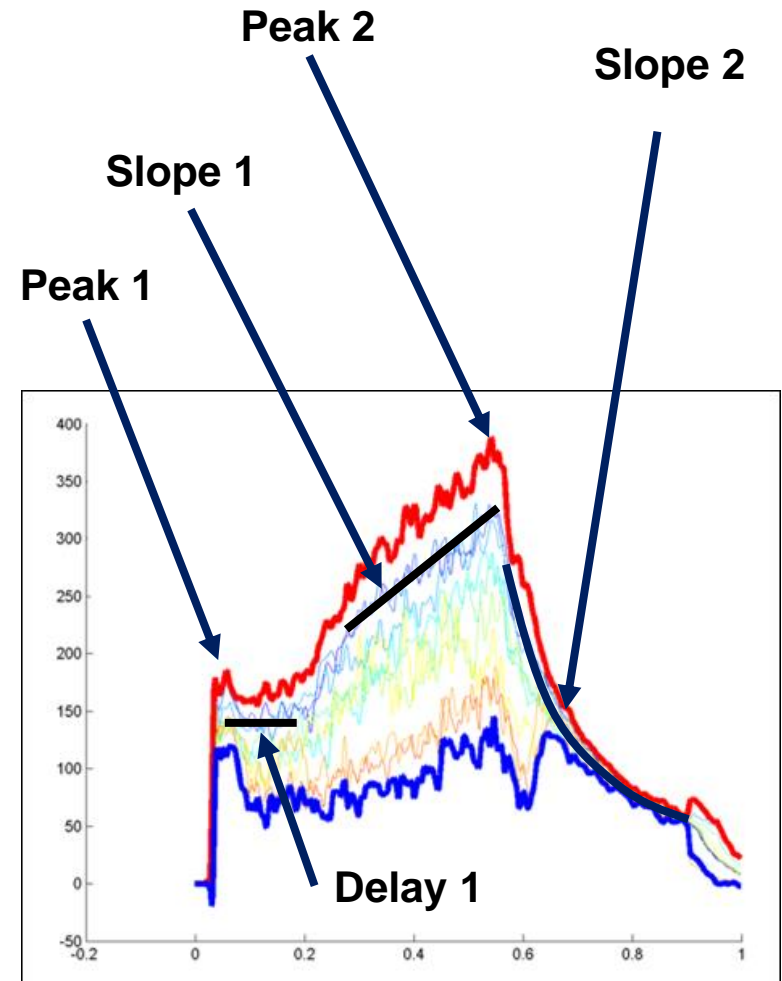


Figure from David Harrison of Lockheed Martin Corporation in 2019 DATAWorks Presentation

## Applying Functional Data Analysis throughout Aerospace Testing

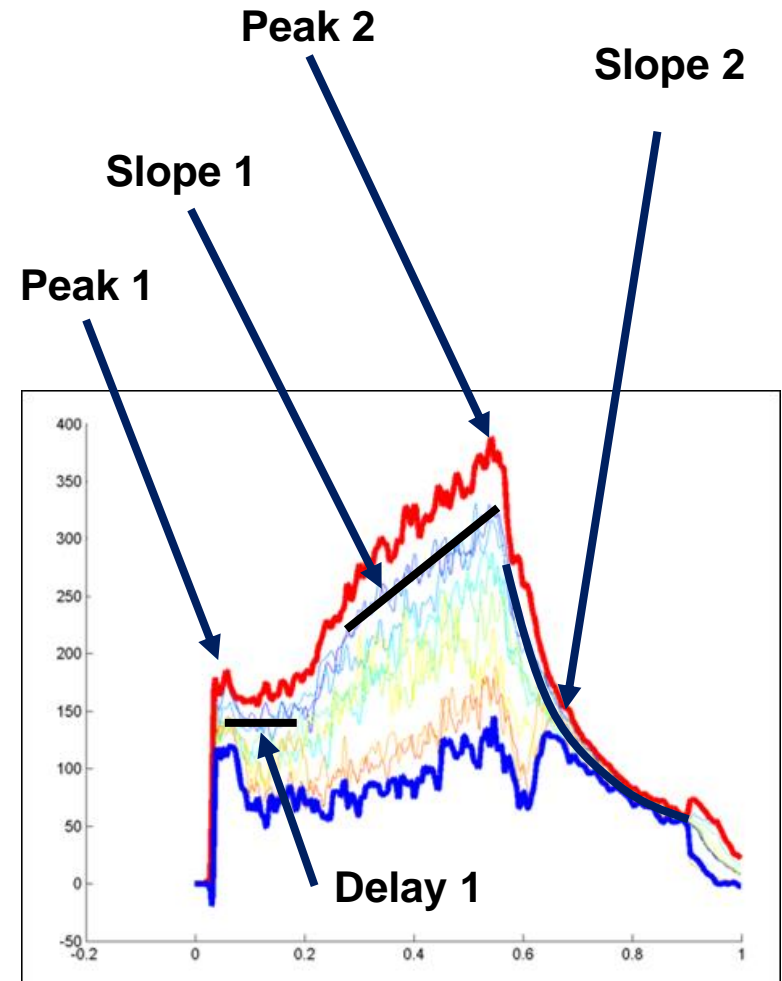
[Link to Abstract:](#)

Sensors abound in aerospace testing and while many scientists look at the data from a physics perspective, the comparative statistics information is what drives decisions. **A multi-company project was comparing launch data from the 1980's to a current set of data that included 30 sensors.** Each sensor was designed to gather 3000 data points during the 3-second launch event. The data included temperature, acceleration, and pressure information. This talk will compare the data analysis methods developed for this project as well as the use of the new Functional Data Analysis tool within JMP for its ability to discern in-family launch performances.



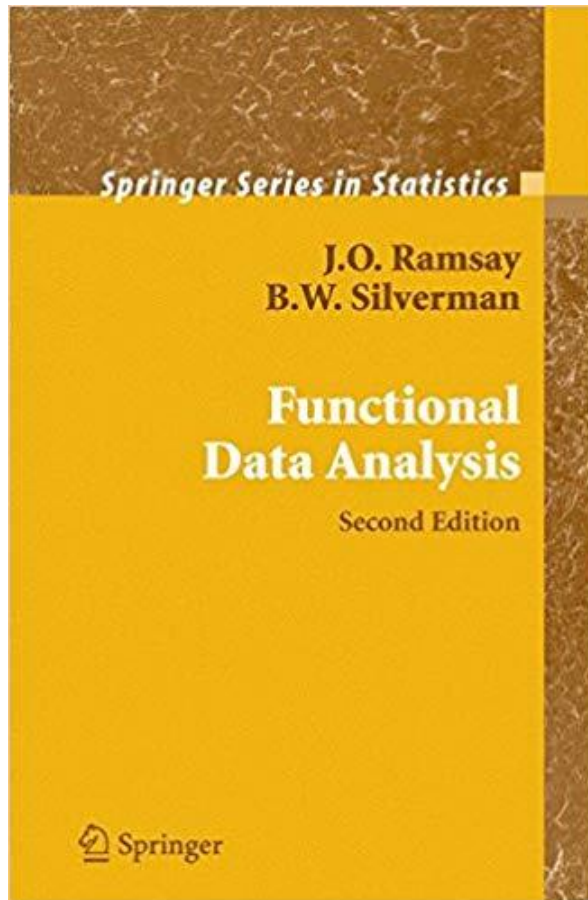
# Analysis Method Overview: Data Landmarks

- Perform distinct statistical analysis on each “landmark”- various peaks, slopes, delay
- Landmarks from new tests were compared to previous runs
- Does not use all the data!
- ***Fairly effective non-FDA option, but FDA improved the analysis***

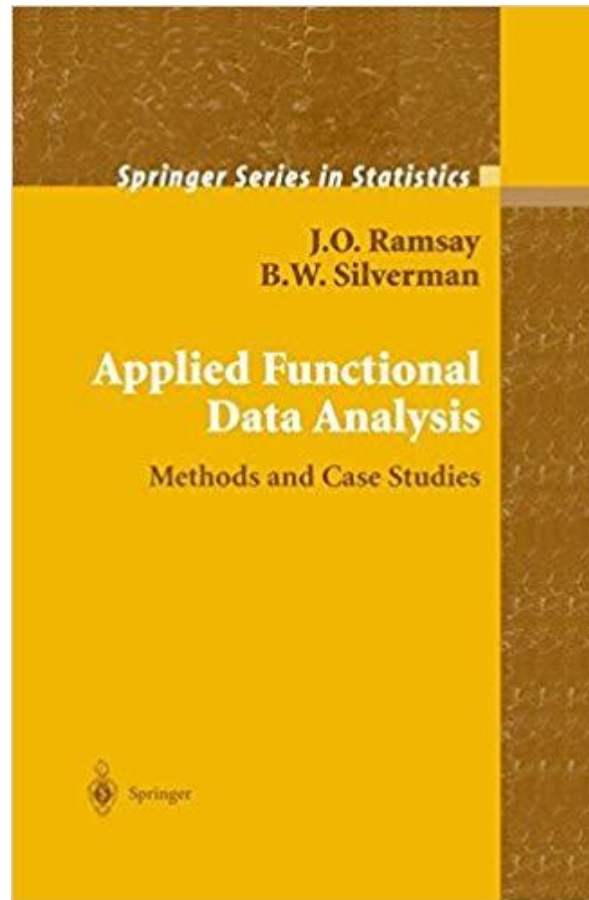




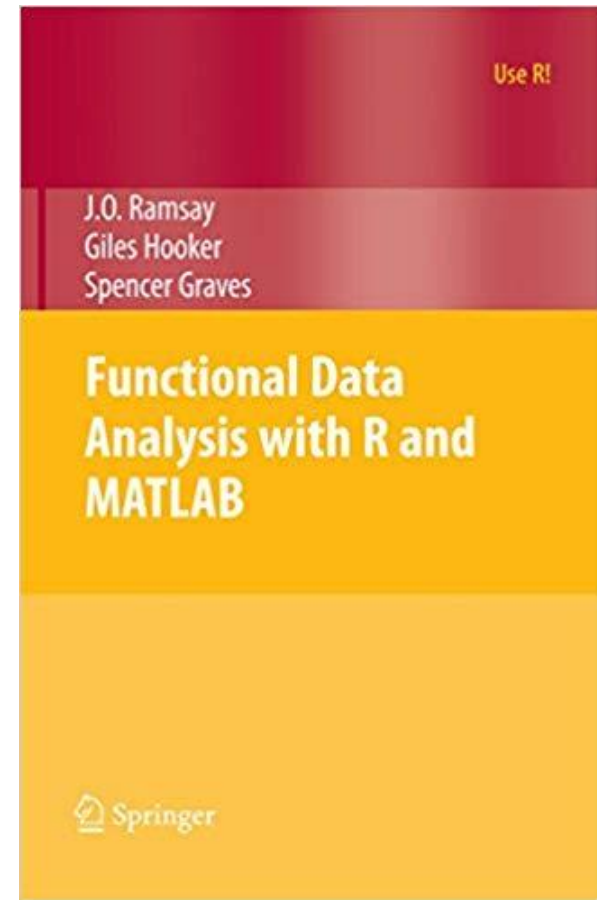
# Functional Data Analysis seminal texts by James O. Ramsay and Bernard W. Silverman



2005  
(1e1997)

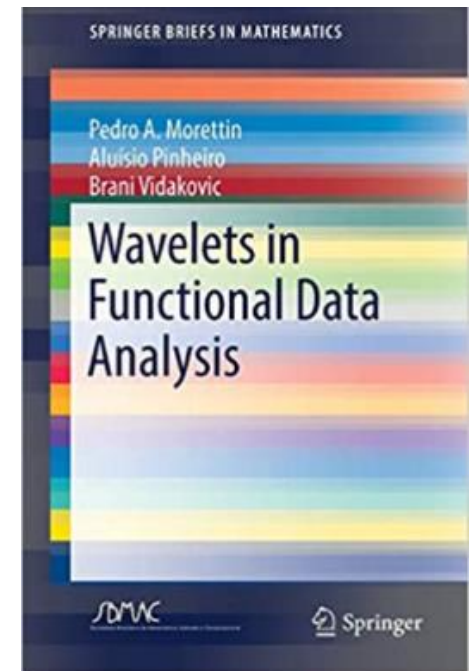
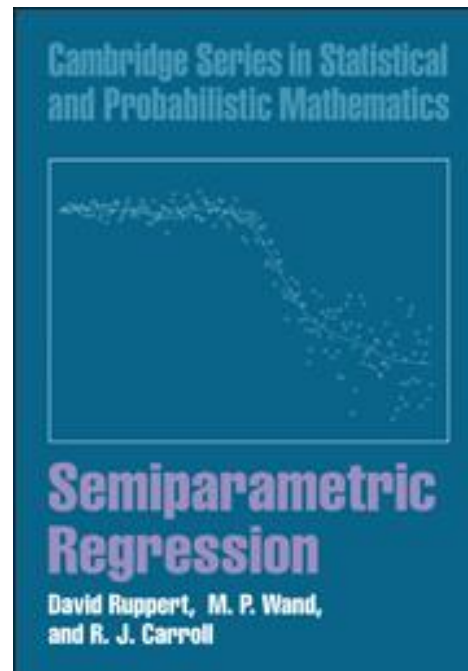
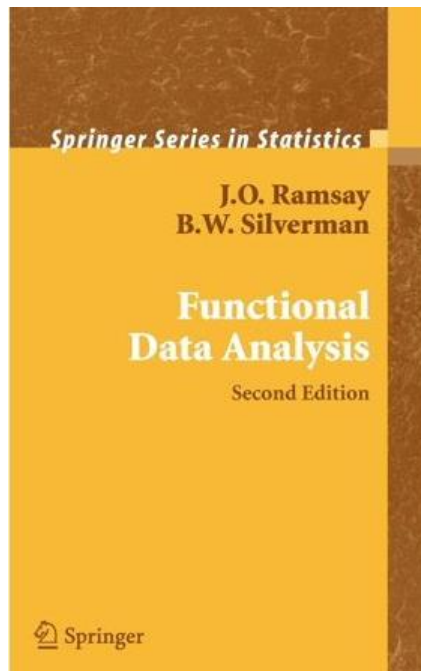


2005



2009

# FUNCTIONAL DATA EXPLORER









# What is Functional Data Analysis?

**Functional data analysis (FDA)** is a branch of statistics that analyzes data providing information about **curves, surfaces** or anything else **varying over a continuum**. In its most general form, under an FDA framework each sample element is considered to be a **function**.

Traditional Rectangular Data

|   | Batch | X1   | X2   | Y    |
|---|-------|------|------|------|
| 1 | 001   | 1.00 | 1.00 | 2.17 |
| 2 | 002   | 0.94 | 1.01 | 0.00 |
| 3 | 003   | 1.06 | 1.01 | 2.70 |
| 4 | 004   | 0.94 | 0.99 | 0.26 |
| 5 | 005   | 1.06 | 0.99 | 2.87 |
| 6 | 006   | 1.00 | 1.00 | 1.97 |

Functional Data

|   | Batch | X1  | X2   | Y   |
|---|-------|-----|------|---|
| 1 | 001   | 5.6 | 6.5  |    |
| 2 | 002   | 5.3 | 8.15 |    |
| 3 | 003   | 8.3 | 6.85 |   |
| 4 | 004   | 6.9 | 7.6  |  |

The **curve** is the fundamental unit of observation

Functional Data can also be Xs.  
When one has curves as outputs  
of a DOE they are usually the Ys.

# Two Ways to Use Functional Data Analysis

1. **Functional Response DOE (F-DOE):** Goal is to use DOE factors to predict the functional response – the *curve*
2. **Functional Response Machine Learning (F-ML):** Goal is to use the functional data – *i.e. the **curve(s)*** – to predict something
  - a) yield of a batch
  - b) probability of detection / failure / hit

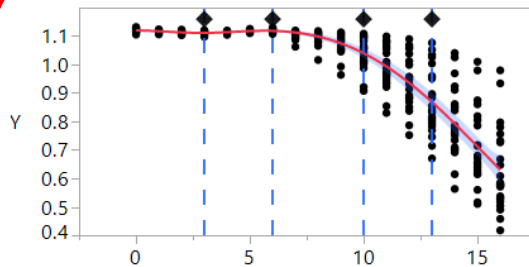
# Functional Data Analysis

- F-DOE & F-ML use functional principal components analysis (F-PCA)
- F-PCA breaks the data into **FPC Scores** and **Eigenfunctions** in a dimension reduction that is closely analogous to classical PCA
- FPC Scores are scalars that explain **function-to-function variation**
- Eigenfunctions explain the **longitudinal variation** (e.g., **time**)
- We fit models with the FPC scores, cluster them, graph them - **just like any other continuous data**
- For F-DOE we **fit the FPC scores as functions of the DOE factors** using (FPC score) X (Eigenfunctions) as intermediate formulas, and (Modeled FPC score) X (Eigenfunctions) as final prediction formula

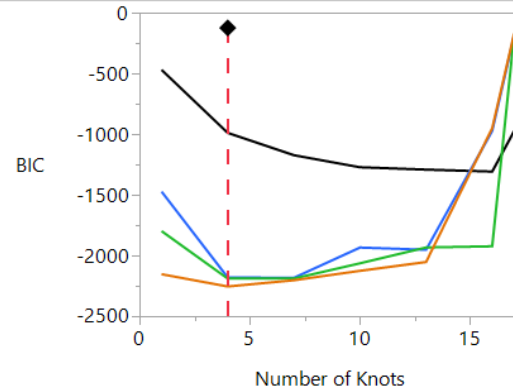
# How do we analyze Functional data?

1. Convert streams of data into a function - Fit Splines or Fourier basis functions
2. Create Functional Principal Components of the basis function - do F-PCA

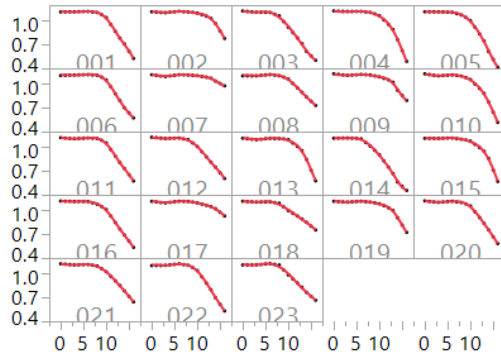
## Model Selection



Legend  
— Prediction  
— Knots



Legend  
— Step Functions  
— Linear  
— Quadratic  
— Cubic

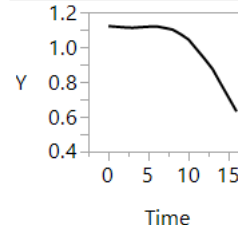


## Summaries

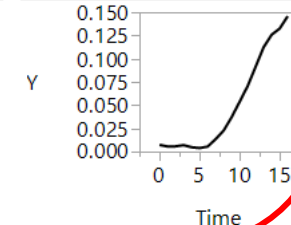
### Overall

|                    |           |
|--------------------|-----------|
| Observations       | 391       |
| Functions          | 23        |
| Mean               | 1.0044514 |
| Standard Deviation | 0.1700357 |
| Minimum            | 0.4175881 |
| Maximum            | 1.1329135 |

### Mean



### Standard Deviation

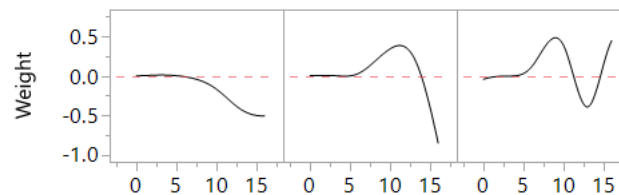


## Function Summaries

| Batch | FPC 1     | FPC 2     | FPC 3     |
|-------|-----------|-----------|-----------|
| 001   | 0.1762753 | 0.0249696 | 0.0391766 |
| 002   | -0.353504 | 0.0180562 | -0.019895 |
| 003   | 0.3748338 | -0.091577 | -0.032289 |
| 004   | 0.0587479 | 0.1194643 | -0.022659 |
| 005   | 0.3493758 | 0.0517594 | -0.011833 |
| 006   | 0.1462727 | 0.0217465 | 0.056155  |
| 007   | -0.534657 | -0.078567 | 0.0063818 |
| 008   | -0.11289  | -0.033664 | 0.0282277 |
| 009   | -0.316959 | 0.0164104 | -0.01071  |
| 010   | 0.0477899 | 0.086351  | -0.019852 |
| 011   | 0.1077061 | 0.0210358 | 0.0472042 |
| 012   | 0.0991761 | -0.041143 | 0.0102085 |
| 013   | -0.112228 | 0.1139909 | -0.024613 |
| 014   | 0.4828437 | -0.104326 | -0.060896 |
| 015   | -0.101294 | 0.0901209 | -0.037628 |
| 016   | 0.2139988 | -0.01602  | 0.0122833 |
| 017   | -0.479219 | -0.073566 | 0.0049346 |
| 018   | -0.099709 | -0.112664 | -0.017042 |
| 019   | -0.254651 | 0.034425  | -0.024642 |
| 020   | 0.0639687 | 0.0282532 | 0.0239146 |
| 021   | 0.0192996 | -0.03352  | 0.0063906 |
| 022   | 0.167006  | 0.040068  | 0.0330496 |
| 023   | 0.0578164 | -0.081604 | 0.0141323 |

## Functional PCA

| FPC | Eigenvalue | 20 40 60 80 | Percent | Cumulative |
|-----|------------|-------------|---------|------------|
| 1   | 0.06855    |             | 92.2%   | 92.2%      |
| 2   | 0.00480    |             | 6.46%   | 98.7%      |
| 3   | 0.00087    |             | 1.17%   | 99.8%      |

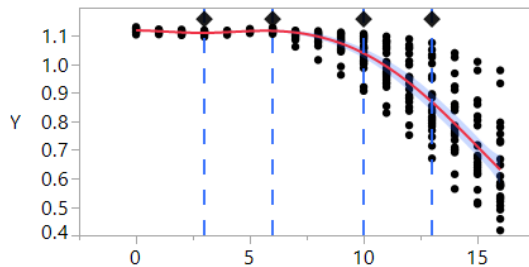


Eigenfunction 1    Eigenfunction 2    Eigenfunction 3

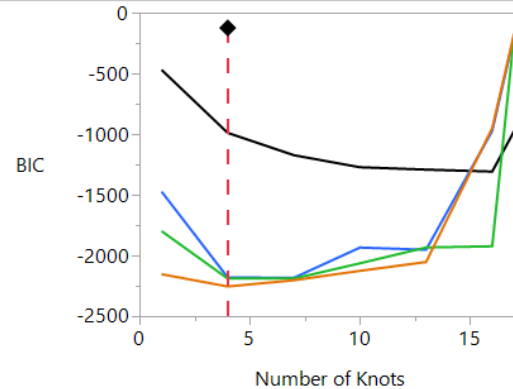
# How do we analyze Functional data?

1. Convert streams of data into a function - Fit Splines or Fourier basis functions
2. Create Functional Principal Components of the basis function – do F-PCA

## Model Selection



Legend  
— Prediction  
— Knots



Legend  
— Step Functions  
— Linear  
— Quadratic  
— Cubic

## Functional Principal Component Scores

### Function Summaries

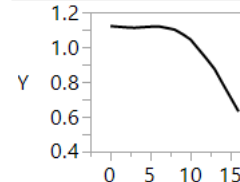
| Batch | FPC 1     | FPC 2     | FPC 3     |
|-------|-----------|-----------|-----------|
| 001   | 0.1762753 | 0.0249696 | 0.0391766 |
| 002   | -0.353504 | 0.0180562 | -0.019895 |
| 003   | 0.3748338 | -0.091577 | -0.032289 |
| 004   | 0.0587479 | 0.1194643 | -0.022659 |
| 005   | 0.3493758 | 0.0517594 | -0.011833 |
| 006   | 0.1462727 | 0.0217465 | 0.056155  |
| 007   | -0.534657 | -0.078567 | 0.0063818 |
| 008   | -0.11289  | -0.033664 | 0.0282277 |
| 009   | -0.316959 | 0.0164104 | -0.01071  |
| 010   | 0.0477899 | 0.086351  | -0.019852 |
| 011   | 0.1077061 | 0.0210358 | 0.0472042 |
| 012   | 0.0991761 | -0.041143 | 0.0102085 |
| 013   | -0.112228 | 0.1139909 | -0.024613 |
| 014   | 0.4828437 | -0.104326 | -0.060896 |
| 015   | -0.101294 | 0.0901209 | -0.037628 |
| 016   | 0.2139988 | -0.01602  | 0.0122833 |
| 017   | -0.479219 | -0.073566 | 0.0049346 |
| 018   | -0.099709 | -0.112664 | -0.017042 |
| 019   | -0.254651 | 0.034425  | -0.024642 |
| 020   | 0.0639687 | 0.0282532 | 0.0239146 |
| 021   | 0.0192996 | -0.03352  | 0.0063906 |
| 022   | 0.167006  | 0.040068  | 0.0330496 |
| 023   | 0.0578164 | -0.081604 | 0.0141323 |

## Summaries

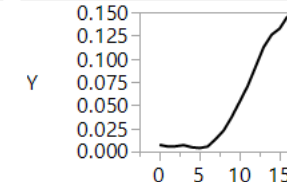
### Overall

|                    |           |
|--------------------|-----------|
| Observations       | 391       |
| Functions          | 23        |
| Mean               | 1.0044514 |
| Standard Deviation | 0.1700357 |
| Minimum            | 0.4175881 |
| Maximum            | 1.1329135 |

### Mean

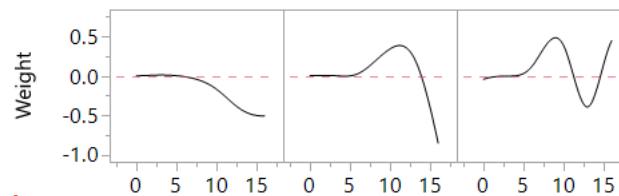


### Standard Deviation



## Functional PCA

| FPC | Eigenvalue | 20 40 60 80 | Percent | Cumulative |
|-----|------------|-------------|---------|------------|
| 1   | 0.06855    |             | 92.2%   | 92.2%      |
| 2   | 0.00480    |             | 6.46%   | 98.7%      |
| 3   | 0.00087    |             | 1.17%   | 99.8%      |



## Eigenfunctions

Eigenfunction 1    Eigenfunction 2    Eigenfunction 3

# How do we analyze Functional data?

3. Eigenfunctions explain the longitudinal variation.

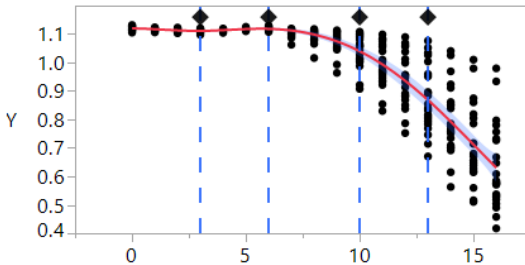
4. Function Summaries (FPC scores) explain function-to-function variation.

Explains  
function-to-  
function  
variation

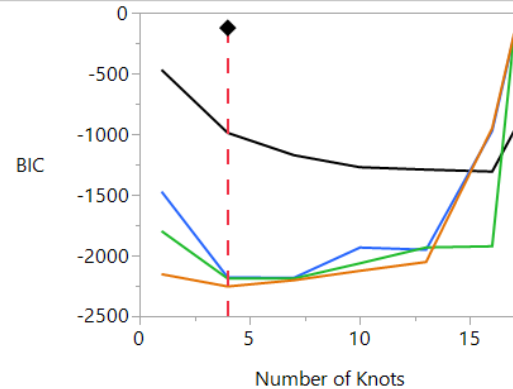
Function Summaries

| Batch | FPC 1     | FPC 2     | FPC 3     |
|-------|-----------|-----------|-----------|
| 001   | 0.1762753 | 0.0249696 | 0.0391766 |
| 002   | -0.353504 | 0.0180562 | -0.019895 |
| 003   | 0.3748338 | -0.091577 | -0.032289 |
| 004   | 0.0587479 | 0.1194643 | -0.022659 |
| 005   | 0.3493758 | 0.0517594 | -0.011833 |
| 006   | 0.1462727 | 0.0217465 | 0.056155  |
| 007   | -0.534657 | -0.078567 | 0.0063818 |
| 008   | -0.11289  | -0.033664 | 0.0282277 |
| 009   | -0.316959 | 0.0164104 | -0.01071  |
| 010   | 0.0477899 | 0.086351  | -0.019852 |
| 011   | 0.1077061 | 0.0210358 | 0.0472042 |
| 012   | 0.0991761 | -0.041143 | 0.0102085 |
| 013   | -0.112228 | 0.1139909 | -0.024613 |
| 014   | 0.4828437 | -0.104326 | -0.060896 |
| 015   | -0.101294 | 0.0901209 | -0.037628 |
| 016   | 0.2139988 | -0.01602  | 0.0122833 |
| 017   | -0.479219 | -0.073566 | 0.0049346 |
| 018   | -0.099709 | -0.112664 | -0.017042 |
| 019   | -0.254651 | 0.034425  | -0.024642 |
| 020   | 0.0639687 | 0.0282532 | 0.0239146 |
| 021   | 0.0192996 | -0.03352  | 0.0063906 |
| 022   | 0.167006  | 0.040068  | 0.0330496 |
| 023   | 0.0578164 | -0.081604 | 0.0141323 |

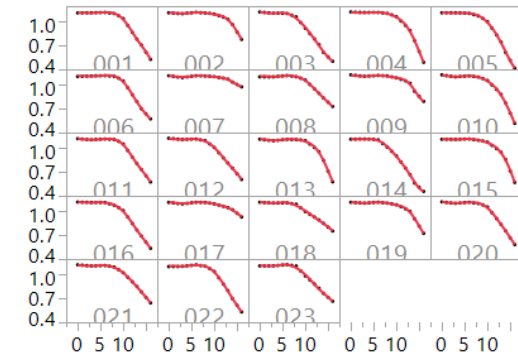
Model Selection



Legend  
— Prediction  
— Knots



Legend  
— Step Functions  
— Linear  
— Quadratic  
— Cubic

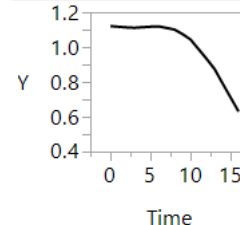


Summaries

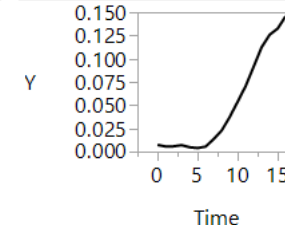
Overall

|                    |           |
|--------------------|-----------|
| Observations       | 391       |
| Functions          | 23        |
| Mean               | 1.0044514 |
| Standard Deviation | 0.1700357 |
| Minimum            | 0.4175881 |
| Maximum            | 1.1329135 |

Mean

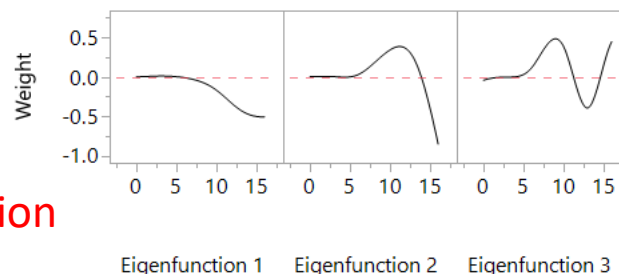


Standard Deviation



Functional PCA

| FPC | Eigenvalue | 20 40 60 80 | Percent | Cumulative |
|-----|------------|-------------|---------|------------|
| 1   | 0.06855    |             | 92.2%   | 92.2%      |
| 2   | 0.00480    |             | 6.46%   | 98.7%      |
| 3   | 0.00087    |             | 1.17%   | 99.8%      |



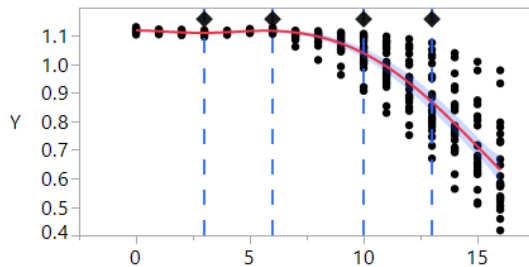
Explains longitudinal variation

# How do we analyze Functional data?

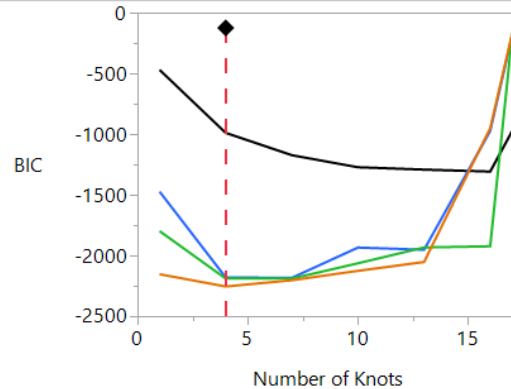
3. Eigenfunctions explain the longitudinal variation.

4. Function Summaries (FPC scores) explain function-to-function variation.

## Model Selection

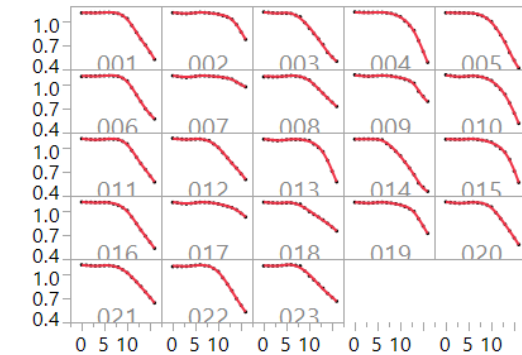


Legend  
— Prediction  
— Knots



Legend  
— Step Functions  
— Linear  
— Quadratic  
— Cubic

Explains  
function-to-  
function  
variation

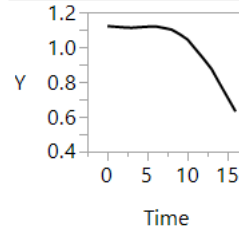


## Summaries

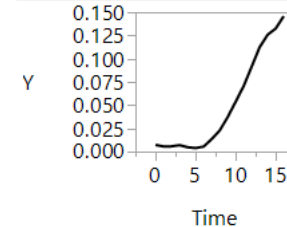
### Overall

|                    |           |
|--------------------|-----------|
| Observations       | 391       |
| Functions          | 23        |
| Mean               | 1.0044514 |
| Standard Deviation | 0.1700357 |
| Minimum            | 0.4175881 |
| Maximum            | 1.1329135 |

### Mean

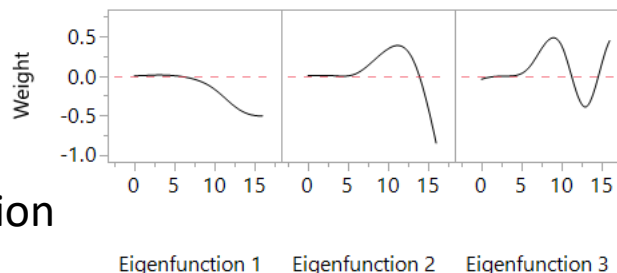


### Standard Deviation



## Functional PCA

| FPC | Eigenvalue | 20 40 60 80 | Percent | Cumulative |
|-----|------------|-------------|---------|------------|
| 1   | 0.06855    |             | 92.2%   | 92.2%      |
| 2   | 0.00480    |             | 6.46%   | 98.7%      |
| 3   | 0.00087    |             | 1.17%   | 99.8%      |



Explains longitudinal variation

## Function Summaries

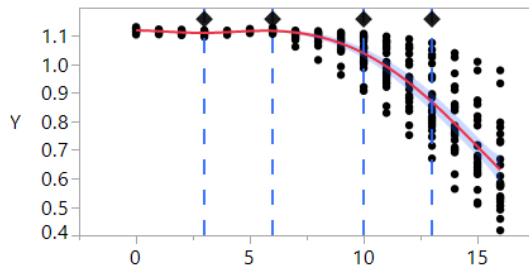
| Batch | FPC 1     | FPC 2     | FPC 3     |
|-------|-----------|-----------|-----------|
| 001   | 0.1762753 | 0.0249696 | 0.0391766 |
| 002   | -0.353504 | 0.0180562 | -0.019895 |
| 003   | 0.3748338 | -0.091577 | -0.032289 |
| 004   | 0.0587479 | 0.1194643 | -0.022659 |
| 005   | 0.3493758 | 0.0517594 | -0.011833 |
| 006   | 0.1462727 | 0.0217465 | 0.056155  |
| 007   | -0.534657 | -0.078567 | 0.0063818 |
| 008   | -0.11289  | -0.033664 | 0.0282277 |
| 009   | -0.316959 | 0.0164104 | -0.01071  |
| 010   | 0.0477899 | 0.086351  | -0.019852 |
| 011   | 0.1077061 | 0.0210358 | 0.0472042 |
| 012   | 0.0991761 | -0.041143 | 0.0102085 |
| 013   | -0.112228 | 0.1139909 | -0.024613 |
| 014   | 0.4828437 | -0.104326 | -0.060896 |
| 015   | -0.101294 | 0.0901209 | -0.037628 |
| 016   | 0.2139988 | -0.01602  | 0.0122833 |
| 017   | -0.479219 | -0.073566 | 0.0049346 |
| 018   | -0.099709 | -0.112664 | -0.017042 |
| 019   | -0.254651 | 0.034425  | -0.024642 |
| 020   | 0.0639687 | 0.0282532 | 0.0239146 |
| 021   | 0.0192996 | -0.03352  | 0.0063906 |
| 022   | 0.167006  | 0.040068  | 0.0330496 |
| 023   | 0.0578164 | -0.081604 | 0.0141323 |



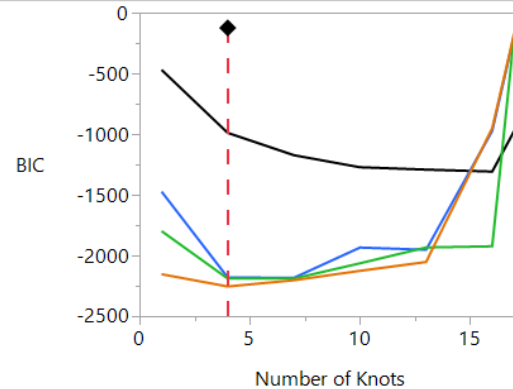
# How do we analyze Functional data?

5. Products of FPC scores multiplying their corresponding eigenfunctions, when added to the Mean closely reproduce the individual function (batch) curves.

## Model Selection

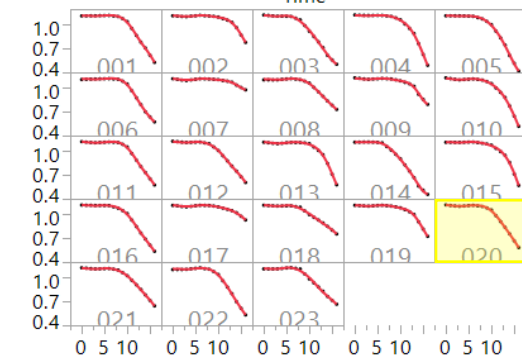


Legend  
— Prediction  
— Knots



Legend  
— Step Functions  
— Linear  
— Quadratic  
— Cubic

Explains  
function-to-  
function  
variation

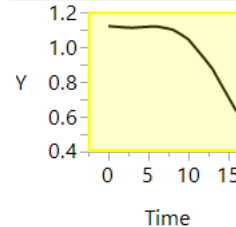


## Summaries

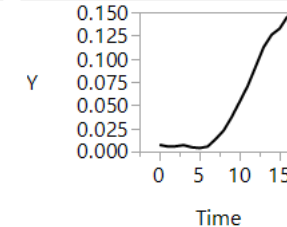
### Overall

|                    |           |
|--------------------|-----------|
| Observations       | 391       |
| Functions          | 23        |
| Mean               | 1.0044514 |
| Standard Deviation | 0.1700357 |
| Minimum            | 0.4175881 |
| Maximum            | 1.1329135 |

### Mean

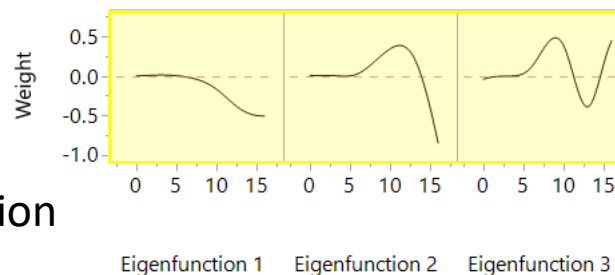


### Standard Deviation



## Functional PCA

| FPC | Eigenvalue | 20 40 60 80 | Percent | Cumulative |
|-----|------------|-------------|---------|------------|
| 1   | 0.06855    |             | 92.2%   | 92.2%      |
| 2   | 0.00480    |             | 6.46%   | 98.7%      |
| 3   | 0.00087    |             | 1.17%   | 99.8%      |

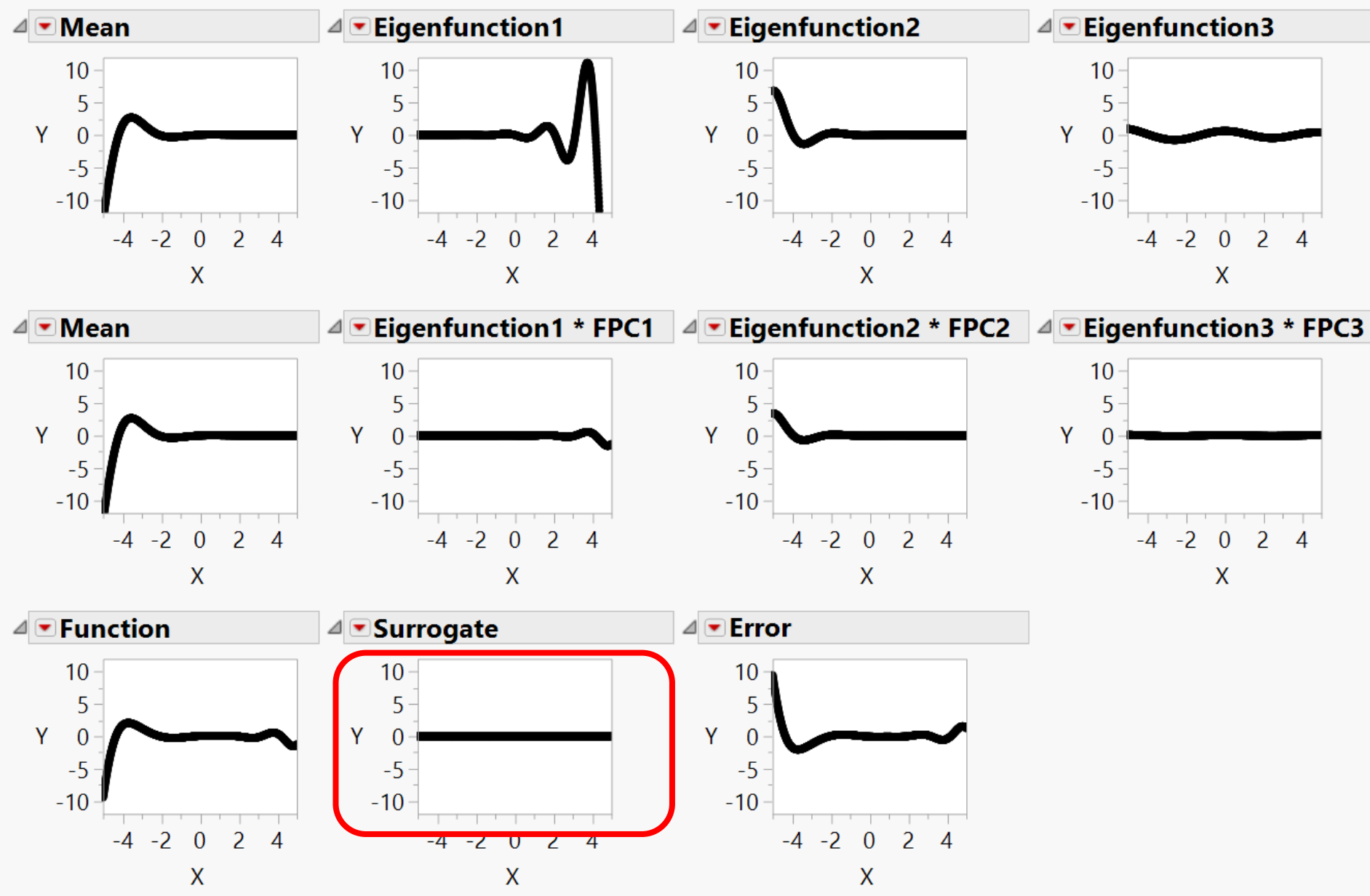


Explains longitudinal variation

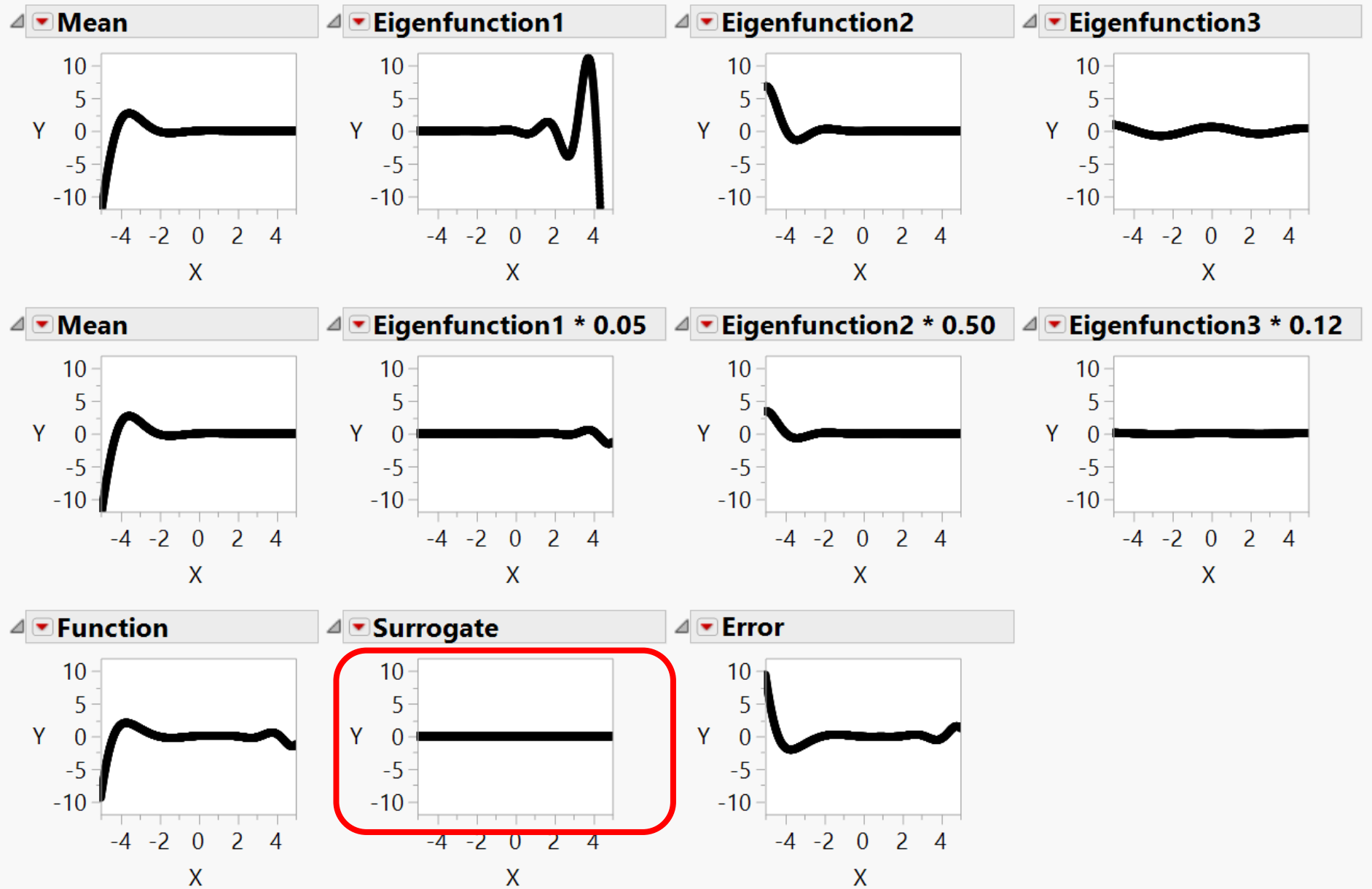
## Function Summaries

| Batch | FPC 1     | FPC 2     | FPC 3     |
|-------|-----------|-----------|-----------|
| 001   | 0.1762753 | 0.0249696 | 0.0391766 |
| 002   | -0.353504 | 0.0180562 | -0.019895 |
| 003   | 0.3748338 | -0.091577 | -0.032289 |
| 004   | 0.0587479 | 0.1194643 | -0.022659 |
| 005   | 0.3493758 | 0.0517594 | -0.011833 |
| 006   | 0.1462727 | 0.0217465 | 0.056155  |
| 007   | -0.534657 | -0.078567 | 0.0063818 |
| 008   | -0.11289  | -0.033664 | 0.0282277 |
| 009   | -0.316959 | 0.0164104 | -0.01071  |
| 010   | 0.0477899 | 0.086351  | -0.019852 |
| 011   | 0.1077061 | 0.0210358 | 0.0472042 |
| 012   | 0.0991761 | -0.041143 | 0.0102085 |
| 013   | -0.112228 | 0.1139909 | -0.024613 |
| 014   | 0.4828437 | -0.104326 | -0.060896 |
| 015   | -0.101294 | 0.0901209 | -0.037628 |
| 016   | 0.2139988 | -0.01602  | 0.0122833 |
| 017   | -0.479219 | -0.073566 | 0.0049346 |
| 018   | -0.099709 | -0.112664 | -0.017042 |
| 019   | -0.254651 | 0.034425  | -0.024642 |
| 020   | 0.0639687 | 0.0282532 | 0.0239146 |
| 021   | 0.0192996 | -0.03352  | 0.0063906 |
| 022   | 0.167006  | 0.040068  | 0.0330496 |
| 023   | 0.0578164 | -0.081604 | 0.0141323 |

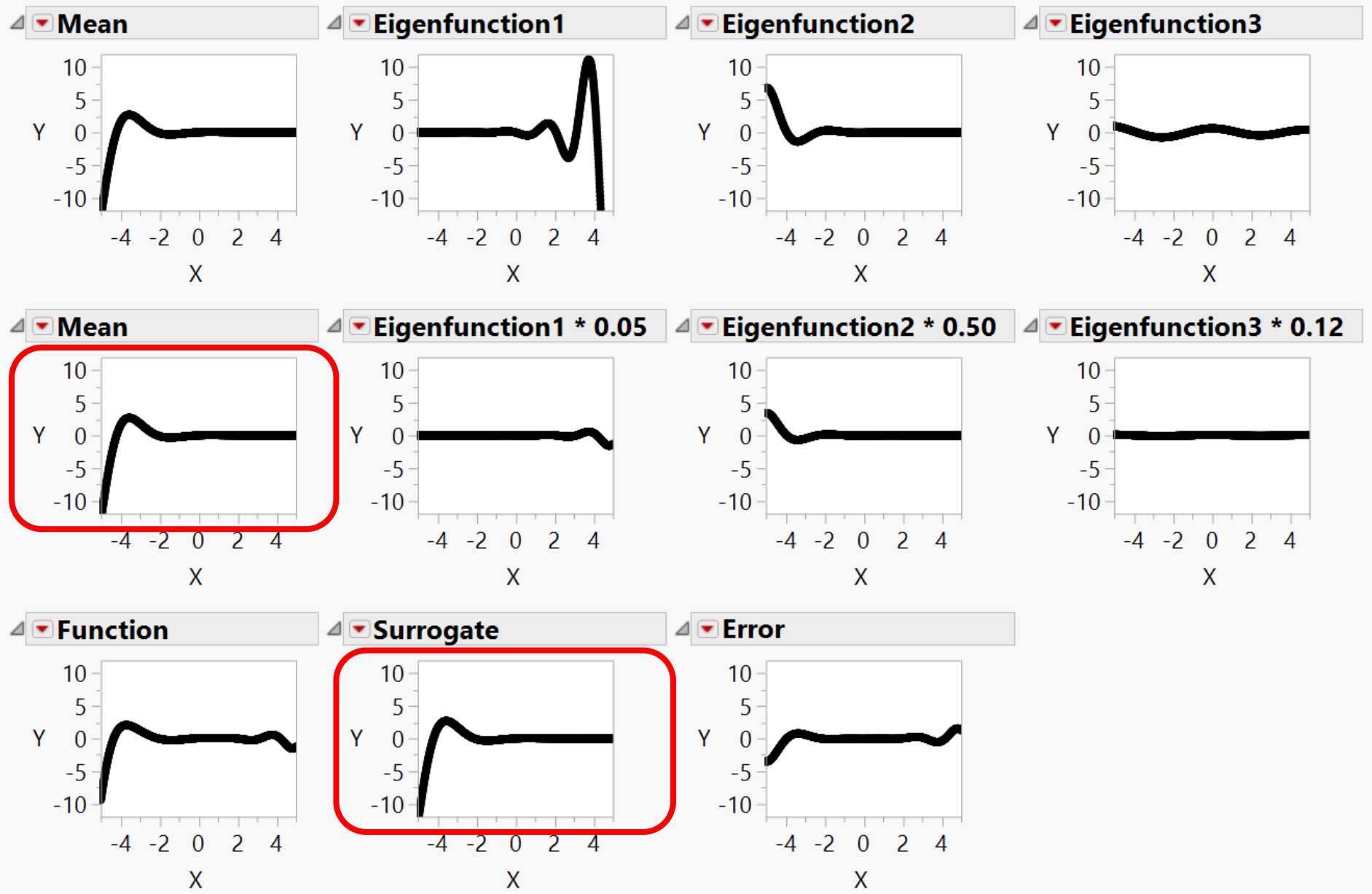




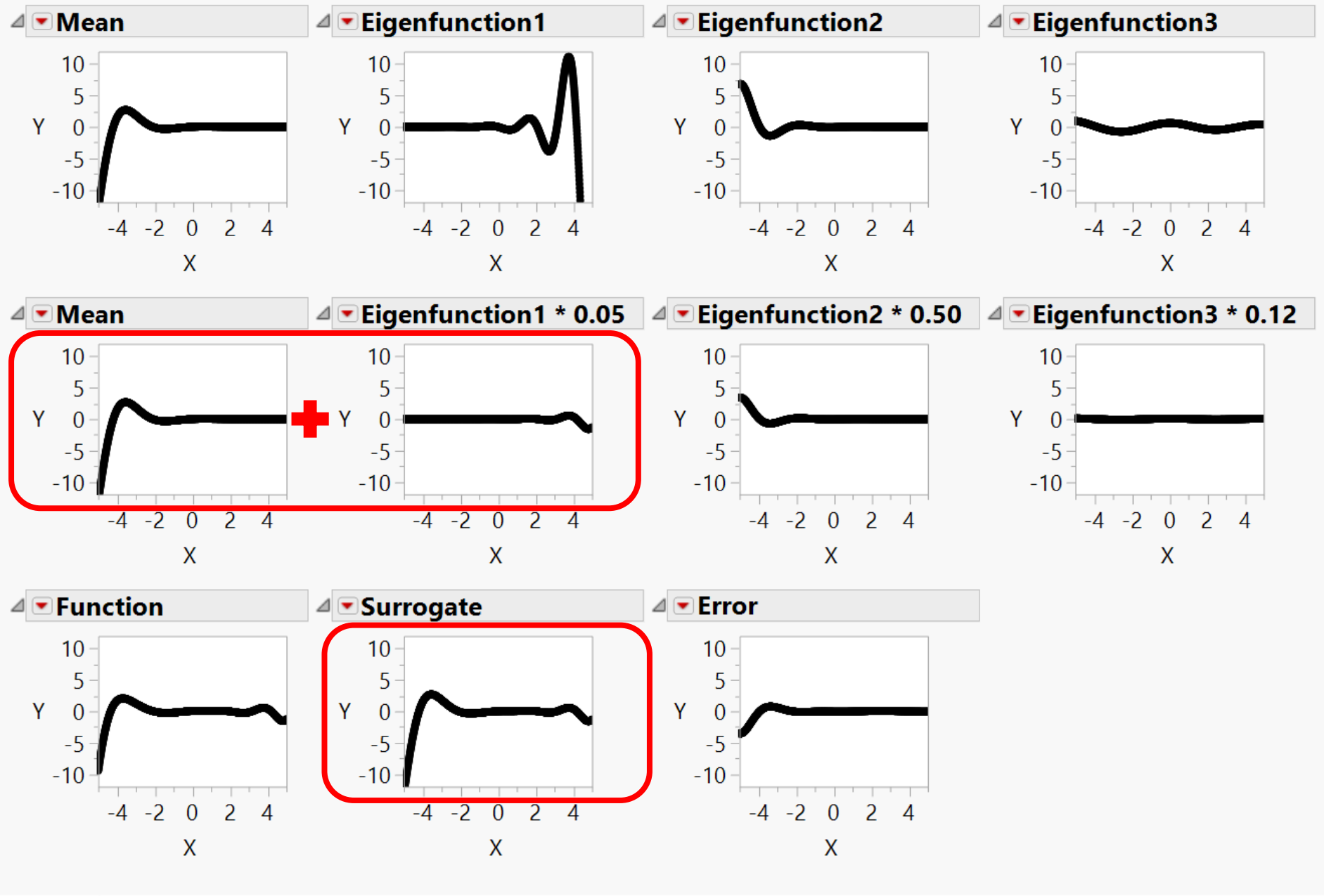
$$Y(X) = \text{Surrogate}$$



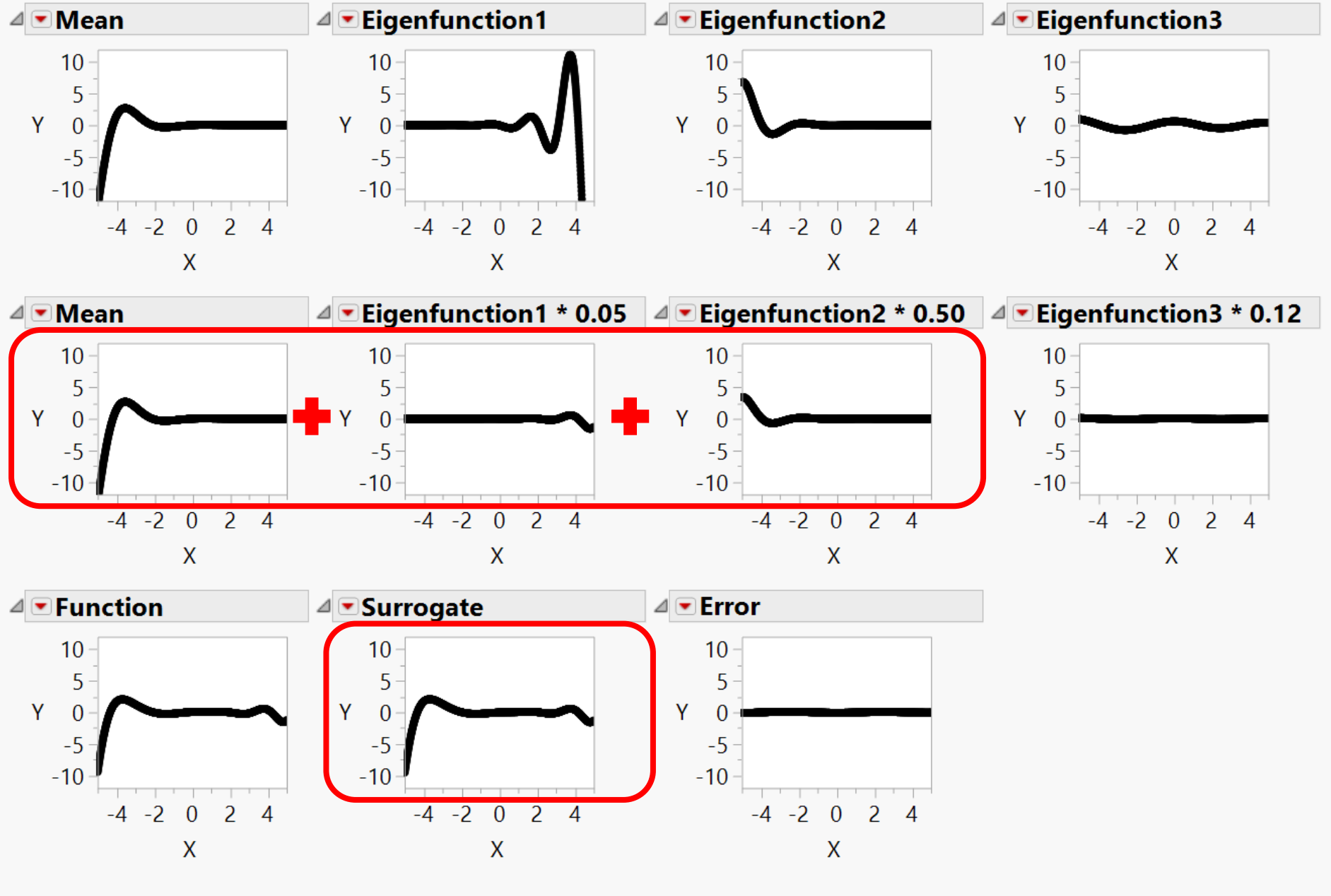
$Y(X) = \text{Surrogate}$



$Y(X) = \mu(X)$

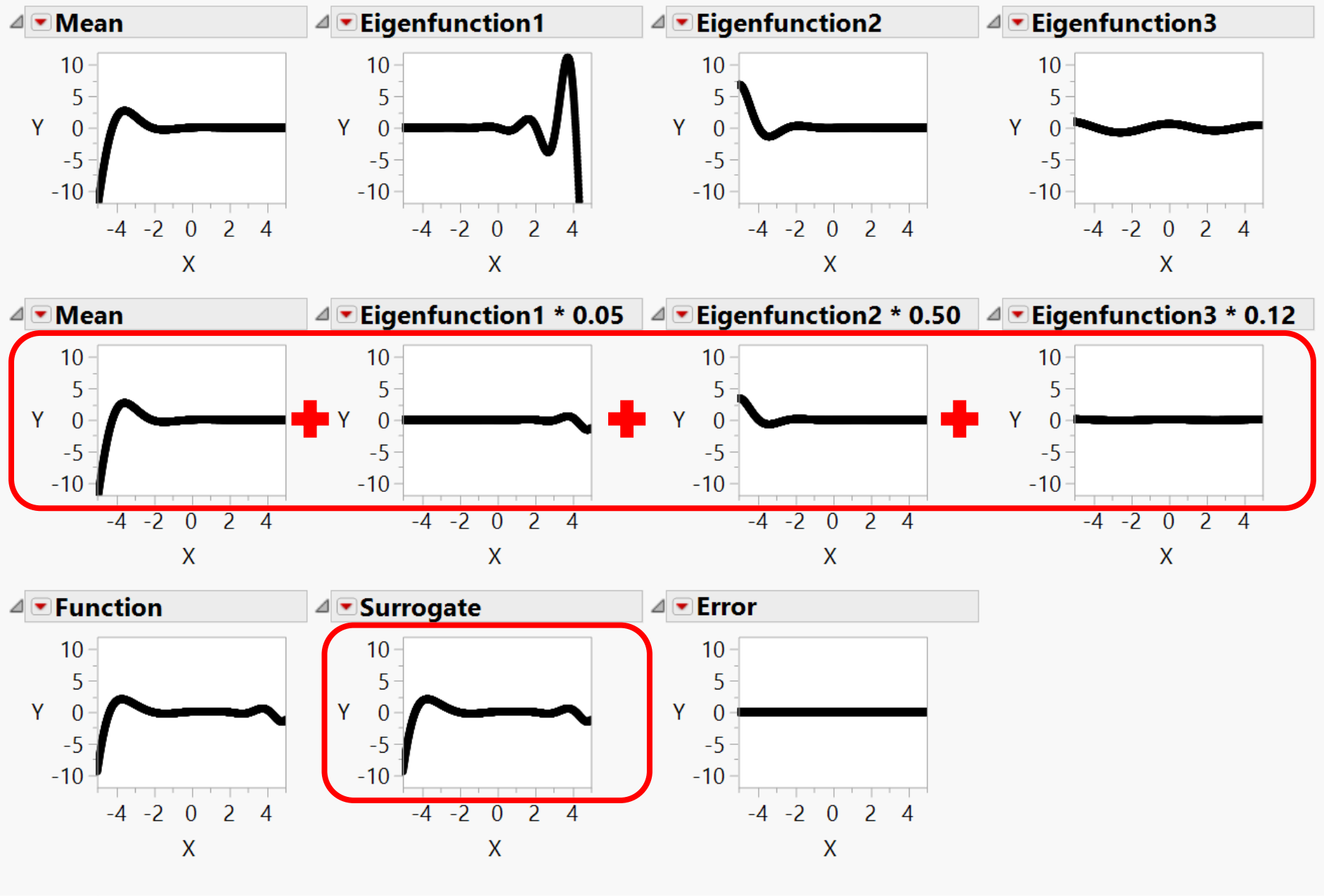


$$Y(X) = \mu(X) + 0.05 \cdot E_1(X)$$



$$Y(X) = \mu(X) + 0.05 \cdot E_1(X) + 0.50 \cdot E_2(X)$$

Functions

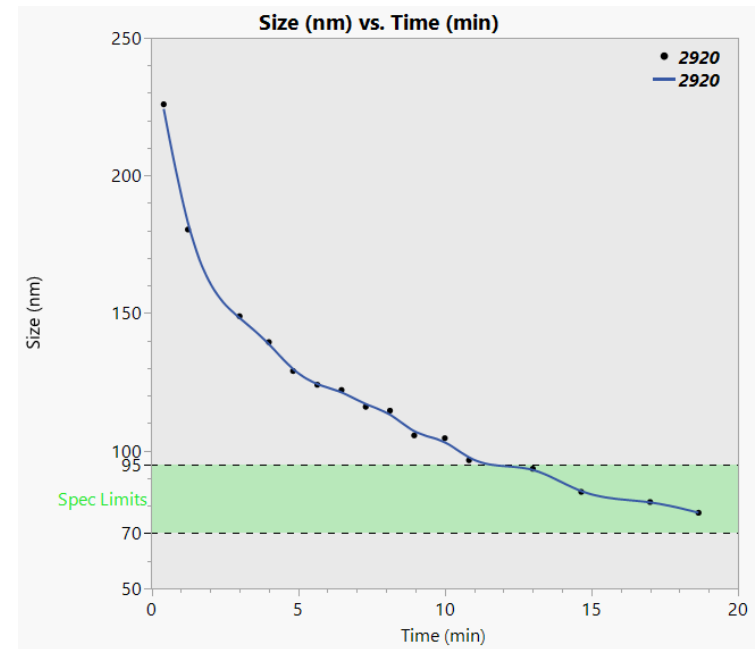
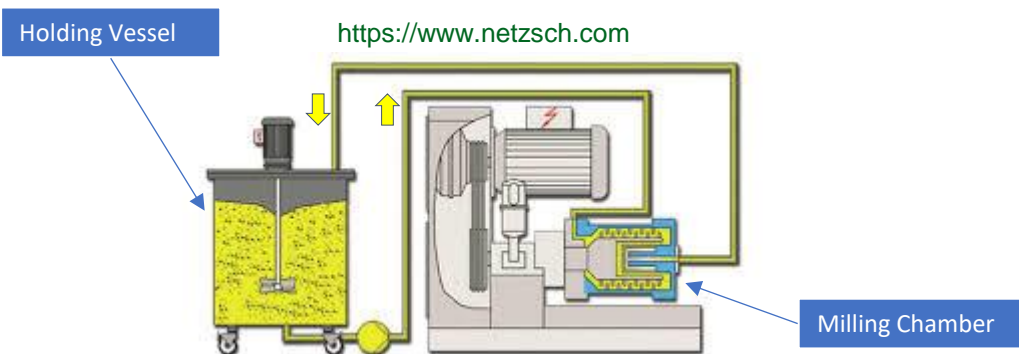


$$Y(X) = \mu(X) + 0.05 \cdot E_1(X) + 0.50 \cdot E_2(X) + 0.12 \cdot E_3(X)$$

# Simple Case Study Based on Real Data Using Functional Principal Components

FPCs efficiently summarize your functional data in a few components, but how do we use these to help analyze our data?

Example DoE response

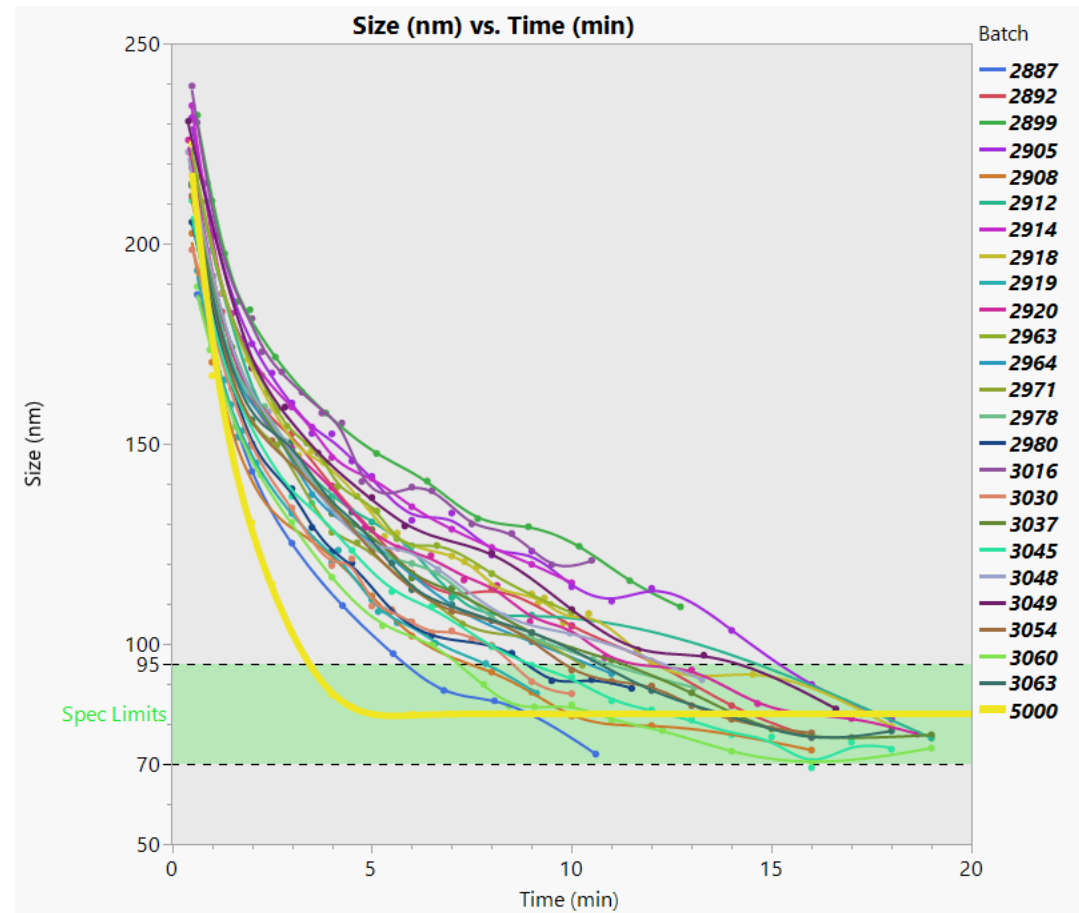


# Example DoE Response

LEFT: Definitive Screening Design plus Confirmation Trials

RIGHT: Measured Batch Profiles - Thick yellow line is “Ideal” response, aka “Golden Curve”

| Batch | Run Order | %Beads | %Active | Flow | Temperature | Trial Type   |
|-------|-----------|--------|---------|------|-------------|--------------|
| 2887  | 1         | 90     | 25      | 150  | 45          | Design       |
| 2892  | 2         | 80     | 25      | 350  | 15          | Design       |
| 2899  | 3         | 80     | 15      | 550  | 15          | Design       |
| 2905  | 4         | 80     | 15      | 150  | 45          | Design       |
| 2908  | 5         | 90     | 25      | 150  | 15          | Design       |
| 2912  | 6         | 90     | 15      | 150  | 30          | Design       |
| 2914  | 7         | 85     | 15      | 150  | 15          | Design       |
| 2918  | 8         | 90     | 15      | 550  | 15          | Design       |
| 2919  | 9         | 90     | 25      | 550  | 15          | Design       |
| 2920  | 10        | 90     | 15      | 350  | 45          | Design       |
| 2963  | 11        | 80     | 20      | 150  | 15          | Design       |
| 2964  | 12        | 85     | 20      | 350  | 30          | Design       |
| 2971  | 13        | 80     | 25      | 150  | 45          | Design       |
| 2978  | 14        | 80     | 25      | 550  | 30          | Design       |
| 2980  | 15        | 85     | 25      | 550  | 45          | Design       |
| 3016  | 16        | 80     | 15      | 550  | 45          | Design       |
| 3030  | 17        | 90     | 20      | 550  | 45          | Design       |
| 3037  | 18        | 87.5   | 17.5    | 450  | 37.5        | Confirmation |
| 3045  | 19        | 87.5   | 22.5    | 450  | 22.5        | Confirmation |
| 3048  | 20        | 87.5   | 17.5    | 250  | 22.5        | Confirmation |
| 3049  | 21        | 82.5   | 17.5    | 450  | 22.5        | Confirmation |
| 3054  | 22        | 82.5   | 22.5    | 250  | 37.5        | Confirmation |
| 3060  | 23        | 90     | 25      | 550  | 45          | Confirmation |
| 3063  | 24        | 85     | 20      | 350  | 30          | Confirmation |
| 5000  | 25        | •      | •       | •    | •           | Confirmation |



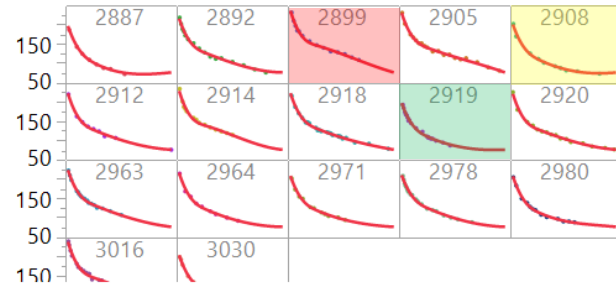
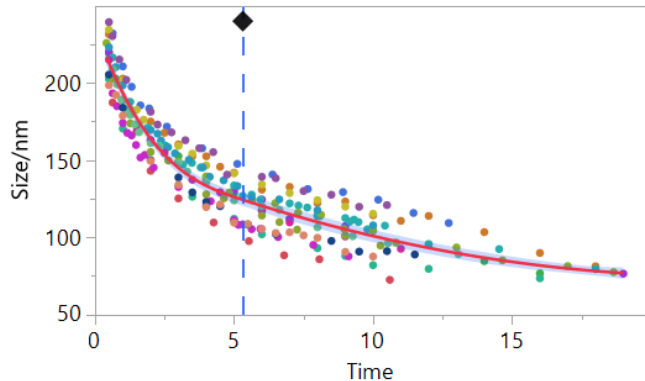


# Single Eigenfunction and Associated FPC Scores for each Batch

## B-Spline on Initial data

### Model Controls

### Model Selection



## Functional PCA

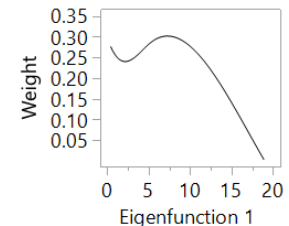
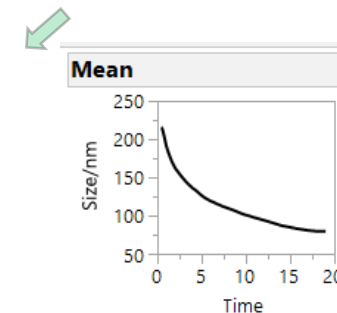
| FPC | Eigenvalue | 20 40 60 80 | Percent | Cumulative |
|-----|------------|-------------|---------|------------|
| 1   | 2122.9     |             | 99%     | 99%        |

## Function Summaries

| Batch of Mill<br>DOE responses | FPC 1     |
|--------------------------------|-----------|
| 2887                           | -85.9675  |
| 2892                           | 2.2051353 |
| 2899                           | 79.337968 |
| 2905                           | 58.272695 |
| 2908                           | -60.35218 |
| 2912                           | 7.7564941 |
| 2914                           | 49.652658 |
| 2918                           | 22.510929 |
| 2919                           | -56.34453 |
| 2920                           | 9.1386688 |
| 2963                           | 29.363581 |
| 2964                           | -13.79295 |
| 2971                           | -21.19822 |
| 2978                           | -6.822422 |
| 2980                           | -38.32591 |
| 3016                           | 71.771035 |
| 3030                           | -47.20546 |

More  
← Different: +79

↖ Similar:  $-58 \pm 2$

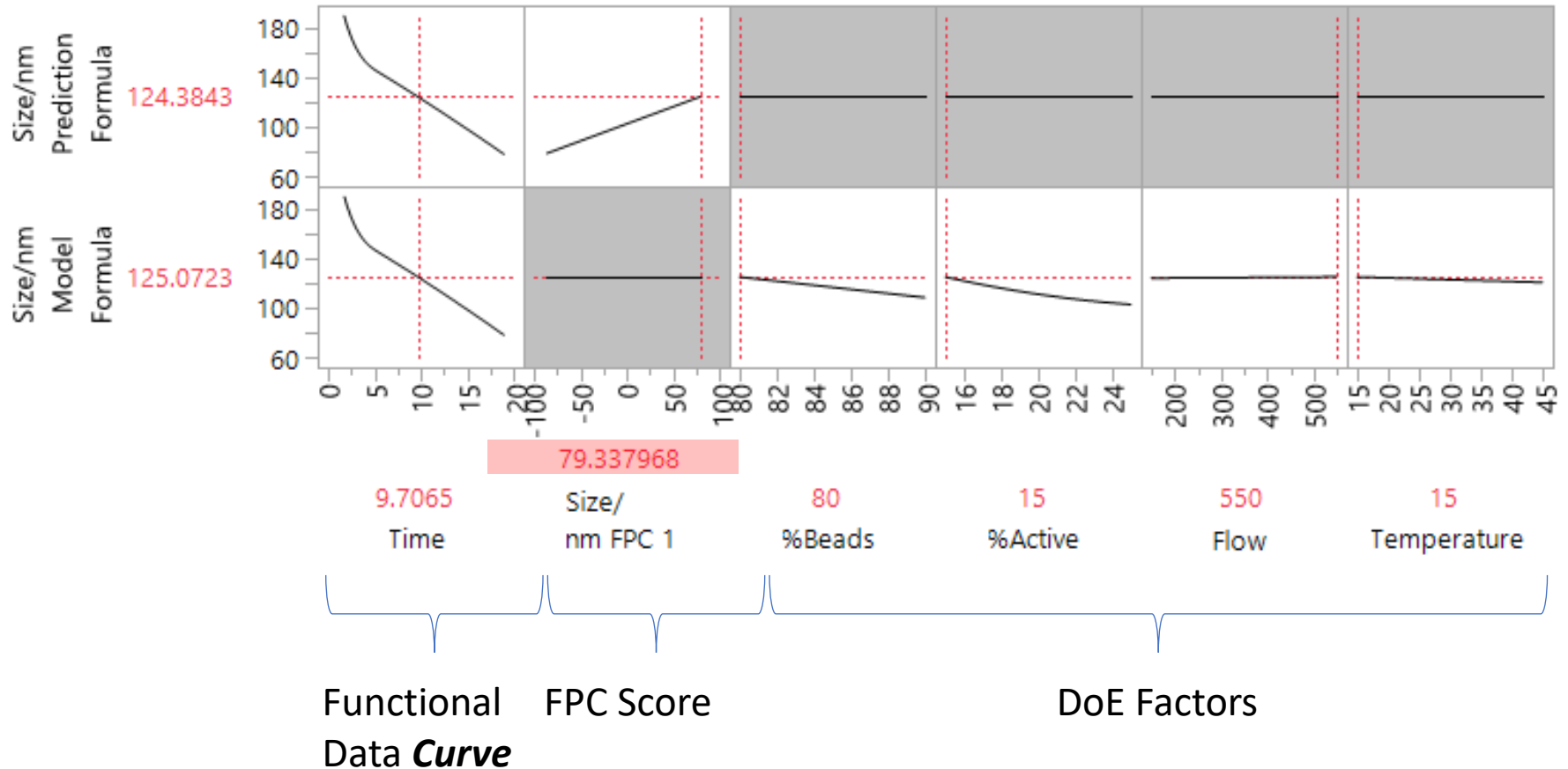


$$Y(X) = \mu(X) + \text{FPC1} \cdot E_1(X)$$

# Model the FPC Scores as functions of the DOE factors

Batch 2899

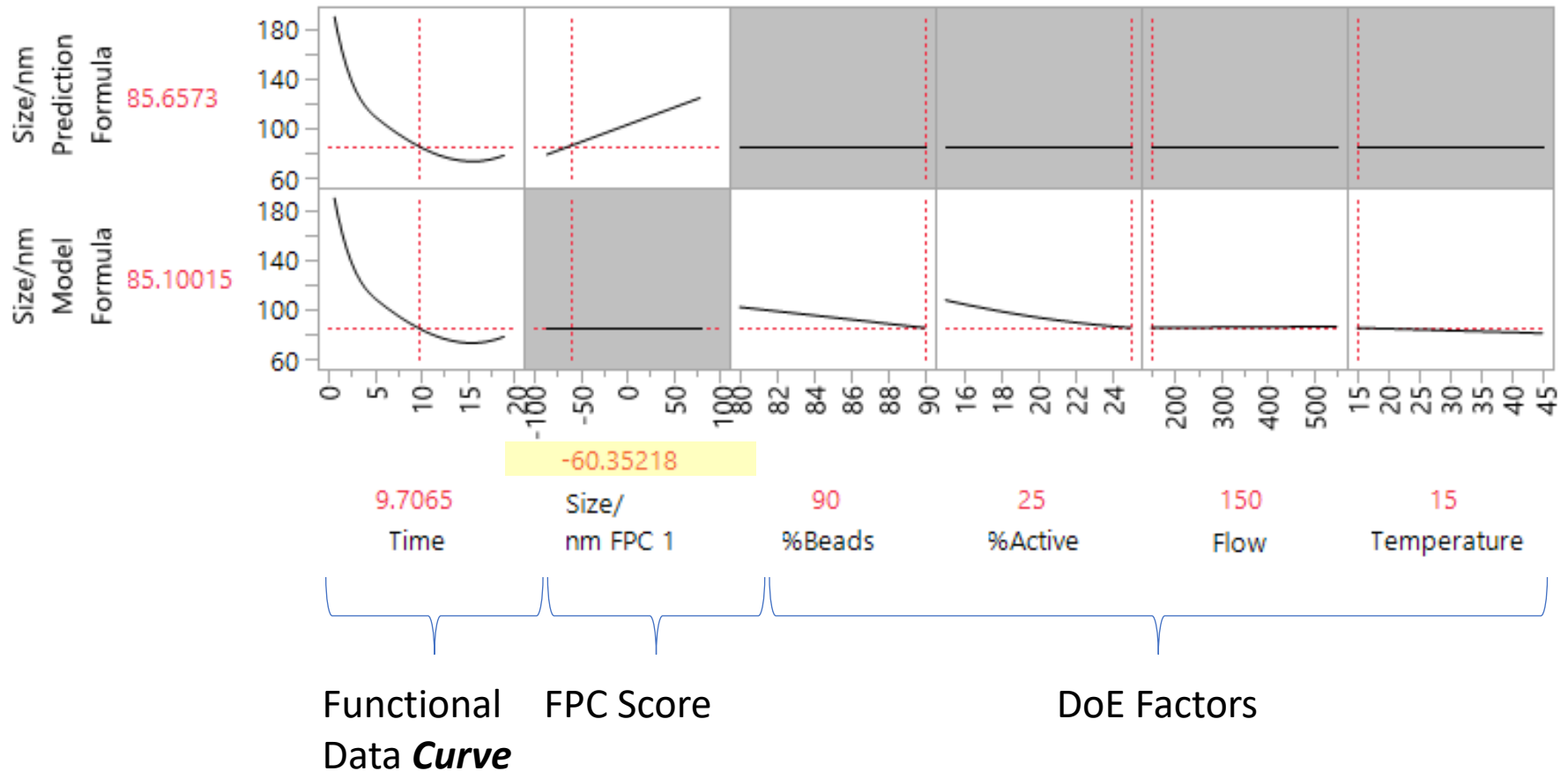
## Prediction Profiler



# Model the FPC Scores as functions of the DOE factors

## Batch 2908

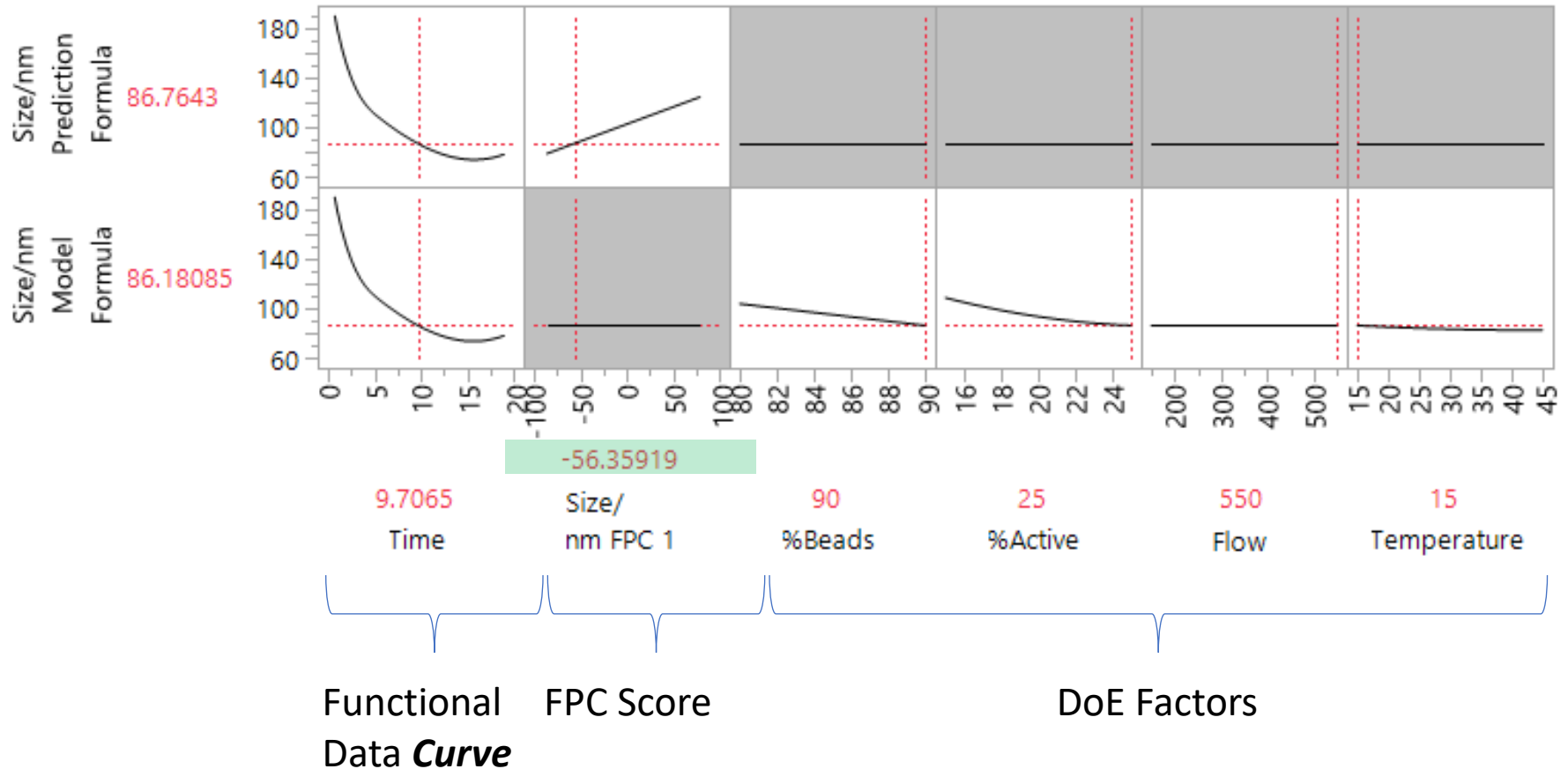
### Prediction Profiler

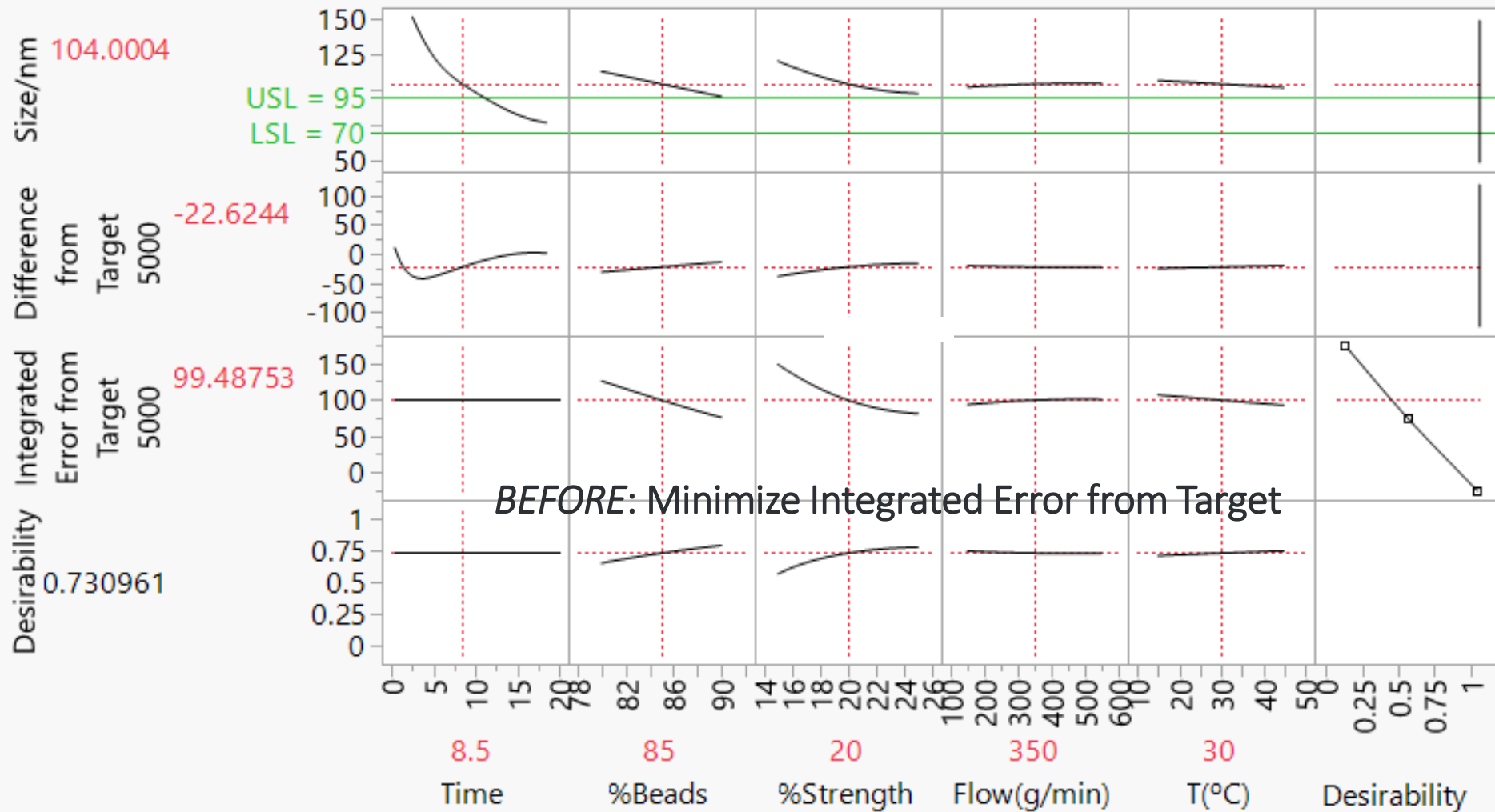


# Model the FPC Scores as functions of the DOE factors

## Batch 2919

### Prediction Profiler

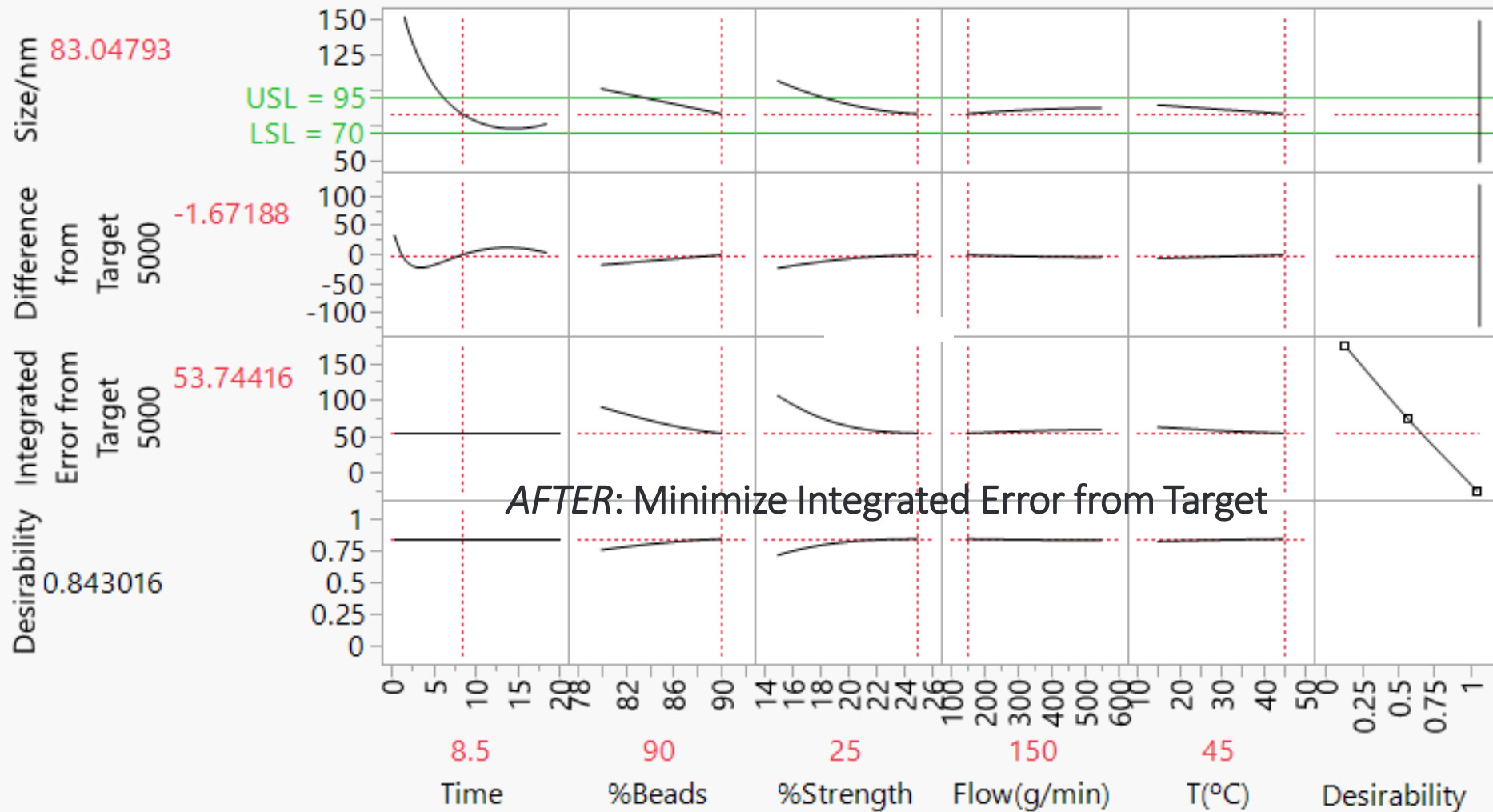




Functional Response,  
Difference from Target,  
**Integrated Error**

DoE Factors

Desirability  
(0=Bad, 1=Good)



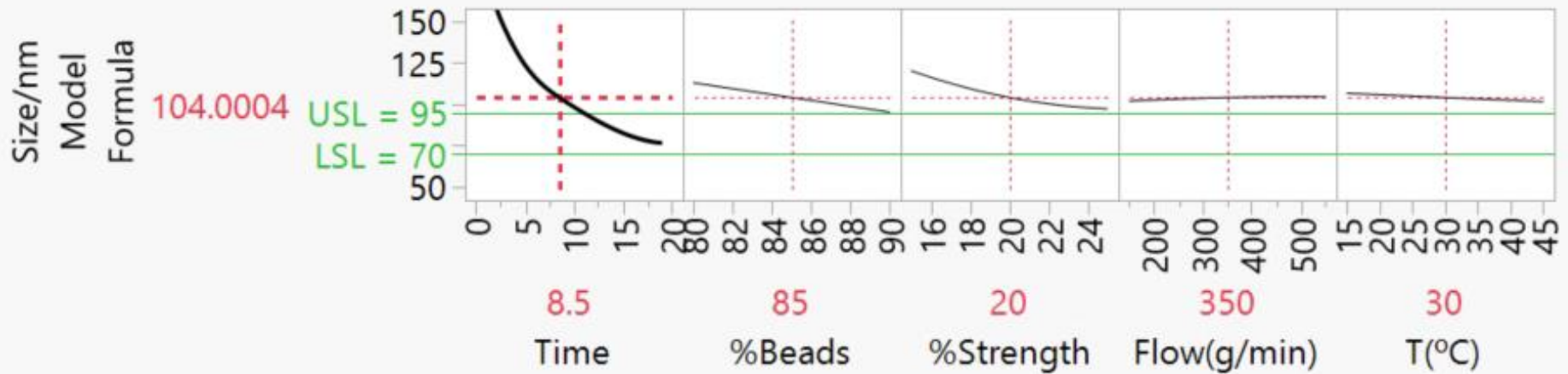
Functional Response,  
Difference from Target,  
**Integrated Error**

DoE Factors

Desirability  
(0=Bad, 1=Good)

# Final Prediction Model

## Prediction Profiler



Functional  
Data **Curve**

DoE Factors

# Use JMP to analyze Mill\_DOE.jmp data

1. Analyze > Specialized Modeling > Functional Data Explorer
2. Populate Dialog with Column Names > Click OK
3. Cleanup Data (Not required with these data)
4. Load Target Function – Batch 5000
5. Add Spec Limits to Size axis
6. Hot Spot Functional Data Explorer > Models > Model Controls > B-Spline Controls
7. Click Go
8. Inspect Function Summaries
9. Hot Spot Function Summaries > Customize Function Summaries > Deselect All > Check Save Formulas Click “OK” or “OK and Save”
10. Hot Spot B-Spline on Load Targets > Functional DOE Analysis
11. Hot Spot FDOE Profiler > Optimization and Desirability > Desirabilities Function
12. Hot Spot FDOE Profiler > Optimization and Desirability > Maximize Desirability
13. Hot Spot Customize Function Summaries > Deselect All > Check Save Formulas
14. Hot Spot Function Summaries > Save Summaries (If not done in step 9)
15. Hot Spot Functional Data Explorer > Save Script to Data Table



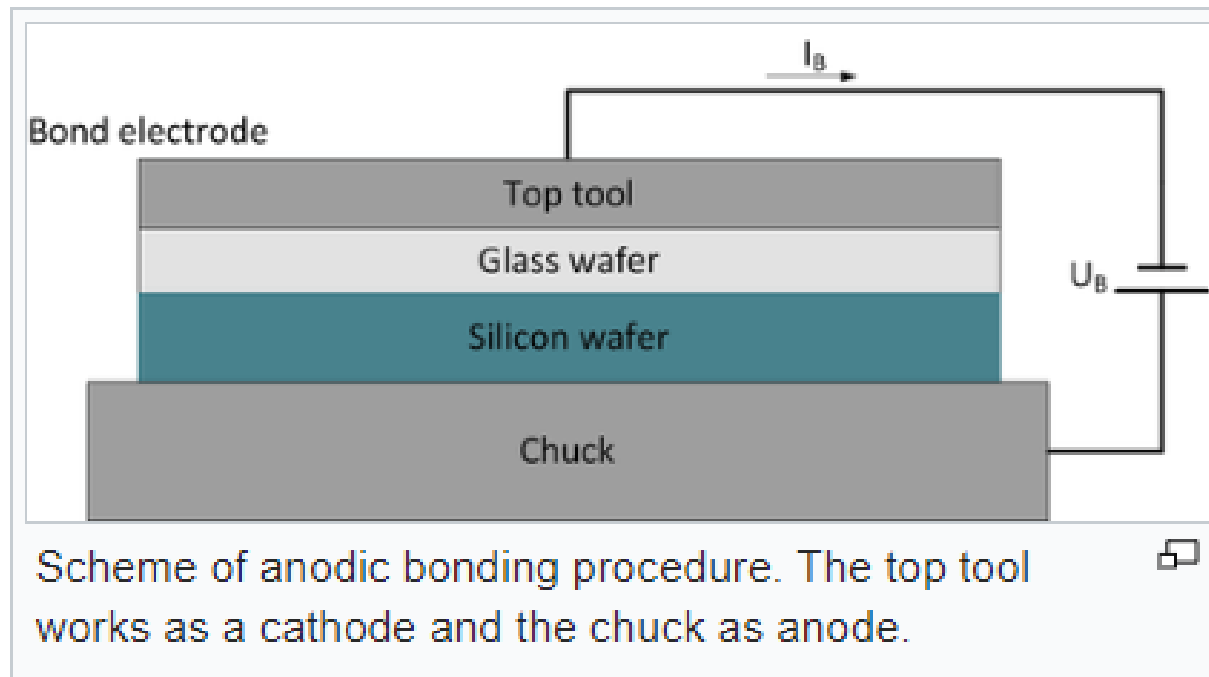
# Two Ways to Use Functional Data Analysis

- 1. Functional Response DOE (F-DOE):** Goal is to use DOE factors to predict the functional response – the *curve*
- 2. Functional Response Machine Learning (F-ML):** Goal is to use the functional data – *i.e. the curve(s)* – to predict something
  - a) yield of a batch
  - b) probability of detection / failure / hit

# Two Ways to Use Functional Data Analysis

1. **Functional Response DOE (F-DOE):** Goal is to use DOE factors to predict the functional response – the *curve*
2. **Functional Response Machine Learning (F-ML):**  
Goal is to use the functional data –  
*i.e. the **curve(s)*** – to predict something
  - a) yield of a batch
  - b) probability of detection / failure / hit

# Case Study Using Five Sensor Streams of Functional Data to Predict Wafer Condition after Anodic Bonding of Glass to Wafer



Picture from *Wikipedia...*

**Glass Bonded to Silicon Wafer**

**ISSUE: 12% of Wafers become Defective**

***BUT won't Know for Weeks which have Failed!***

# Anodic Bond Data:

2000 Wafers X 61 Time Steps = 122,000 Rows

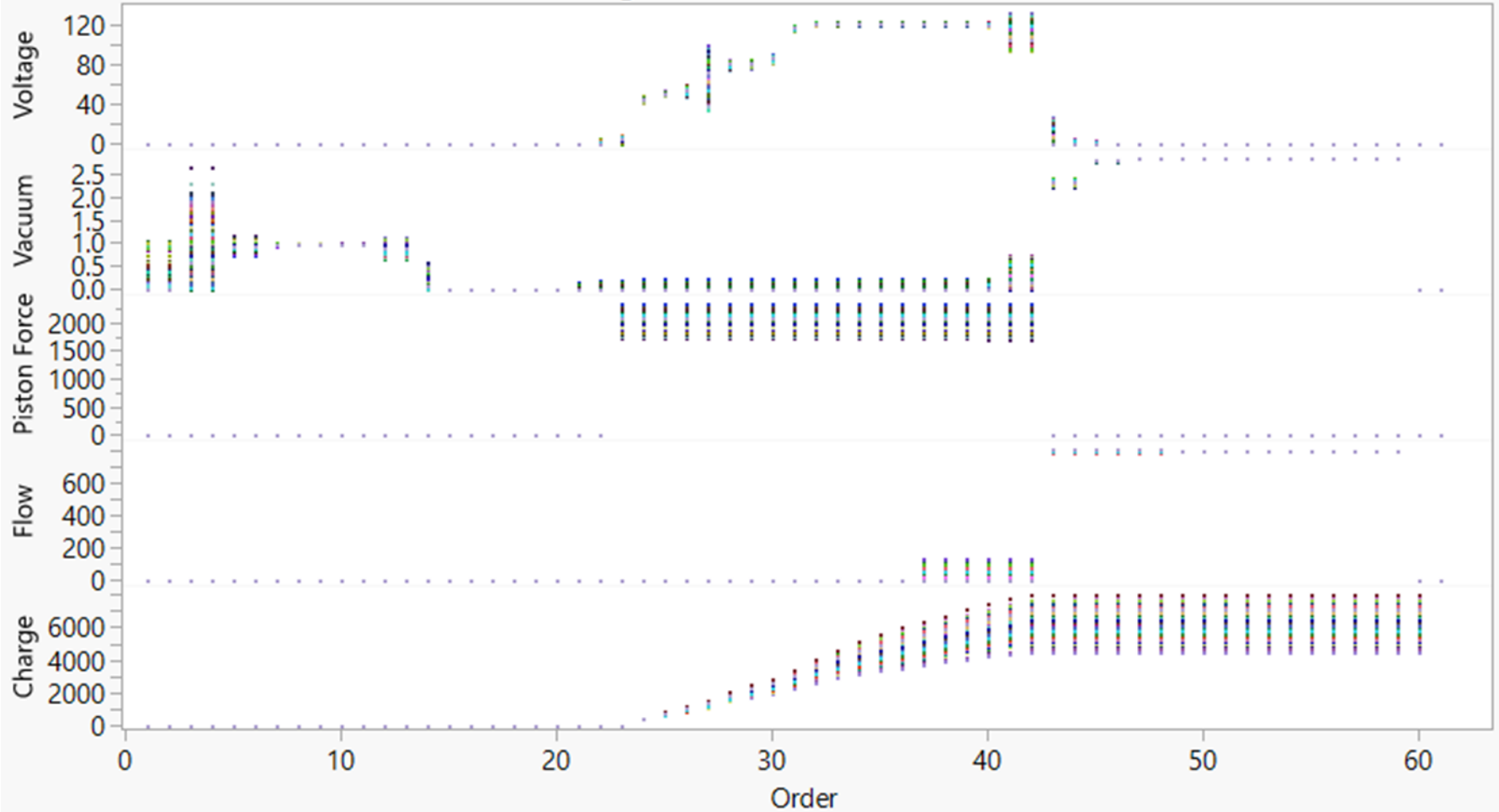
The bonding tool has several sensors that take real-time measurements of *Charge*, *Flow*, *Piston Force*, *Vacuum*, & *Voltage*.

|    | Wafer Id | Condition | Validation | Order | Charge | Flow   | Piston Force | Vacuum | Voltage |
|----|----------|-----------|------------|-------|--------|--------|--------------|--------|---------|
| 1  | 1        | GOOD      | Training   | 1     | 0.00   | 0.3013 | 0            | 0.00   | 0.00    |
| 2  | 1        | GOOD      | Training   | 2     | 0.00   | 0.3013 | 0            | 0.00   | 0.00    |
| 3  | 1        | GOOD      | Training   | 3     | 0.00   | 0.3013 | 0            | 0.50   | 0.00    |
| 4  | 1        | GOOD      | Training   | 4     | 0.00   | 0.3013 | 0            | 0.50   | 0.00    |
| 5  | 1        | GOOD      | Training   | 5     | 0.00   | 0.3013 | 0            | 0.94   | 0.00    |
| 6  | 1        | GOOD      | Training   | 6     | 0.00   | 0.3013 | 0            | 0.94   | 0.00    |
| 7  | 1        | GOOD      | Training   | 7     | 0.00   | 0.0008 | 0            | 0.99   | 0.00    |
| 8  | 1        | GOOD      | Training   | 8     | 0.00   | 0.0008 | 0            | 1.00   | 0.00    |
| 9  | 1        | GOOD      | Training   | 9     | 0.00   | 0.0008 | 0            | 1.00   | 0.00    |
| 10 | 1        | GOOD      | Training   | 10    | 0.00   | 0.0008 | 0            | 0.99   | 0.00    |
| 11 | 1        | GOOD      | Training   | 11    | 0.00   | 0.0008 | 0            | 0.99   | 0.00    |

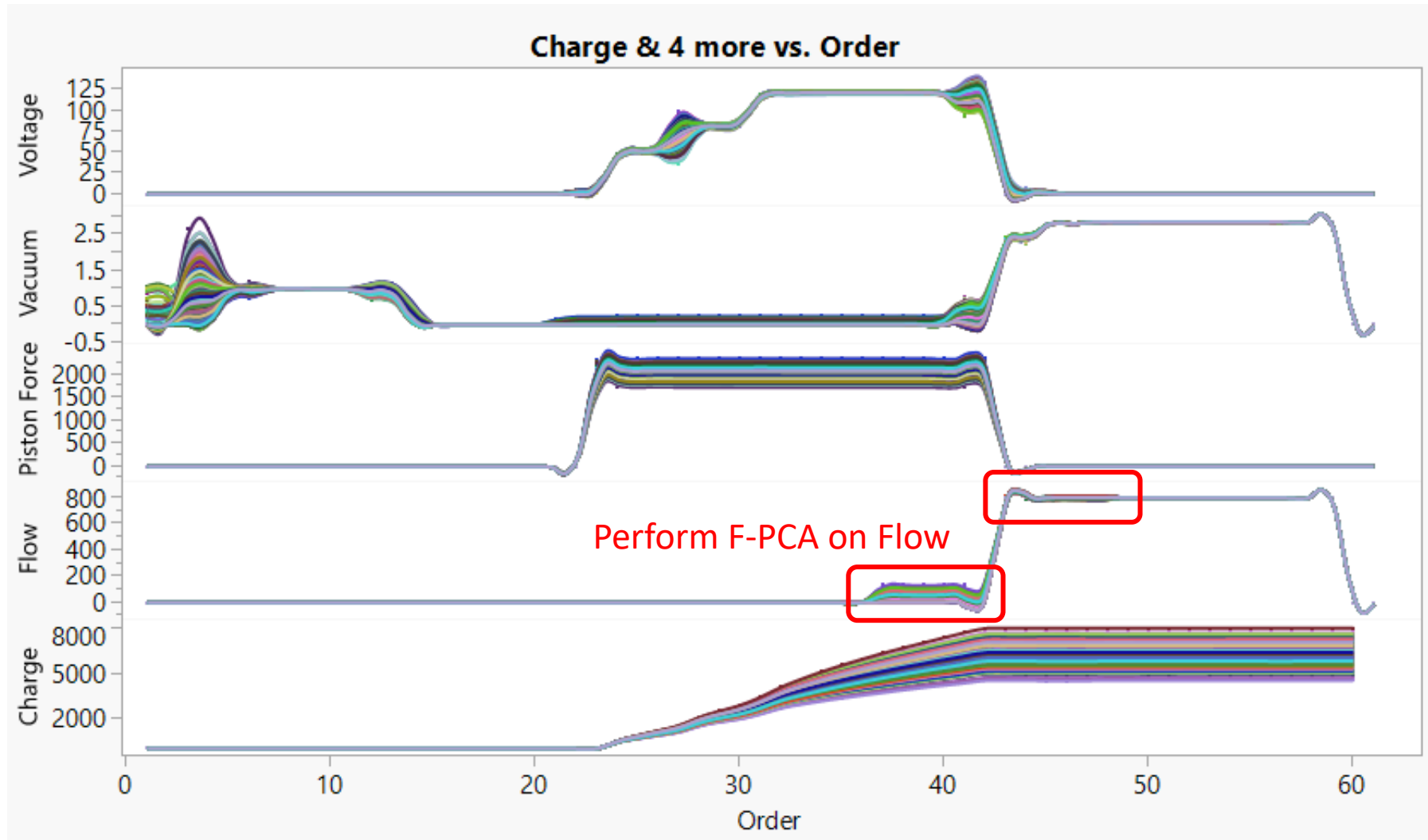
Can we use these sensor data to predict with high probability the wafers that were damaged by the bonding process – right now?

# Anodic Bond Data: Discrete Observations

Charge & 4 more vs. Order



# Anodic Bond Data: Smoothed Data Streams from 2000 Glass-to-Wafer Bonds



# What is Functional Data Analysis?

**Functional data analysis (FDA)** is a branch of statistics that analyzes data providing information about **curves, surfaces** or anything else **varying over a continuum**. In its most general form, under an FDA framework each sample element is considered to be a **function**.

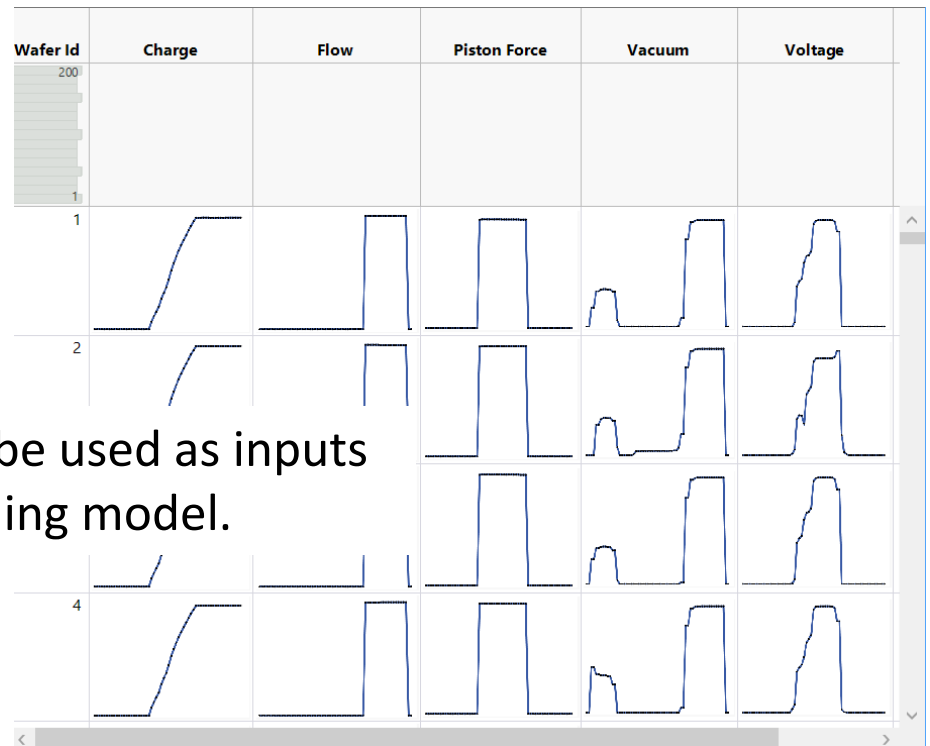
2000 wafers X 61 rows/wafer

| Wafer Id     | Order | Charge | Flow | Piston Force | Vacuum | Voltage |
|--------------|-------|--------|------|--------------|--------|---------|
| 1            | 61    | 8073   | 813  | 2340         | 2.85   | 133     |
| 2            |       |        |      |              |        |         |
| 3            |       |        |      |              |        |         |
| 4            |       |        |      |              |        |         |
| 5            |       |        |      |              |        |         |
| 1,995 others | 1     | 0      | 0    | 0            | 0      | 0       |
| 1            | 1     | 1      | 0.00 | 0.3013       | 0      | 0.00    |
| 2            | 1     | 2      | 0.00 | 0.3013       | 0      | 0.00    |
| 3            | 1     | 3      | 0.00 | 0.3013       | 0      | 0.50    |
| 4            | 1     | 4      | 0.00 | 0.3013       | 0      | 0.50    |
| 5            | 1     | 5      | 0.00 | 0.3013       | 0      | 0.94    |
| 6            | 1     | 6      | 0.00 | 0.3013       | 0      | 0.94    |
| 7            | 1     | 7      | 0.00 | 0            |        |         |
| 8            | 1     | 8      | 0.00 | 0            |        |         |
| 9            | 1     | 9      | 0.00 | 0            |        |         |
| 10           | 1     | 10     | 0.00 | 0            |        |         |
| 11           | 1     | 11     | 0.00 | 0            |        |         |
| 12           | 1     | 12     | 0.00 | 0.0008       | 0      | 0.93    |
| 13           | 1     | 13     | 0.00 | 2.3e-6       | 0      | 0.93    |
| 14           | 1     | 14     | 0.00 | 2.3e-6       | 0      | 0.16    |
| 15           | 1     | 15     | 0.00 | 2.3e-6       | 0      | 0.00    |
| 16           | 1     | 16     | 0.00 | 2.3e-6       | 0      | 0.00    |
| 17           |       |        |      |              |        |         |

122,000 rows of data

Functional data to be used as inputs  
to a Machine Learning model.

2000 Functional Data Streams

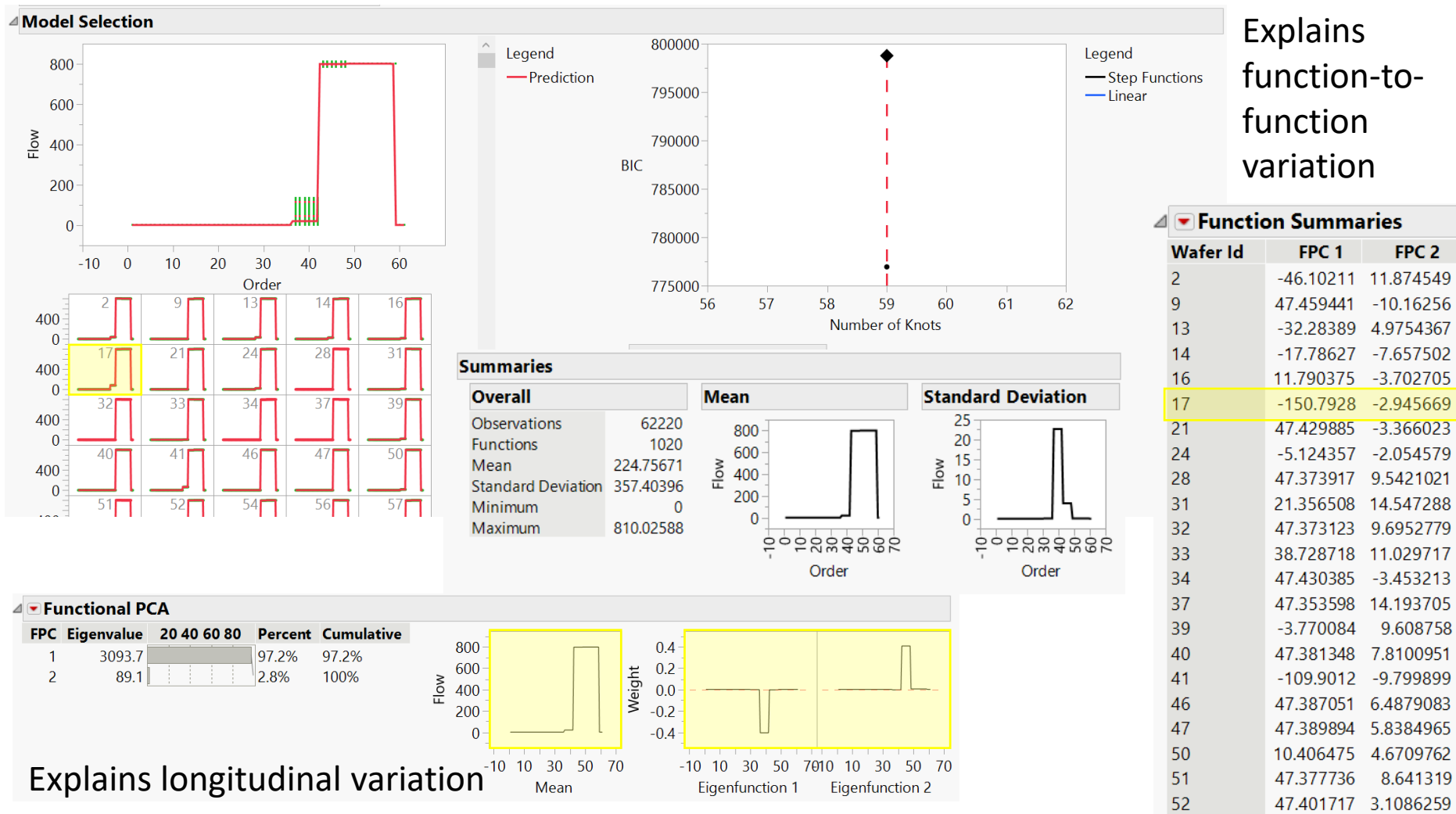


**Curve** is the fundamental unit of observation



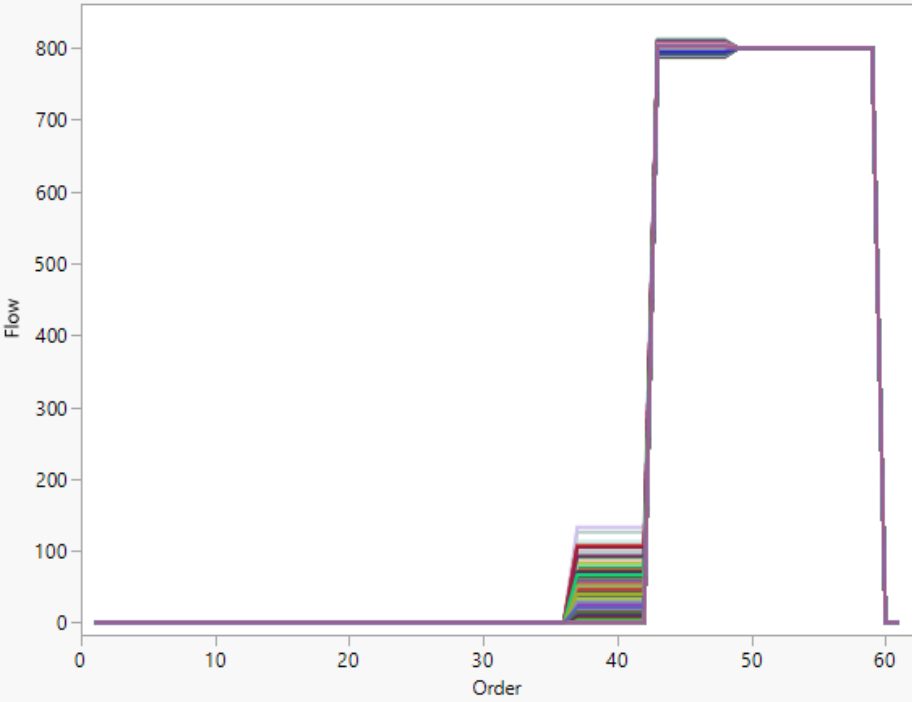
# How do we analyze Functional data?

- Products of FPC scores multiplying their corresponding eigenfunctions, when added to the Mean closely reproduce the individual function (Flow) curves.

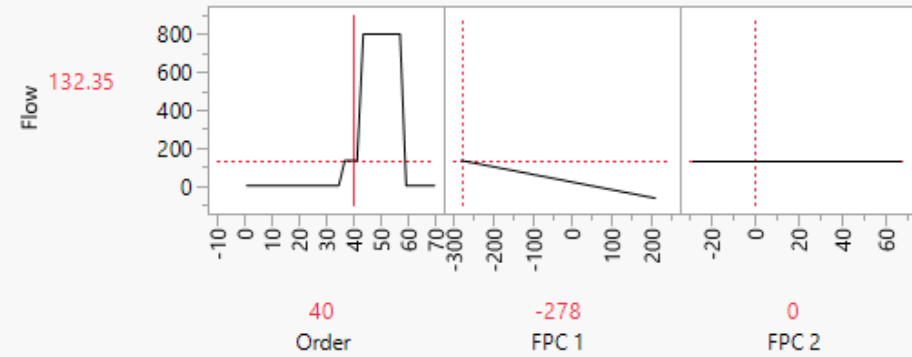


# Flow FPC1

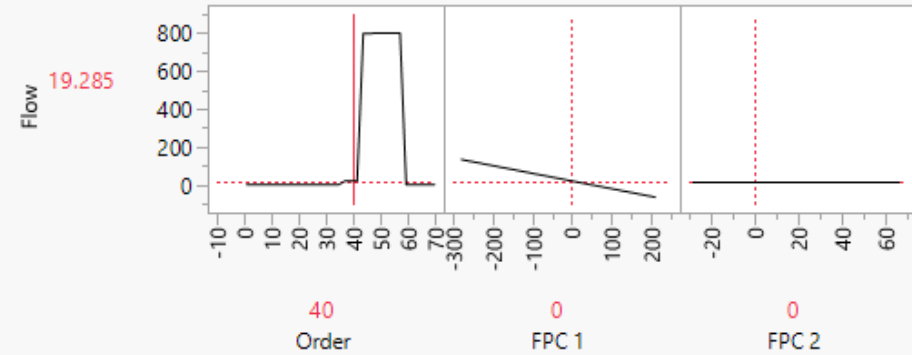
Flow vs. Order



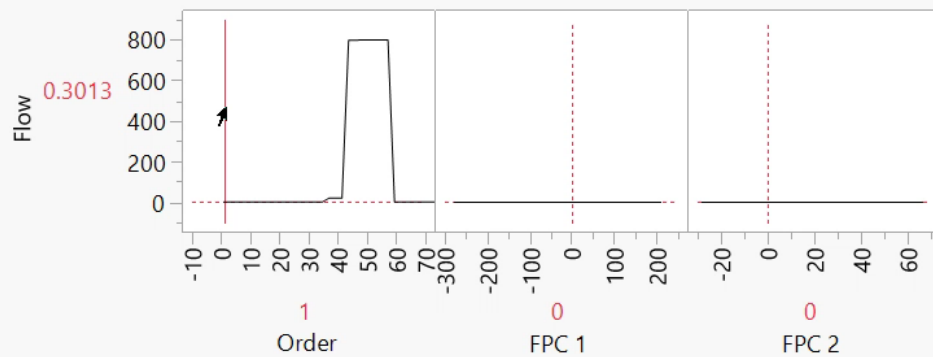
FPC Profiler



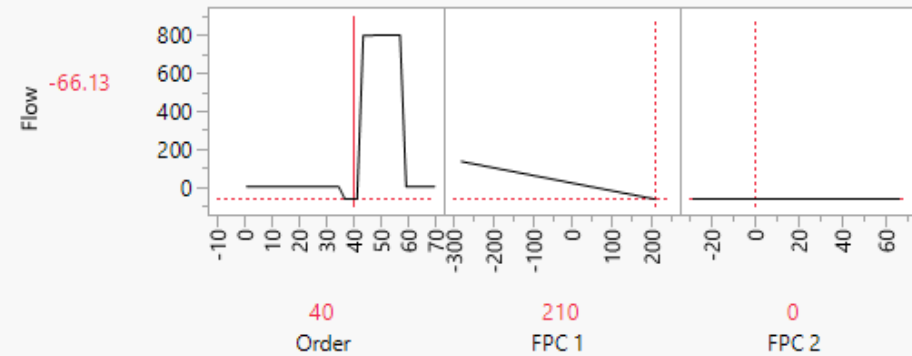
FPC Profiler



FPC Profiler



FPC Profiler

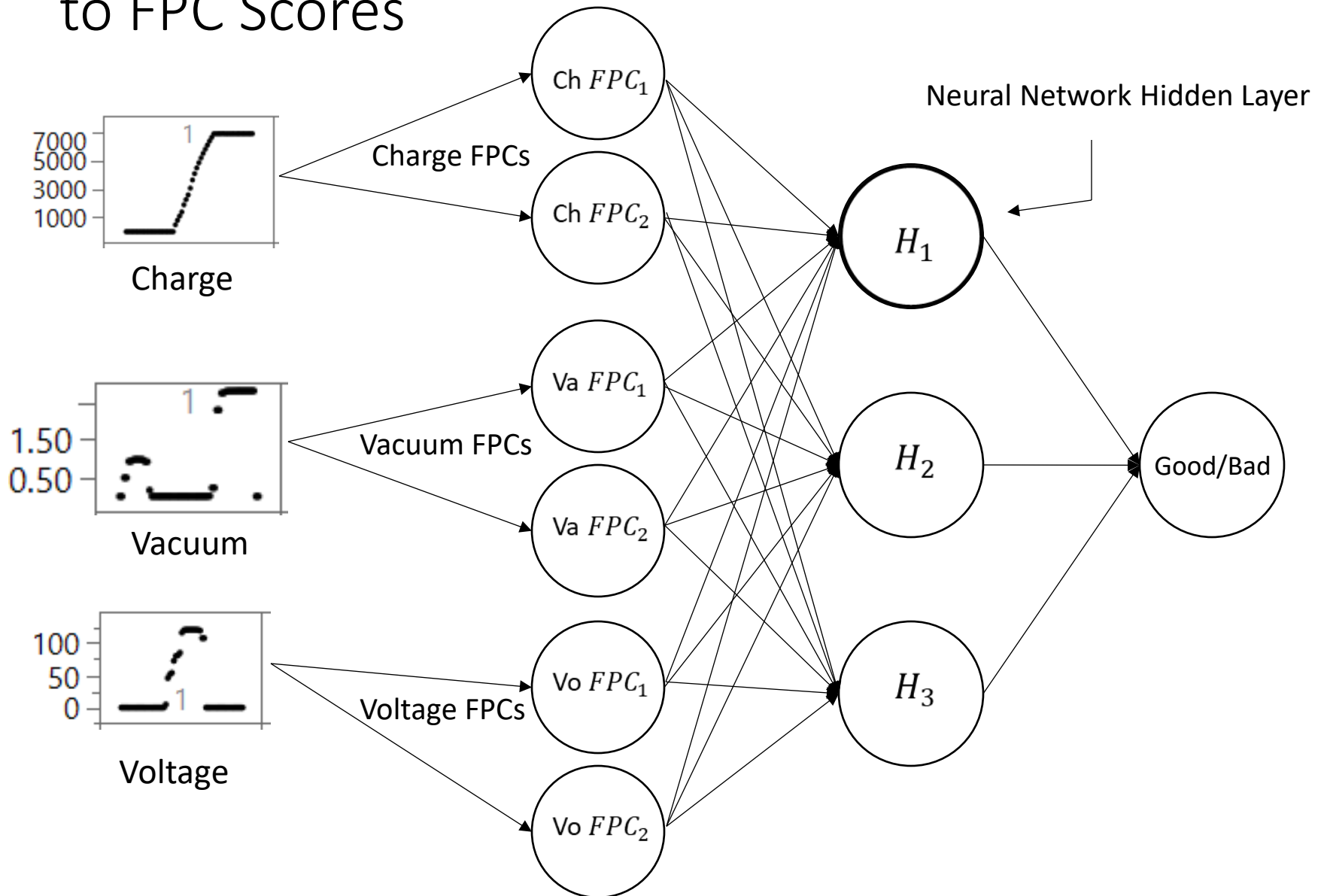


# Table of 12\* FPC Scores used to Model Condition

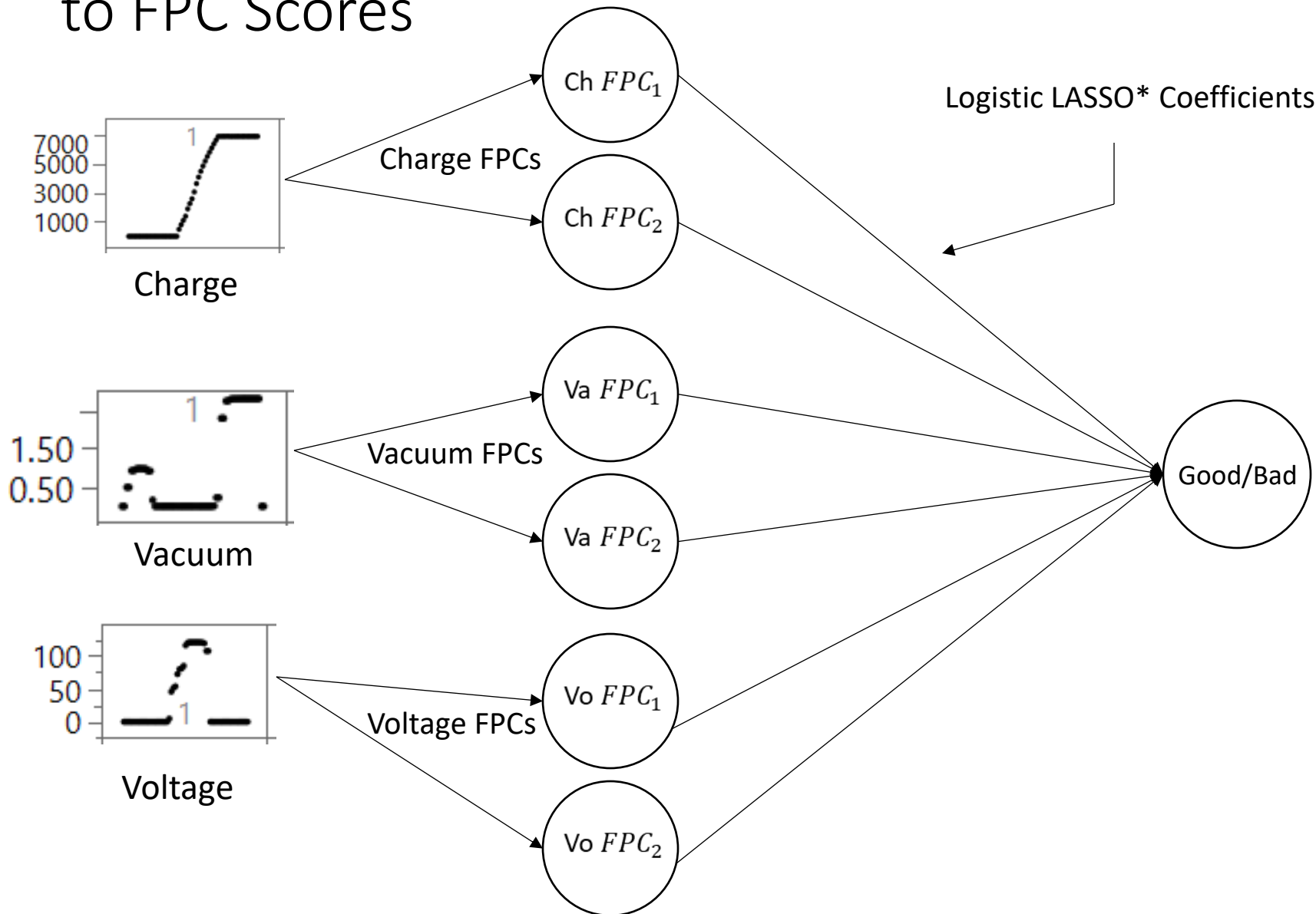
| Wafer Id     | Condition | Validation | Charge FPC 1 | Flow FPC 1   | Flow FPC 2   | Piston Force FPC 1 | Vacuum FPC 1 | Vacuum FPC 2 | Vacuum FPC 3 |
|--------------|-----------|------------|--------------|--------------|--------------|--------------------|--------------|--------------|--------------|
| 1            | GOOD      | Training   | 8295         | 45.7         | 34.7         | 1311               | 2.97         | 0.58         | 0.3          |
| 2            | BAD       | Validation |              |              |              |                    |              |              |              |
| 3            |           | Test       |              |              |              |                    |              |              |              |
| 4            |           |            |              |              |              |                    |              |              |              |
| 5            |           |            |              |              |              |                    |              |              |              |
| 1,995 others |           |            | -8903        | -281         | -28.6        | -1442              | -1.04        | -0.72        | -0.59        |
| 2            | GOOD      | Training   | 1939.8213944 | -47.99467507 | 11.835188042 | 896.52111845       | -1.01396436  | -0.367923438 | -0.296827224 |
| 6            | GOOD      | Training   | 212.05037509 | -25.13715846 | -0.995827966 | -84.14117634       | 0.5477329858 | -0.345665597 | -0.124382493 |
| 9            | GOOD      | Training   | -34.75688741 | 45.575552449 | -10.16506939 | 315.25975102       | -0.538688448 | 0.1508261265 | 0.1519855668 |
| 11           | BAD       | Training   | -1213.835105 | 45.495664474 | 6.680705469  | 50.911046557       | -0.064379983 | -0.01993877  | 0.0981726233 |
| 12           | GOOD      | Training   | -1013.153308 | -39.90355163 | 2.3623655717 | -340.5227485       | 0.9016693131 | -0.178605516 | 0.0605873881 |
| 16           | GOOD      | Training   | -3985.006867 | 9.903945259  | -3.719261334 | -587.5458989       | 0.7955624065 | 0.0297828066 | 0.042004684  |
| 17           | GOOD      | Training   | 1832.3340433 | -152.6794891 | -3.02626366  | 134.08863681       | -0.621848953 | -0.486645362 | 0.121393253  |
| 18           | GOOD      | Training   | -340.0108681 | -12.98109355 | -16.7930519  | -175.6851773       | 0.2920817989 | 0.0409672523 | -0.004401981 |
| 20           | GOOD      | Training   | 1178.6360003 | -114.1446381 | -5.71271533  | -81.37628428       | 0.0064651167 | -0.318735656 | -0.120126295 |
| 23           | BAD       | Training   | -3015.826554 | 45.496077049 | 6.5969881112 | -393.9696867       | 0.8434918197 | 0.2002559472 | -0.0234644   |
| 27           | BAD       | Training   | 647.92427295 | 45.526409332 | 0.2060713613 | -1.736804312       | -0.087564419 | 0.3828626381 | -0.076428983 |
| 28           | BAD       | Training   | -2331.969564 | 45.482267145 | 9.539588982  | 149.29748575       | -0.258712307 | 0.4084688316 | -0.057425217 |
| 32           | BAD       | Training   | 1110.549478  | 45.481412652 | 9.692744135  | 173.94633371       | -0.309896253 | 0.1897807551 | 0.1005481727 |
| 33           | GOOD      | Training   | 545.08163682 | 36.83648222  | 11.02376959  | -68.08890553       | 0.3480580578 | 0.1435364586 | -0.069774713 |
| 34           | BAD       | Training   | 1153.6816932 | 45.543853482 | -3.455710358 | 2.0682952872       | 0.1741396938 | 0.3336479119 | -0.056667203 |

\*5 Columns of FPC Scores NOT shown

# Predict Wafer Condition by Fitting Neural Model to FPC Scores



# Predict Wafer Condition by Fitting Logistic Model to FPC Scores



\* "LASSO" stands for **L**east **A**bsolute **S**hrinkage and **S**election **O**perator

# Results of Fitting Logistic and Neural Models

## Binomial Logistic Regression with Validation Column

### Model Summary

| Response               | Condition           |            |
|------------------------|---------------------|------------|
| Distribution           | Binomial            |            |
| Estimation Method      | Logistic Regression |            |
| Validation Method      | Validation Column   |            |
| Probability Model Link | Logit               |            |
| Measure                | Training            | Validation |
| Number of rows         | 1000                | 500        |
| Sum of Frequencies     | 1000                | 500        |
| -LogLikelihood         | 228.54132           | 114.45861  |
| Number of Parameters   | 13                  | 13         |
| BIC                    | 546.88346           | 309.70713  |
| AICc                   | 483.45181           | 255.6662   |
| Generalized RSquare    | 0.4523866           | 0.4555533  |

## Neural

Validation Column: Validation

### Model Launch

### Model NTanH(1)NLinear(1)NGaussian(1)NBoost(16)

#### Training

##### Condition

| Measures               | Value     |
|------------------------|-----------|
| Generalized RSquare    | 0.5325325 |
| Entropy RSquare        | 0.4432134 |
| RMSE                   | 0.2428032 |
| Mean Abs Dev           | 0.1309093 |
| Misclassification Rate | 0.074     |
| -LogLikelihood         | 200.94695 |
| Sum Freq               | 1000      |

#### Validation

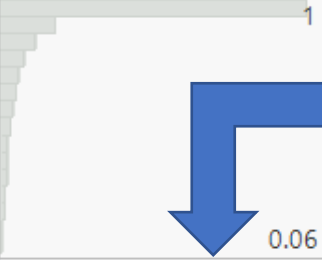
##### Condition

| Measures               | Value     |
|------------------------|-----------|
| Generalized RSquare    | 0.5607994 |
| Entropy RSquare        | 0.4707256 |
| RMSE                   | 0.2421686 |
| Mean Abs Dev           | 0.1288395 |
| Misclassification Rate | 0.074     |
| -LogLikelihood         | 96.042421 |
| Sum Freq               | 500       |

|            |           | GenReg Binom 3-way Most Likely Condition |     |
|------------|-----------|--|-----|
| Validation | Condition | GOOD                                     | BAD |
| Training   | GOOD      | 864                                      | 19  |
|            | BAD       | 74                                       | 43  |
| Validation | GOOD      | 433                                      | 8   |
|            | BAD       | 42                                       | 17  |
| Test       | GOOD      | 428                                      | 13  |
|            | BAD       | 37                                       | 22  |

|            |           | BN 1-1-1(16) Most Likely Condition |     |
|------------|-----------|------------------------------------|-----|
| Validation | Condition | GOOD                               | BAD |
| Training   | GOOD      | 872                                | 11  |
|            | BAD       | 63                                 | 54  |
| Validation | GOOD      | 437                                | 4   |
|            | BAD       | 33                                 | 26  |
| Test       | GOOD      | 434                                | 7   |
|            | BAD       | 42                                 | 17  |

# Table of Neural Model Predictions of Condition

| 18/0 Cols |      | Condition | Validation | Probability(Condition=GOOD )   | BN 1-1-1(16) Most Likely Condition |
|-----------|------|-----------|------------|--|------------------------------------|
|           |      | GOOD      | Training   |  | GOOD                               |
|           |      | BAD       | Validation |  | BAD                                |
|           |      |           | Test       |  |                                    |
| ○         | 1564 | GOOD      | Test       | 0.9961988532   | GOOD                               |
| ○         | 1565 | GOOD      | Test       | 0.9415687312   | GOOD                               |
| ○         | 1566 | GOOD      | Test       | 0.9670452594   | GOOD                               |
| ○         | 1567 | GOOD      | Test       | 0.5550716285   | GOOD                               |
| +         | 1568 | BAD       | Test       | 0.4930775442   | BAD                                |
| +         | 1569 | BAD       | Test       | 0.5335959152   | GOOD                               |
| ○         | 1570 | GOOD      | Test       | 0.6131324314   | GOOD                               |
| +         | 1571 | BAD       | Test       | 0.8668675093   | GOOD                               |
| +         | 1572 | BAD       | Test       | 0.2583111765   | BAD                                |
| ○         | 1573 | GOOD      | Test       | 0.8699599618   | GOOD                               |
| +         | 1574 | BAD       | Test       | 0.7201964039   | GOOD                               |
| ○         | 1575 | GOOD      | Test       | 0.99613207   | GOOD                               |
| +         | 1576 | BAD       | Test       | 0.5647955597   | GOOD                               |
| ○         | 1577 | GOOD      | Test       | 0.9891191688   | GOOD                               |
| ○         | 1578 | GOOD      | Test       | 0.987960478  | GOOD                               |
| ○         | 1579 | GOOD      | Test       | 0.9750283598   | GOOD                               |

Model these probabilities to develop decision tree “stoplight.”

Just misses being “Good” prediction

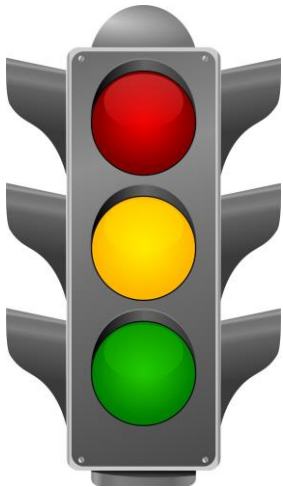
Not nearly a “Good” prediction



# Want to Predict Likely Failed Wafers –

Decision Tree Fit to Neural Network Probability Predictions

Built from Functional Principal Component Scores for  
Five Anodic Bonding Sensors



Less than 0.395

0.395 to 0.754

More than 0.754

|            | RSquare | N    | Number of Splits |
|------------|---------|------|------------------|
| Training   | 0.398   | 1000 | 2                |
| Validation | 0.336   | 500  |                  |

| All Rows                          |        |        |       |
|-----------------------------------|--------|--------|-------|
| <div><div></div><div></div></div> |        |        |       |
| Level                             | Rate   | Prob   | Count |
| GOOD                              | 0.8830 | 0.8830 | 883   |
| BAD                               | 0.1170 | 0.1170 | 117   |

| Probability(Condition=GOOD)<0.7541564548 |        |        |       |
|--|--------|--------|-------|
| <div><div></div><div></div></div>        |        |        |       |
| Level                                    | Rate   | Prob   | Count |
| GOOD                                     | 0.4596 | 0.4622 | 74    |
| BAD                                      | 0.5404 | 0.5378 | 87    |

| Probability(Condition=GOOD)>=0.7541564548 |        |        |       |
|---|--------|--------|-------|
| <div><div></div><div></div></div>         |        |        |       |
| Level                                     | Rate   | Prob   | Count |
| GOOD                                      | 0.9642 | 0.9641 | 809   |
| BAD                                       | 0.0358 | 0.0359 | 30    |
| Candidates                                |        |        |       |

| Probability(Condition=GOOD)<0.3950198207 |        |        |       |
|--|--------|--------|-------|
| <div><div></div><div></div></div>        |        |        |       |
| Level                                    | Rate   | Prob   | Count |
| GOOD                                     | 0.0476 | 0.0661 | 2     |
| BAD                                      | 0.9524 | 0.9339 | 40    |
| Candidates                               |        |        |       |

| Probability(Condition=GOOD)>=0.3950198207 |        |        |       |
|---|--------|--------|-------|
| <div><div></div><div></div></div>         |        |        |       |
| Level                                     | Rate   | Prob   | Count |
| GOOD                                      | 0.6050 | 0.6070 | 72    |
| BAD                                       | 0.3950 | 0.3930 | 47    |
| Candidates                                |        |        |       |

5%/95%

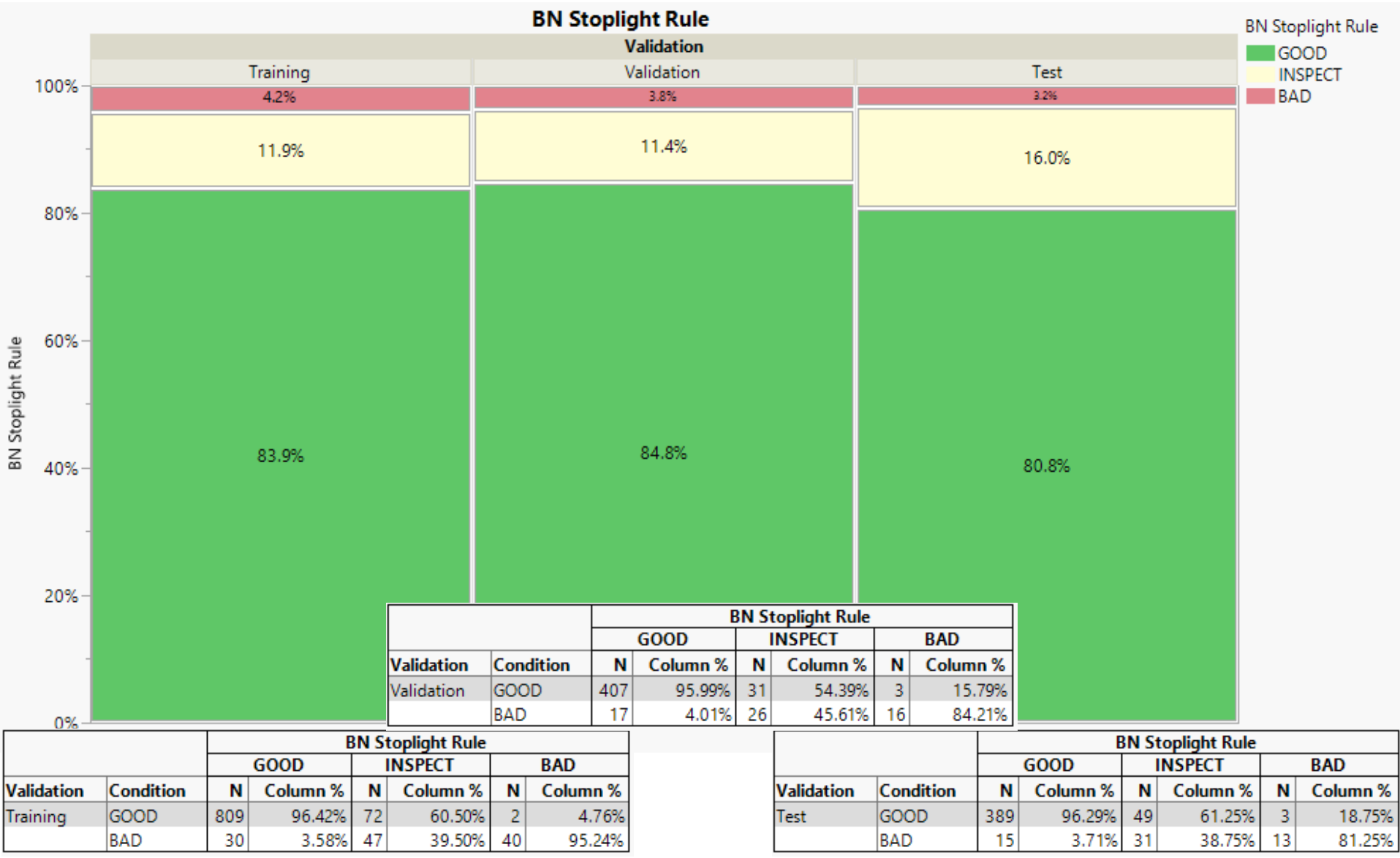
96%/4%

60%/40%

# Table of Model Predictions and Stoplight Rule

| 18/0 Cols |      | Condition | Validation | Probability(Condition=GOOD ) | BN 1-1-1(16) Most Likely Condition | BN Stoplight Rule |
|-----------|------|-----------|------------|------------------------------|------------------------------------|-------------------|
|           |      | GOOD      | Training   | 1                            | GOOD                               | GOOD              |
|           |      | BAD       | Validation |                              | BAD                                | INSPECT           |
|           |      |           | Test       |                              |                                    | BAD               |
|           |      |           |            | 0.06                         |                                    |                   |
| ○         | 1564 | GOOD      | Test       | 0.9961988532                 | GOOD                               | GOOD              |
| ○         | 1565 | GOOD      | Test       | 0.9415687312                 | GOOD                               | GOOD              |
| ○         | 1566 | GOOD      | Test       | 0.9670452594                 | GOOD                               | GOOD              |
| ○         | 1567 | GOOD      | Test       | 0.5550716285                 | GOOD                               | INSPECT           |
| +         | 1568 | BAD       | Test       | 0.4930775442                 | BAD                                | INSPECT           |
| +         | 1569 | BAD       | Test       | 0.5335959152                 | GOOD                               | INSPECT           |
| ○         | 1570 | GOOD      | Test       | 0.6131324314                 | GOOD                               | INSPECT           |
| +         | 1571 | BAD       | Test       | 0.8668675093                 | GOOD                               | GOOD              |
| +         | 1572 | BAD       | Test       | 0.2583111765                 | BAD                                | BAD               |
| ○         | 1573 | GOOD      | Test       | 0.8699599618                 | GOOD                               | GOOD              |
| +         | 1574 | BAD       | Test       | 0.7201964039                 | GOOD                               | INSPECT           |
| ○         | 1575 | GOOD      | Test       | 0.99613207                   | GOOD                               | GOOD              |
| +         | 1576 | BAD       | Test       | 0.5647955597                 | GOOD                               | INSPECT           |
| ○         | 1577 | GOOD      | Test       | 0.9891191688                 | GOOD                               | GOOD              |
| ○         | 1578 | GOOD      | Test       | 0.987960478                  | GOOD                               | GOOD              |
| ○         | 1579 | GOOD      | Test       | 0.9750283598                 | GOOD                               | GOOD              |

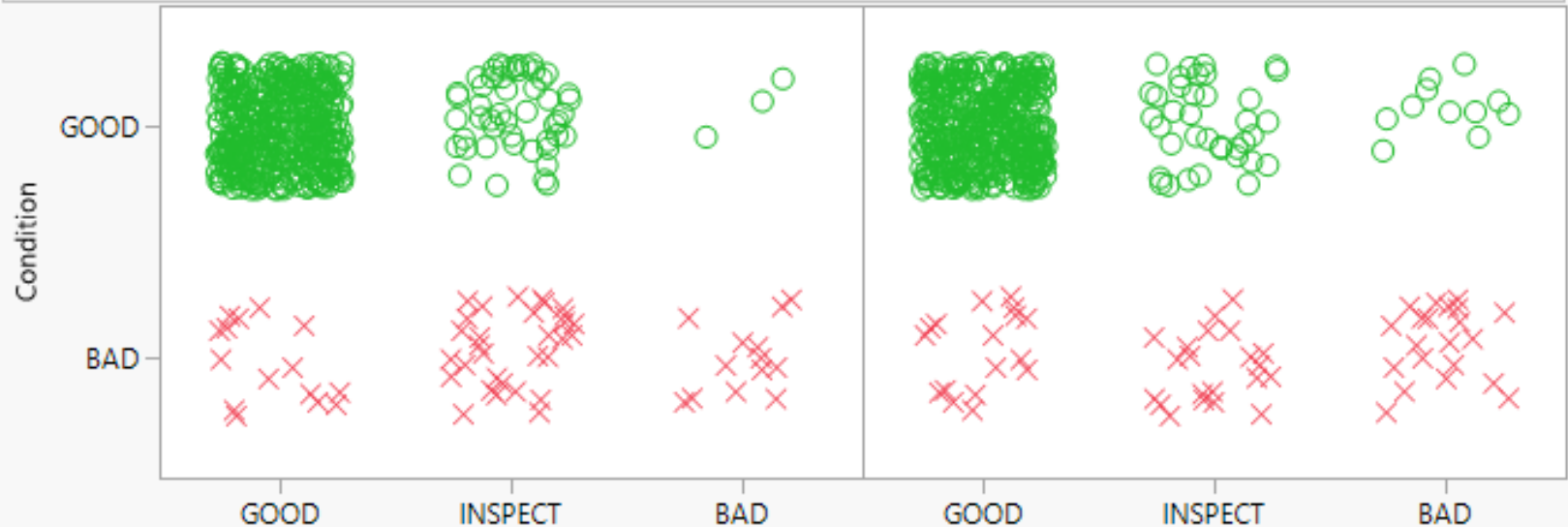
# Percentage Wafers in Each Classification by Training-Validation-Test Group - AND Tabulation of Actual by Predicted Condition



# Scatterplot Actual vs. Prediction in Test Group & Tabulation of Actual by Predicted Condition

Where(Format( :Validation ) == "Test")

**Scatterplot Matrix**



**Validation = Test**

|            |           | BN Stoplight Rule |         |     | GR Stoplight Rule |         |     |
|------------|-----------|-------------------|---------|-----|-------------------|---------|-----|
| Validation | Condition | GOOD              | INSPECT | BAD | GOOD              | INSPECT | BAD |
| Test       | GOOD      | 389               | 49      | 3   | 393               | 37      | 11  |
|            | BAD       | 15                | 31      | 13  | 17                | 21      | 21  |

BN gets fewer correct, but also  
fewer wrong: 3.6% misclassified

GR gets more correct, but also  
more wrong: 5.6% misclassified

# Use JMP to analyze Anodic\_Bond.jmp data

1. Analyze > Specialized Modeling > Functional Data Explorer
2. Populate Dialog with Column Names > Click OK (NOTE: Demo only Flow Response)
3. Cleanup Data (Not required with these data)
4. Hot Spot Functional Data explorer > Models > Model Controls > P-Spline Controls  
Check Step Functions only - set knots to only 59 – Click Go
5. Inspect Function Summaries
6. Hot Spot Function Summaries > Customize Function Summaries > Deselect All >  
Check Save Formulas Click “OK” or “OK and Save”
7. Hot Spot Function Summaries > Save Summaries (If not done in step 6)
8. Hot Spot Functional Data Explorer > Save Script to Data Table

# Functional Data Analysis Performance Tips

- When there are 1000s of batches with 1000s of measurements things can slow down quite a bit.
- Try using a Training set with dozens or a 100 or so batches.
  - Place the remaining batches in Validation.
  - You will still get FPC for all the batches, but the mixed model that is fit behind the scenes will only use the training batches.
- Try subsampling down to every 10<sup>th</sup> or 20<sup>th</sup> measurement. Often you have more measurements than you need.
- Use the subset of the data to ‘fail fast’ in the modeling process.
- You can always go back and refit the better models to a larger version of the data.

# Summary

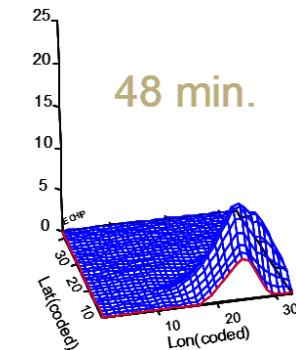
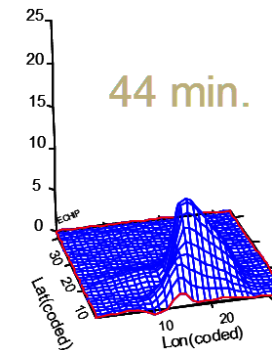
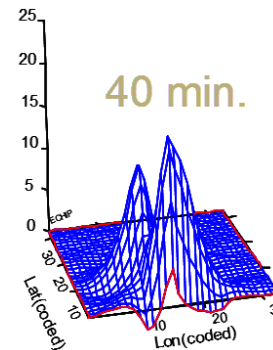
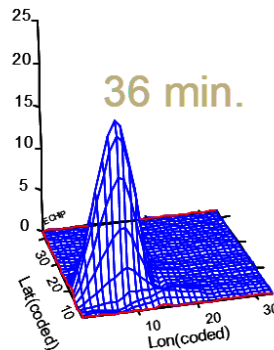
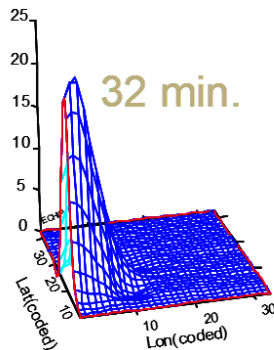
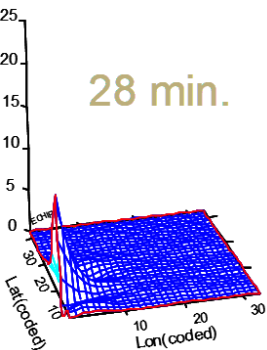
- Functional data shows up in many forms such as sensor data, spectral data, simulation data - almost any response in a longitudinal order
- These data are often summarized to allow for “landmark” analysis. This approach does not take advantage of all the data that has been collected and can lead to missing out on effects of the shape of data.
- When Functional Data Analysis of a response is combined with Design of Experiments one can model the shape of the data stream as a function of the design factors.
- One can use Machine Learning methods to fit the FPC scores derived from data streams (that characterize the run-to-run variation) to build predictive models.

First ran into Functional Data 18 Years ago at the  
Army's *Edgewood Chemical Biological Center*

Now...CCDC CBC



U.S. Army Combat Capabilities Development Command  
Chemical Biological Center



## 10-factor Agent Transport & Dispersion Simulation

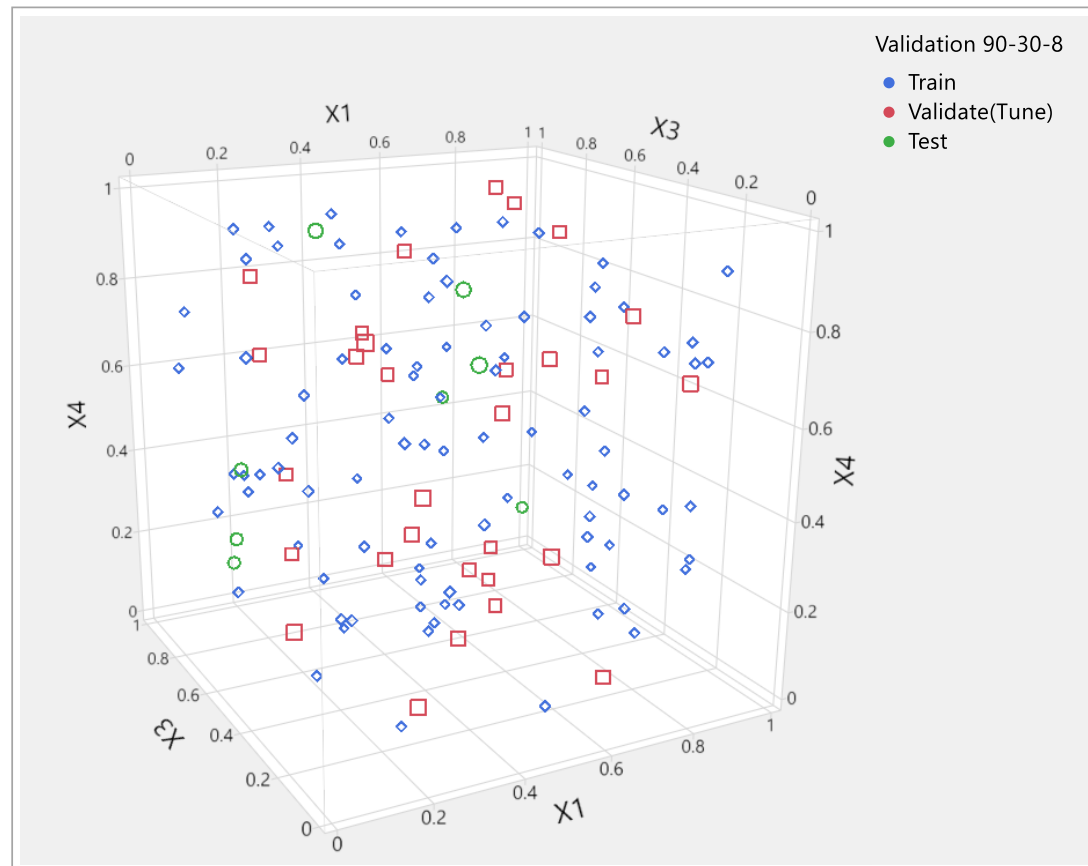
- Able to model Concentration *at a particular time*,
- or Dosage *at end of time*,
- but **NOT** Concentration *shape over time*
- Profs. Jeff Wu & Roshan Joseph from Georgia Tech ISyE suggested using Functional Data Analysis



# Complex Case Study using Simulation Data

## 128-Trial Space-Filling DOE in Six Factors + Time

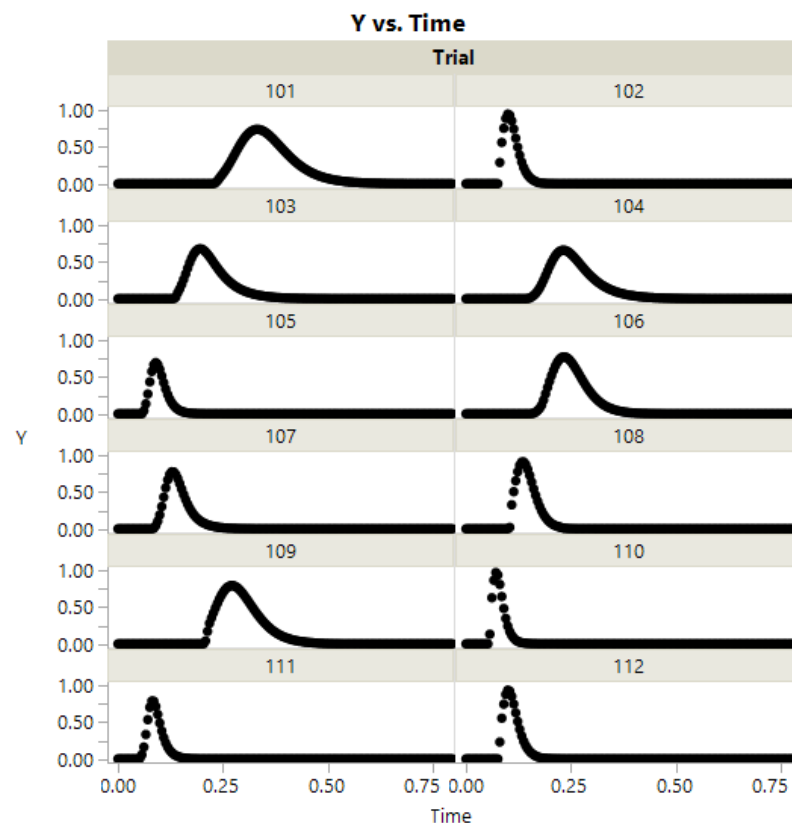
128 Computer Simulations Split into 3 Subsets:  
90 Training, 30 Validation(Tune), and 8 Test



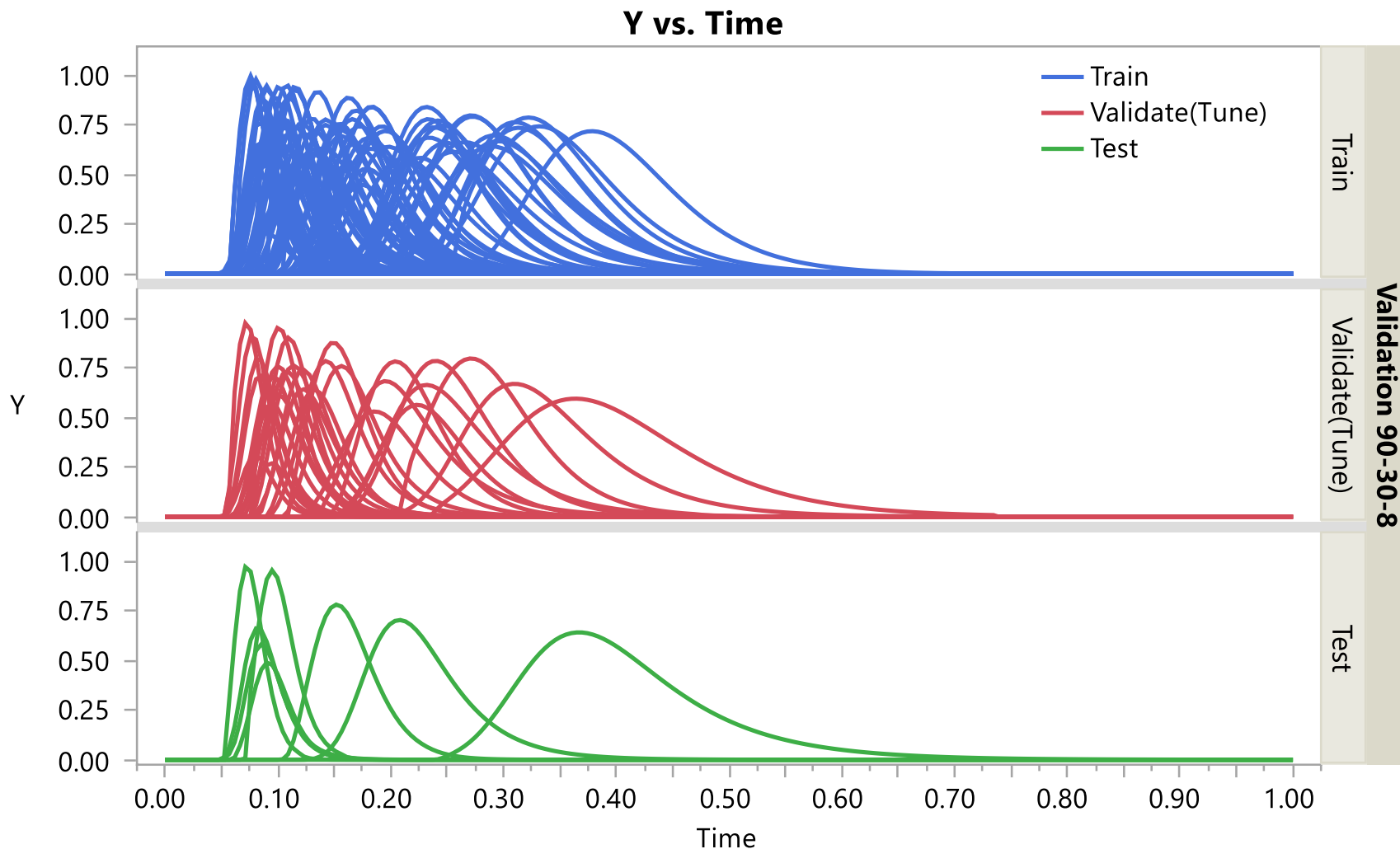
# 128 Unique-Trial Space-Filling Design of Experiments

|     | Trial | X1    | X2    | X3    | X4    | X5    | X6    |
|-----|-------|-------|-------|-------|-------|-------|-------|
| 101 | 101   | 0.244 | 0.469 | 0.000 | 0.393 | 0.500 | 0.000 |
| 102 | 102   | 0.983 | 0.563 | 0.638 | 0.543 | 0.500 | 0.500 |
| 103 | 103   | 0.031 | 0.094 | 0.234 | 0.259 | 0.625 | 0.500 |
| 104 | 104   | 0.158 | 0.719 | 0.170 | 0.836 | 1.000 | 1.000 |
| 105 | 105   | 0.638 | 0.188 | 0.894 | 0.031 | 0.750 | 0.500 |
| 106 | 106   | 0.228 | 0.813 | 0.170 | 0.039 | 0.750 | 0.000 |
| 107 | 107   | 0.858 | 0.031 | 0.468 | 0.660 | 0.375 | 0.500 |
| 108 | 108   | 0.787 | 0.938 | 0.404 | 0.552 | 0.125 | 0.000 |
| 109 | 109   | 0.220 | 0.094 | 0.064 | 0.560 | 0.125 | 0.500 |
| 110 | 110   | 0.606 | 0.906 | 1.000 | 0.504 | 0.625 | 0.500 |
| 111 | 111   | 0.488 | 0.938 | 0.894 | 0.646 | 0.125 | 1.000 |
| 112 | 112   | 0.433 | 0.375 | 0.638 | 0.521 | 0.000 | 1.000 |

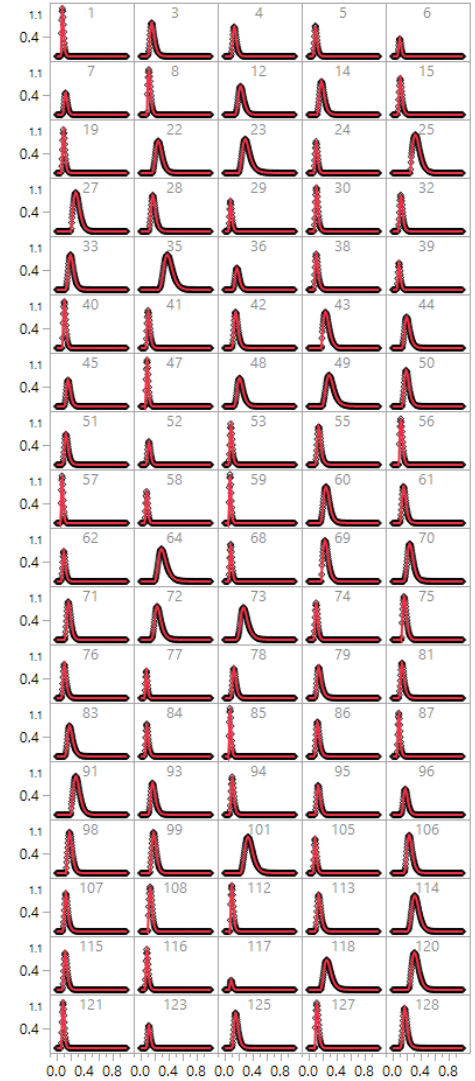
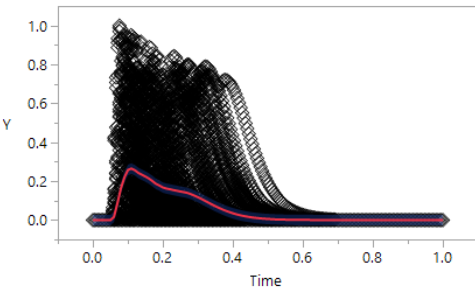
## Y vs Time Data for Each Trial



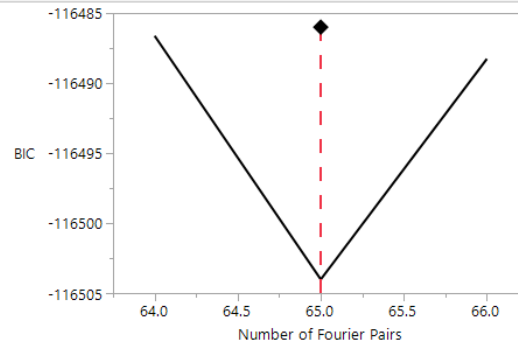
# 128 Simulations Split into 3 Subsets: 90 Training, 30 Validation(Tune), and 8 Test



# Model Selection



Legend  
— Prediction



Legend  
— Period 1

## Fit Statistics

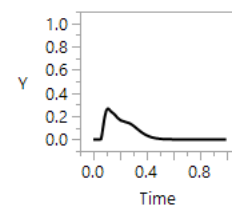
|                   |           |
|-------------------|-----------|
| Pairs             | 65        |
| -2 Log Likelihood | -119076.5 |
| AICc              | -118547.3 |
| BIC               | -116504   |
| GCV               | 0.0002135 |
| Y Std Dev         | 0.003181  |

## Summaries

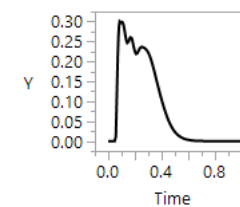
### Overall

|                    |           |
|--------------------|-----------|
| Observations       | 19080     |
| Functions          | 90        |
| Mean               | 0.0518738 |
| Standard Deviation | 0.1545753 |
| Minimum            | 0         |
| Maximum            | 1.0000967 |

### Mean



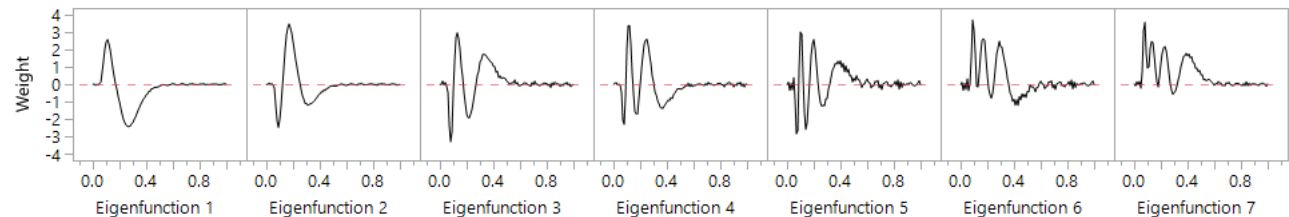
### Standard Deviation



| FPC | Eigenvalue | 20 | 40 | 60 | 80 | Percent | Cumulative |
|-----|------------|----|----|----|----|---------|------------|
| 1   | 0.00771    |    |    |    |    | 42.8%   | 42.8%      |
| 2   | 0.00514    |    |    |    |    | 28.5%   | 71.3%      |
| 3   | 0.00240    |    |    |    |    | 13.3%   | 84.6%      |
| 4   | 0.00155    |    |    |    |    | 8.61%   | 93.2%      |
| 5   | 0.00054    |    |    |    |    | 3.02%   | 96.2%      |
| 6   | 0.00032    |    |    |    |    | 1.78%   | 98%        |
| 7   | 0.00019    |    |    |    |    | 1.06%   | 99.1%      |

## Function Summaries

| Trial | Validation | FPC 1     | FPC 2     | FPC 3     | FPC 4     | FPC 5     | FPC 6     | FPC 7     |
|-------|------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 1     | Training   | 0.0643203 | -0.079624 | -0.076694 | -0.044372 | -0.022896 | 0.0048829 | 0.0191508 |
| 3     | Training   | 0.0132317 | 0.1034896 | 0.0192129 | -0.027173 | -0.016538 | 0.0118134 | 0.0020169 |
| 4     | Training   | 0.049788  | 0.0464589 | 0.0466466 | 0.0026321 | -0.026482 | -0.012951 | -0.003611 |
| 5     | Training   | 0.065293  | -0.053398 | -0.022798 | 0.0007321 | 0.0107226 | -0.00163  | -0.008597 |
| 6     | Training   | 0.0397146 | -0.036641 | -0.011642 | -0.00407  | 0.0077336 | -0.028904 | -0.036985 |
| 7     | Training   | 0.0466078 | 0.0049868 | 0.0310672 | 0.0081664 | -0.015096 | -0.031369 | -0.021895 |
| 8     | Training   | 0.0940523 | -0.004455 | 0.062271  | 0.060297  | 0.0065529 | -0.00089  | 0.0111631 |

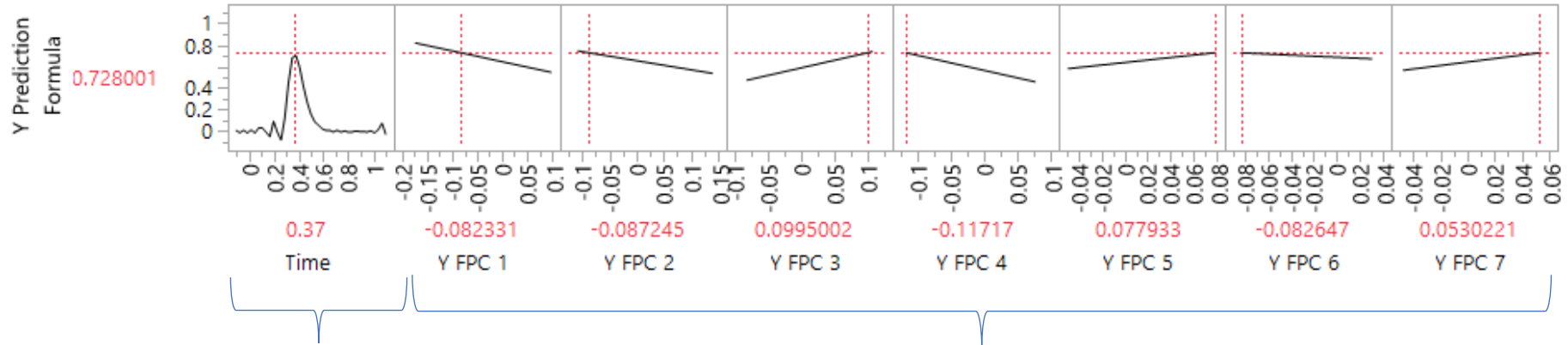


# Fourier Basis Model on Initial Data

90 Training,  
30 Validation,  
8 Test

# FPC Scores fit as function of DOE factors using Gaussian Process Model

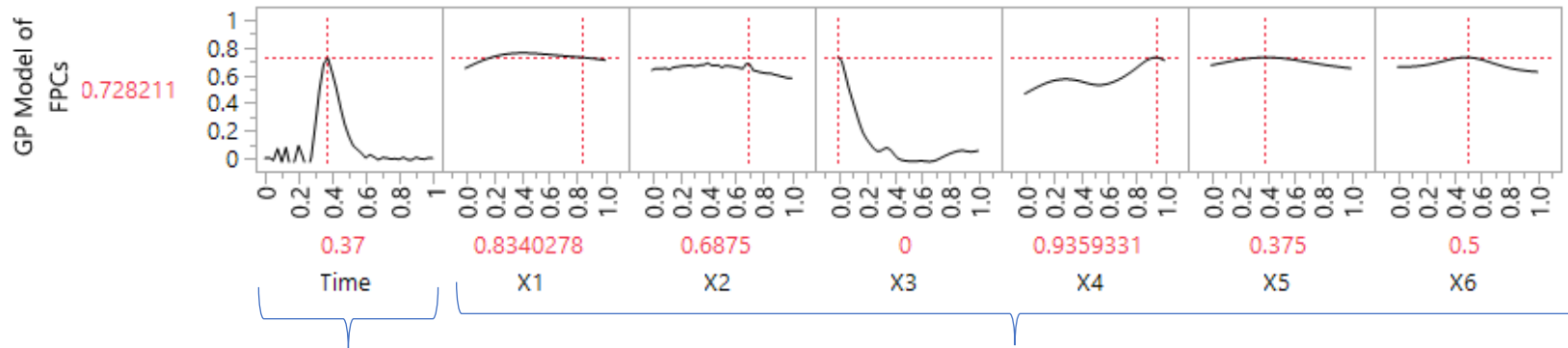
## Prediction Profiler



Functional  
Data **Curve**

FPC Scores

## Prediction Profiler

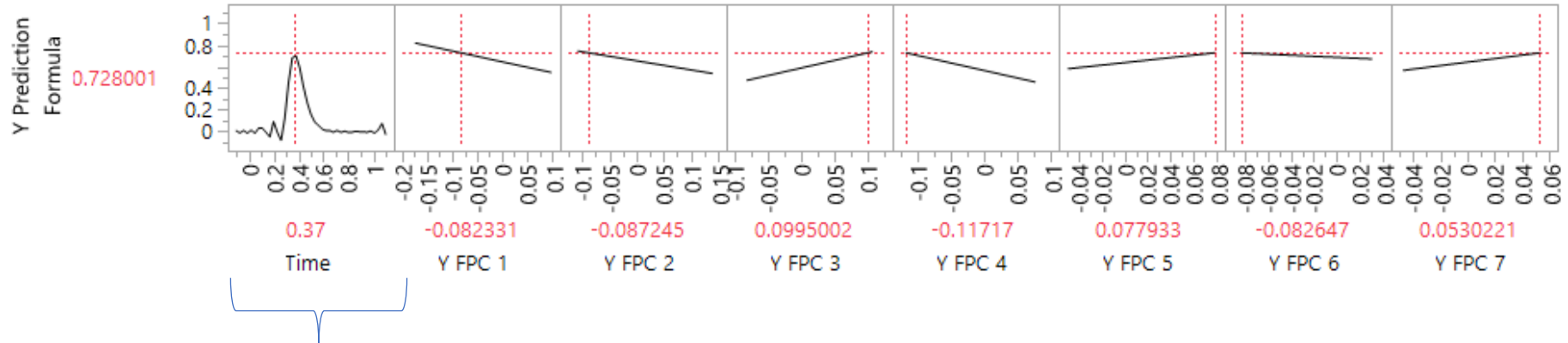


Functional  
Data **Curve**

DoE Factors

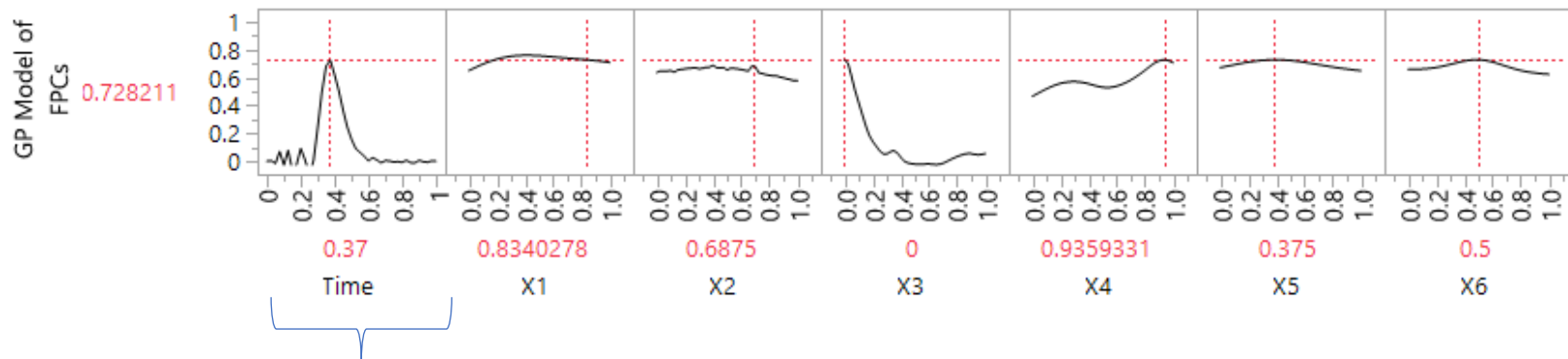
# FPC Scores fit as function of DOE factors using Gaussian Process Model

## Prediction Profiler



Functional Data **Curve** =  $\Sigma(\text{"Y}_i \text{ FPC Score"} * \text{"Y}_i \text{ Eigenfunction"}) + \text{"Y Mean Formula"}$

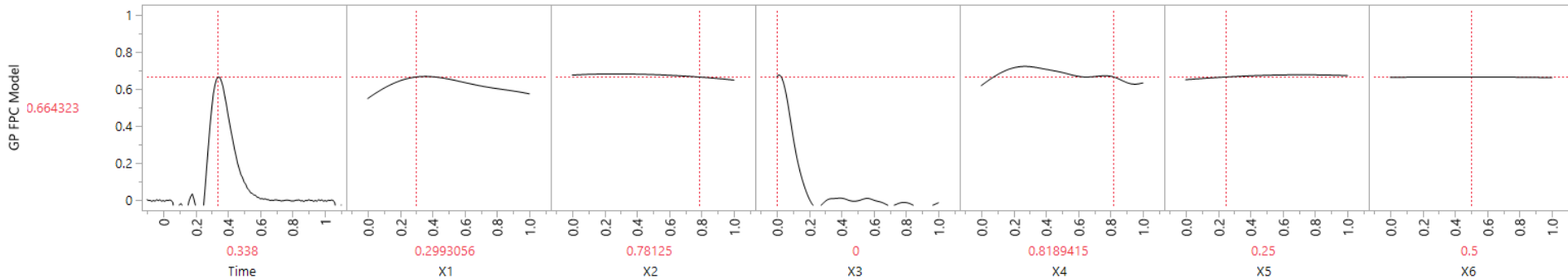
## Prediction Profiler



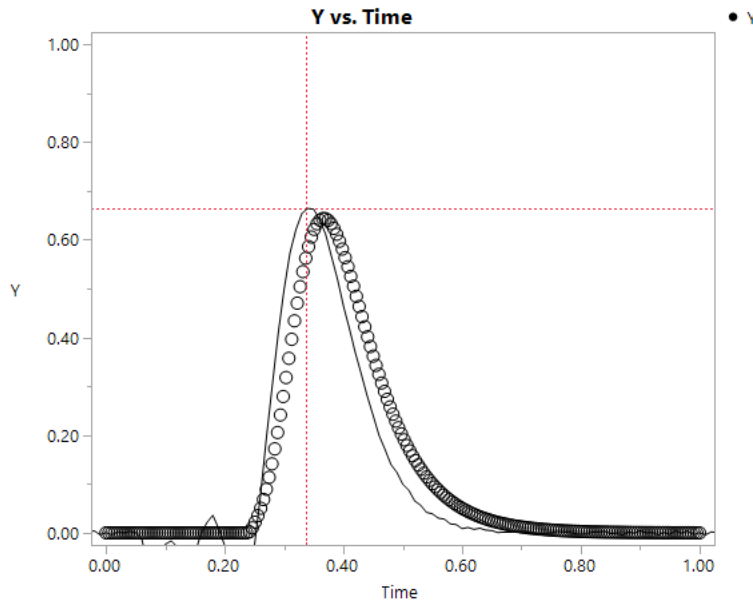
Functional Data **Curve** =  $\Sigma(\text{"Y}_i \text{ FPC Score Prediction Formula"} * \text{"Y}_i \text{ Eigenfunction"}) + \text{"Y Mean Formula"}$

# FPC Scores fit as function of DOE factors using Gaussian Process Model

Prediction Profiler



Functional Data *Curve*



DoE Factors

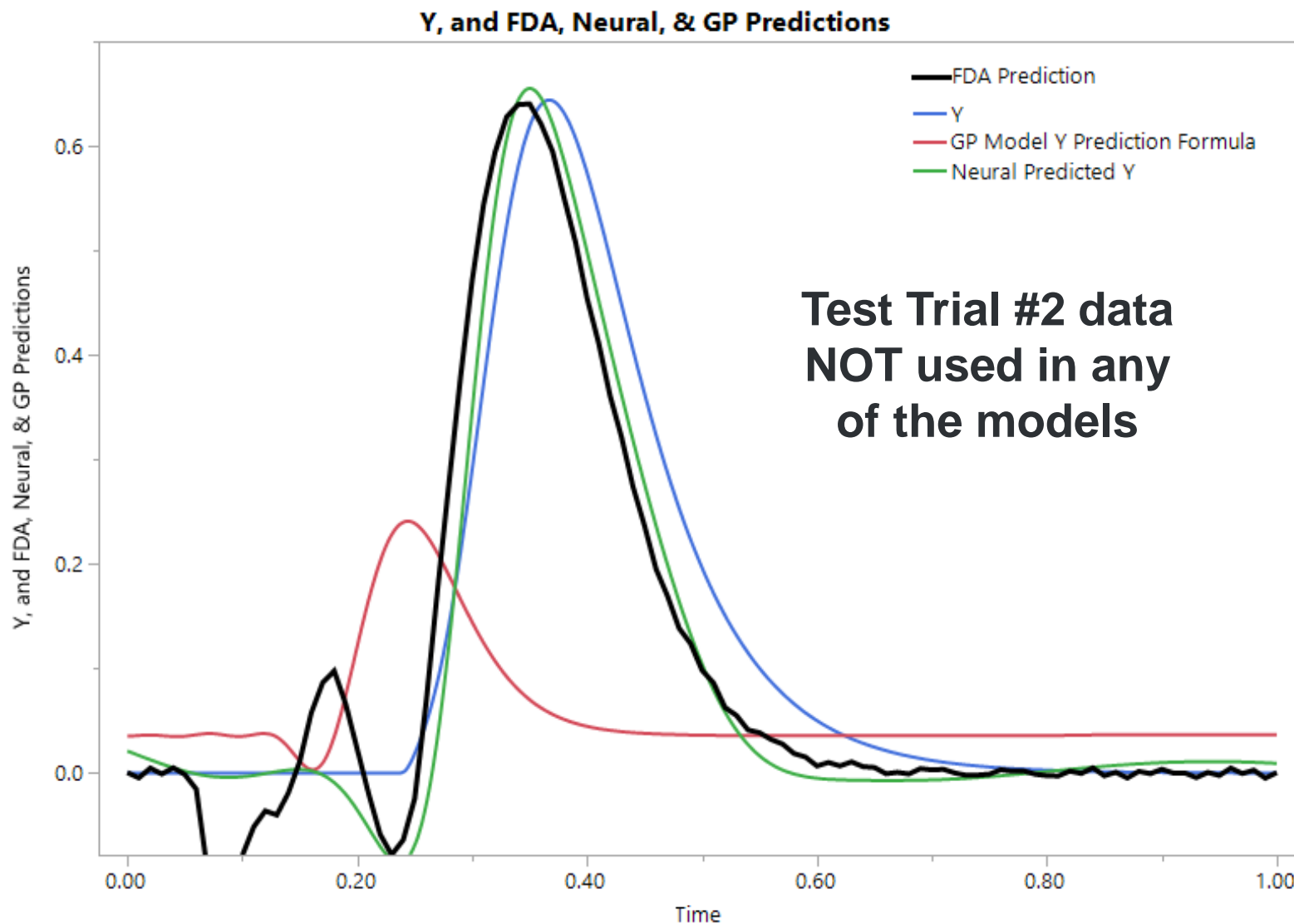
## Test Trial #2

Overlay of simulation data on top of Functional Data Curve

# FDA, Neural, & Gaussian Process Model Predictions

## - All Fit to Same 90-Trial *Training* Subset -

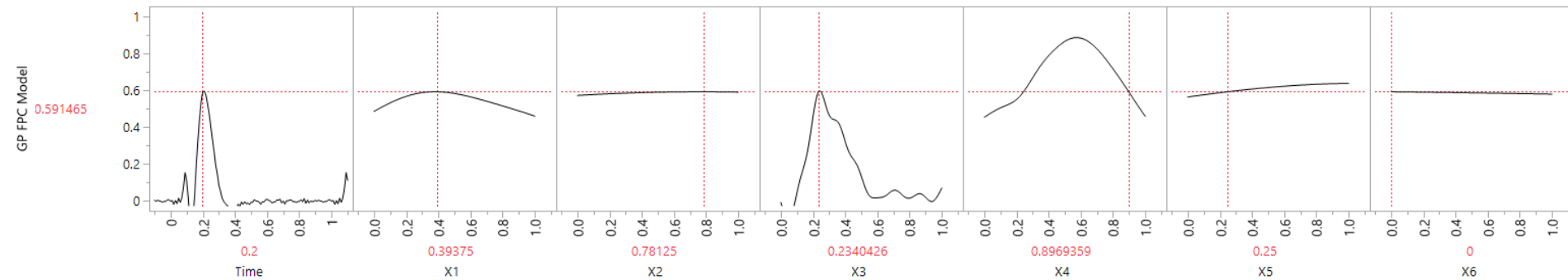
## Overlaid on Y vs. Time



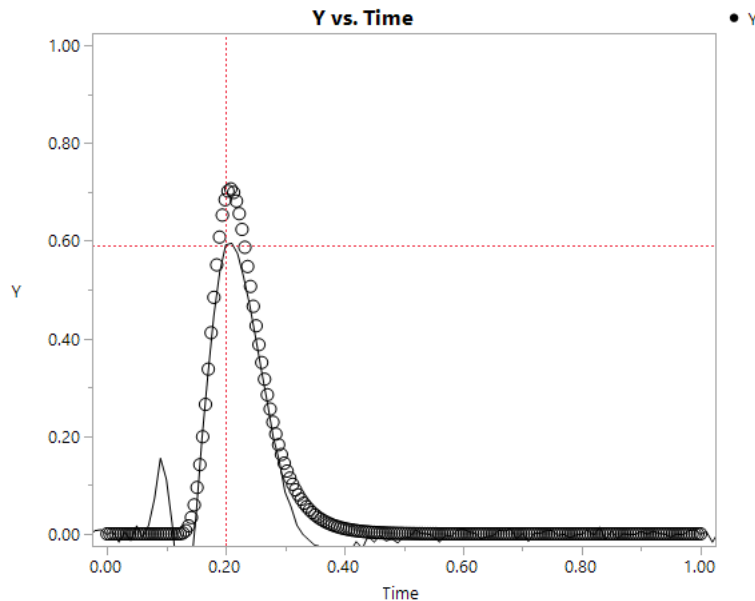


# FPC Scores fit as function of DOE factors using Gaussian Process Model

Prediction Profiler



Functional Data *Curve*



Where(Trial = 16)

DoE Factors

Test Trial #16

Overlay of simulation  
data on top of  
Functional Data  
Curve

erved.

# Use JMP to analyze Mill\_DOE.jmp data

1. Analyze > Specialized Modeling > Functional Data Explorer
2. Populate Dialog with Column Names > Click OK
3. Cleanup Data (Not required with these data)
4. Load Target Function – Batch 5000
5. Add Spec Limits to Size axis
6. Hot Spot Functional Data Explorer > Models > Model Controls > B-Spline Controls
7. Click Go
8. Inspect Function Summaries
9. Hot Spot Function Summaries > Customize Function Summaries > Deselect All > Check Save Formulas Click “OK” or “OK and Save”
10. Hot Spot B-Spline on Load Targets > Functional DOE Analysis
11. Hot Spot FDOE Profiler > Optimization and Desirability > Desirabilities Function
12. Hot Spot FDOE Profiler > Optimization and Desirability > Maximize Desirability
13. Hot Spot Customize Function Summaries > Deselect All > Check Save Formulas
14. Hot Spot Function Summaries > Save Summaries (If not done in step 9)
15. Hot Spot Functional Data Explorer > Save Script to Data Table

# Summary

- Functional data shows up in many forms such as sensor data, spectral data, simulation data - almost any response in a longitudinal order
- These data are often summarized to allow for “landmark” analysis. This approach does not take advantage of all the data that has been collected and can lead to missing out on effects of the shape of data.
- When Functional Data Analysis of a response is combined with Design of Experiments one can model the shape of the data stream as a function of the design factors.
- One can use Machine Learning methods to fit the FPC scores derived from data streams (that characterize the run-to-run variation) to build predictive models.

Links to additional content at [www.jmp.com/fedgov](http://www.jmp.com/fedgov)

### ***Modeling Streamed Sensor Data with Functional Data Analysis***

MORS Data Science & Artificial Intelligence CoP live virtual presentation at 1130 EST on January 27th:

- Weblink: <https://www.gotomeet.me/MORSMeeting50a/data-science-and-ai>
- Conference Line: +1 (669) 224-3412
- Access Code: 716-833-909
- Recording will be posted here as soon as it becomes available.

### ***Efficient M&S Using Sequential DOE Methods***

MORS Modeling & Simulation CoP virtual presentation on 11 March 2020:

[Watch Video](#)

### ***Moving from Data to Decision Faster - Using JMP®***

a 5-minute video highlighting JMP end-to-end analytic workflow

USCENTCOM DATA SYMPOSIUM Feb. 2-4.

[Watch Video](#)

Recordings from Nine December 2020 Webcasts - Using JMP 15 and JMP PRO 15

Click the "[underlined blue title](#)" in each cell to go to the video recording.

| <b><i>Graph Builder</i></b>  | <b><i>Text Exploration</i></b>                                     | <b><i>Functional Data Analysis</i></b>   |
|--|--|--|
| <b><i>Tour of Graph Elements</i></b><br><a href="#">Video GB-1</a>             | <b><i>Intro to Visualization and Modeling</i></b>                  | <b><i>Functional DOE – with Target Function</i></b><br><a href="#">Video FDA-1</a> |
| <b><i>Modeling with GB</i></b><br><a href="#">Video GB-2</a>                   | <b><i>Prepping the Data</i></b><br><a href="#">Video TXT-2</a>     | <b><i>Prepping the Data</i></b><br><a href="#">Video FDA-2</a>                     |
| <b><i>Adding Maps, Images, and Animation</i></b><br><a href="#">Video GB-3</a> | <b><i>Analyses with JMP Pro</i></b><br><a href="#">Video TXT-3</a> | <b><i>Functional Machine Learning</i></b><br><a href="#">Video FDA-3</a>           |

Recording and Slides from August 21st MORS-Talk

## **Text Analytics - Learning from Unstructured Data**

[Watch Video](#) or [Download Slides](#)

Recordings from August 13th Statistically Speaking

## **Demystifying Machine Learning and Artificial Intelligence in the Defense Community**

[Keynote Video](#) - Dr. Laura Freeman, Director of the Intelligent Systems Lab at Virginia Tech's Hume Center for National Security & Technology

[Panel Discussion Video](#) - Dr. Laura Freeman, Dr Ray Hill from AFIT, and Dr. James Wisnowski from Adsurgo LLC

Recordings from Nine August 2020 Webcasts - Using JMP 15 and JMP PRO 15

Click the "[underlined blue title](#)" in each cell to go to the video recording.

| <b>Data Wrangling</b>   | <b>Machine Learning</b>  | <b>Design of Experiments</b>  |
|---|--|---|
| <b>Getting Data into JMP</b><br><a href="#">Video DW-1</a>        | <b>Intro &amp; Honest Assessment</b><br><a href="#">Video ML-1</a> | <b>Custom DOE – Making Design Fit Problem</b><br><a href="#">Video DOE-1</a>        |
| <b>Prepping Data</b><br><a href="#">Video DW-2</a>                | <b>ML Analyses with JMP</b><br><a href="#">Video ML-2</a>          | <b>Screening Designs – Get More for Less</b><br><a href="#">Video DOE-2</a>         |
| <b>Using Data Table Tools Addin</b><br><a href="#">Video DW-3</a> | <b>ML Analyses with JMP Pro</b><br><a href="#">Video ML-3</a>      | <b>Comparing Designs &amp; Fixing Broken Designs</b><br><a href="#">Video DOE-3</a> |

JMP Discovered webcast on May 19, 2020

## **Modeling Streamed Sensor Data with Functional Data Analysis**

[View full 1-hour video](#); or [Download Slides](#)

Pressed for time? View short case-studies featured in full 1-hour webcast.

[Functional Data Analysis – DOE](#) (5-min)

Case 1 - Predicting Shape of Sensor Stream using DOE & Golden Curve Analysis

[Functional Data Analysis – ML](#) (7-min)

Case 2 - Using the Sensor Stream as an Input to a Machine Learning Model

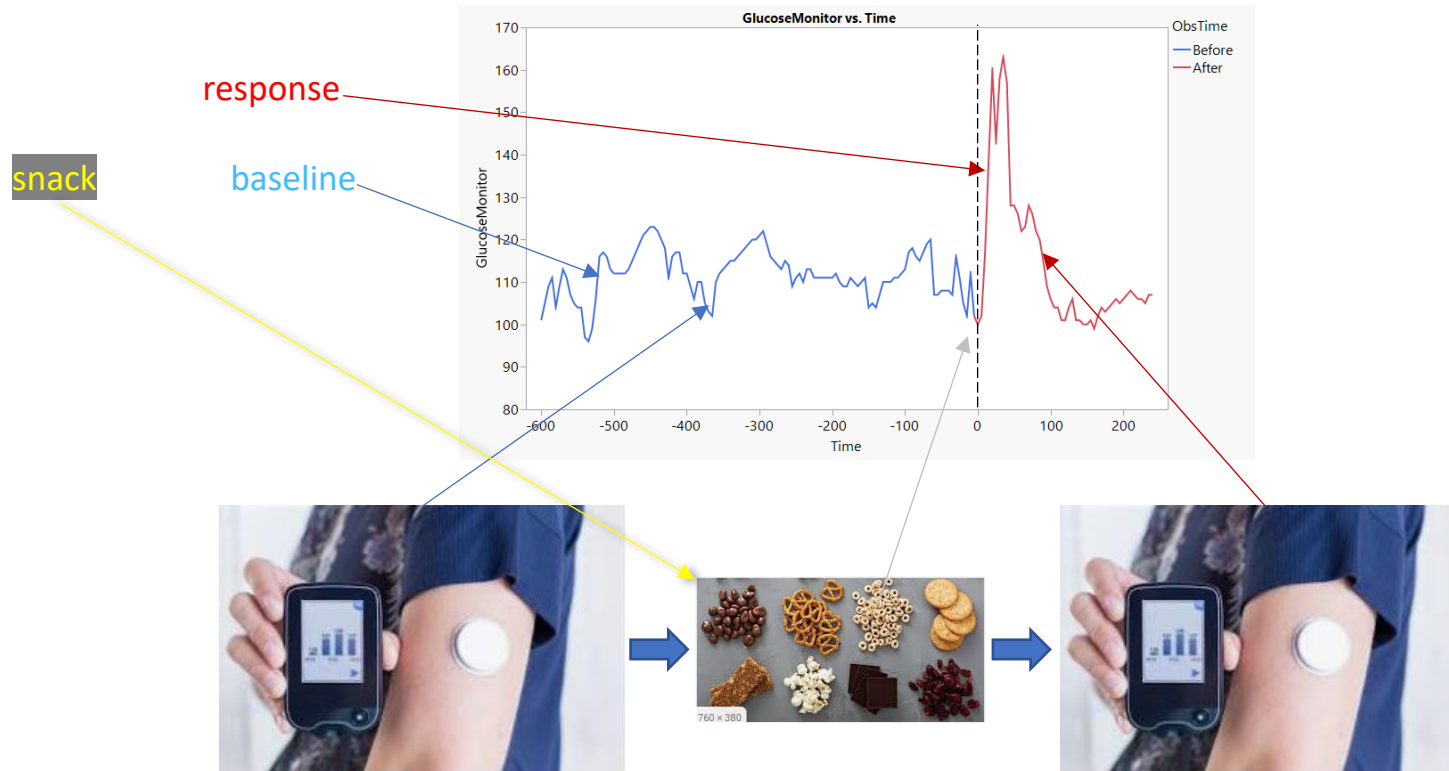
- Functional Data as both Input to & Output from a Glycemic Response Model

# The Goal – Build a Glycemic Response Model

We want to predict a patient's *response-over-time* after first fasting to establish a *baseline-over-time* and then consuming a **snack**.



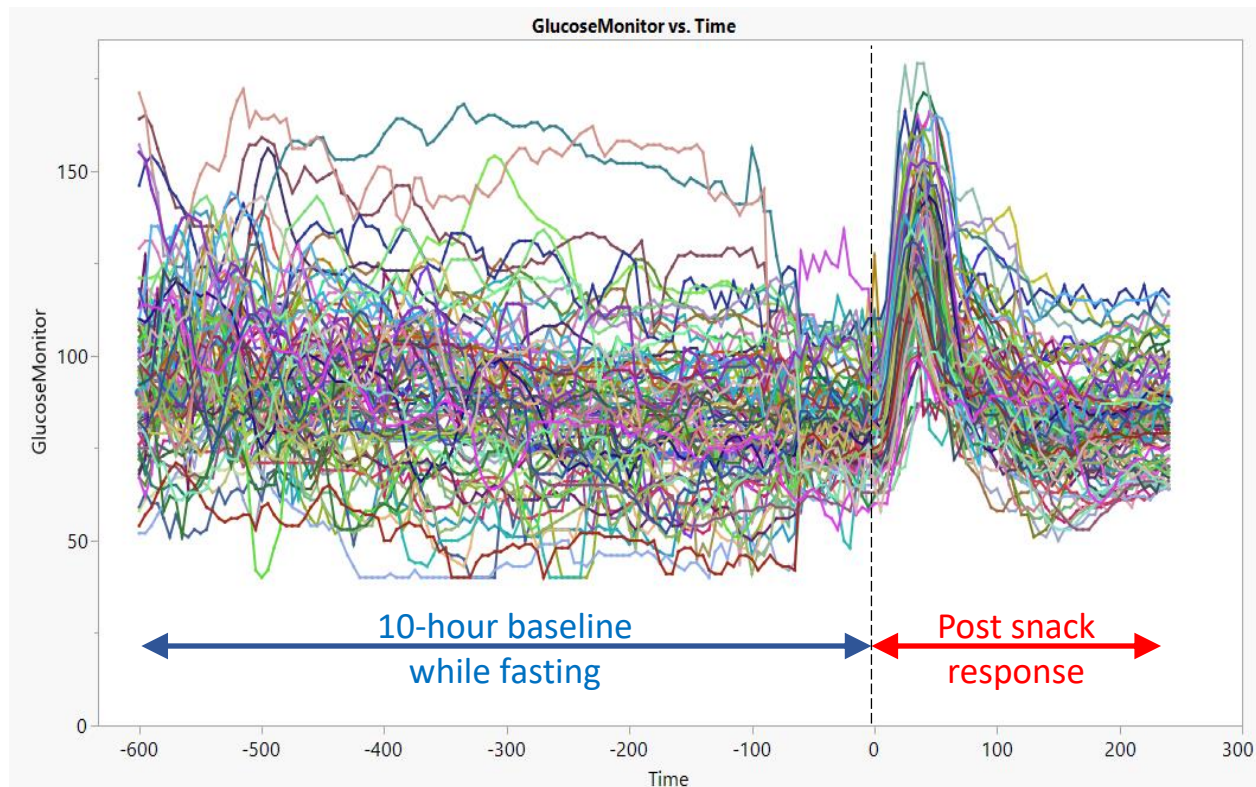
# Wearable Glucose Monitor Collects Glycemic Response Data Over Time



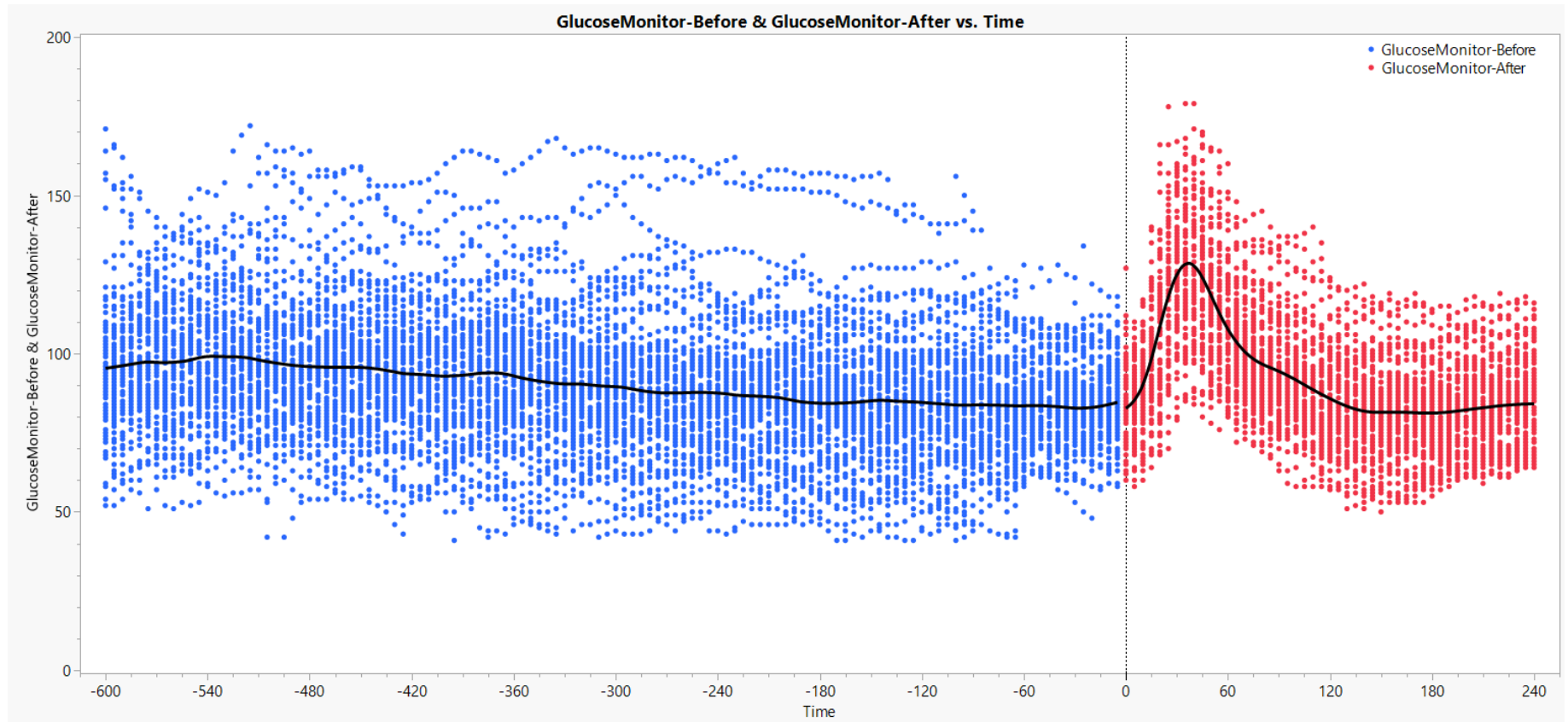


# Glucose Monitoring Case Study

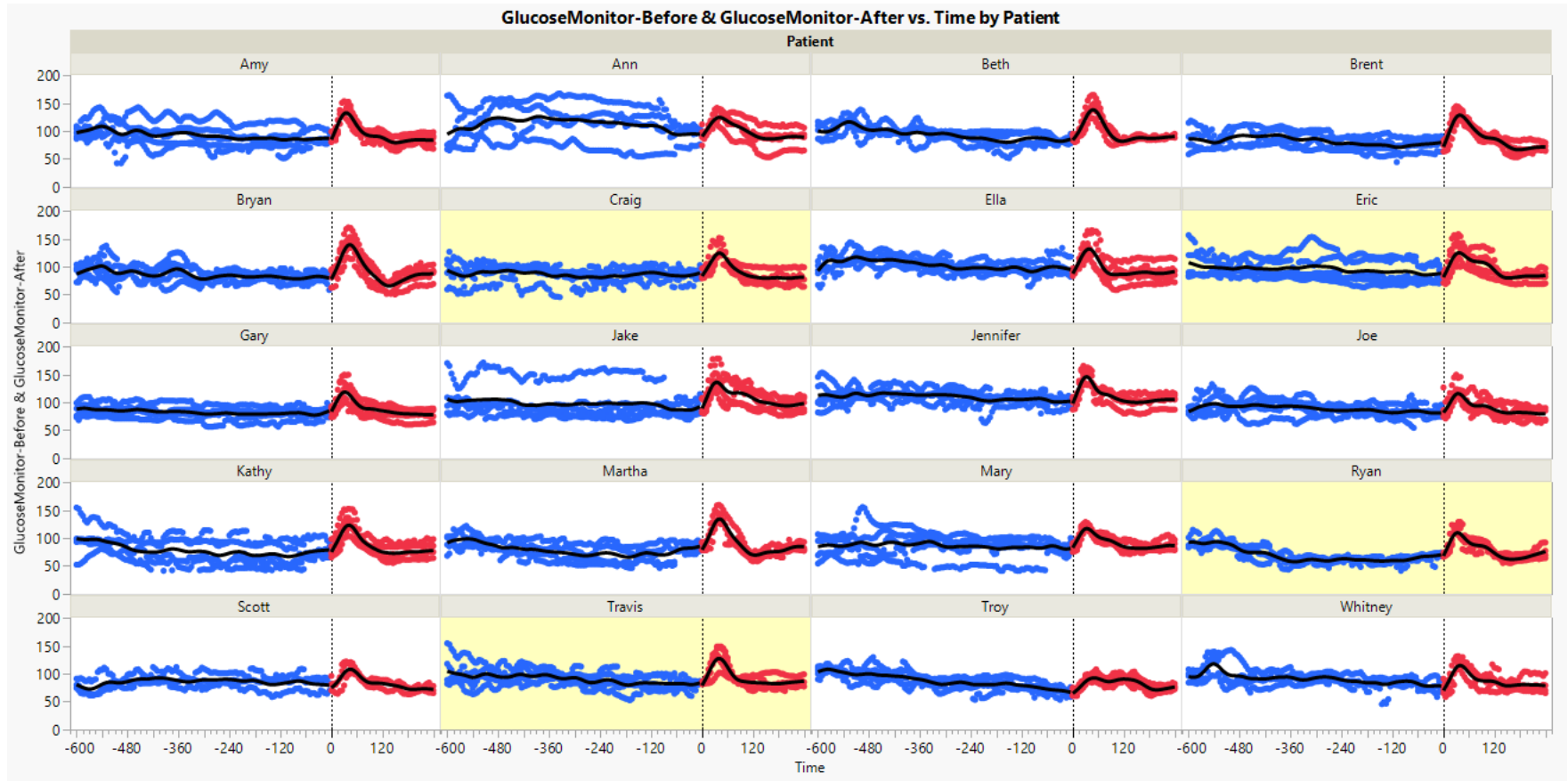
- Continuous glucose testing is used, with monitor reading updated every 5 minutes.
- This study has patients generate 10-hours of baseline data - while fasting - then eat a snack. The glucose levels are then measured for another 4 hours after the snack.
- Participants in the 14-hour study were given 1 of 5 different snacks. Patients repeated the testing with a total of 3, 4, or 5 snacks.
- **Goal is to predict the glycemic response “curve” – i.e., “shape-over-time” – in NEW patients from their baseline “curve” and type of snack consumed.**



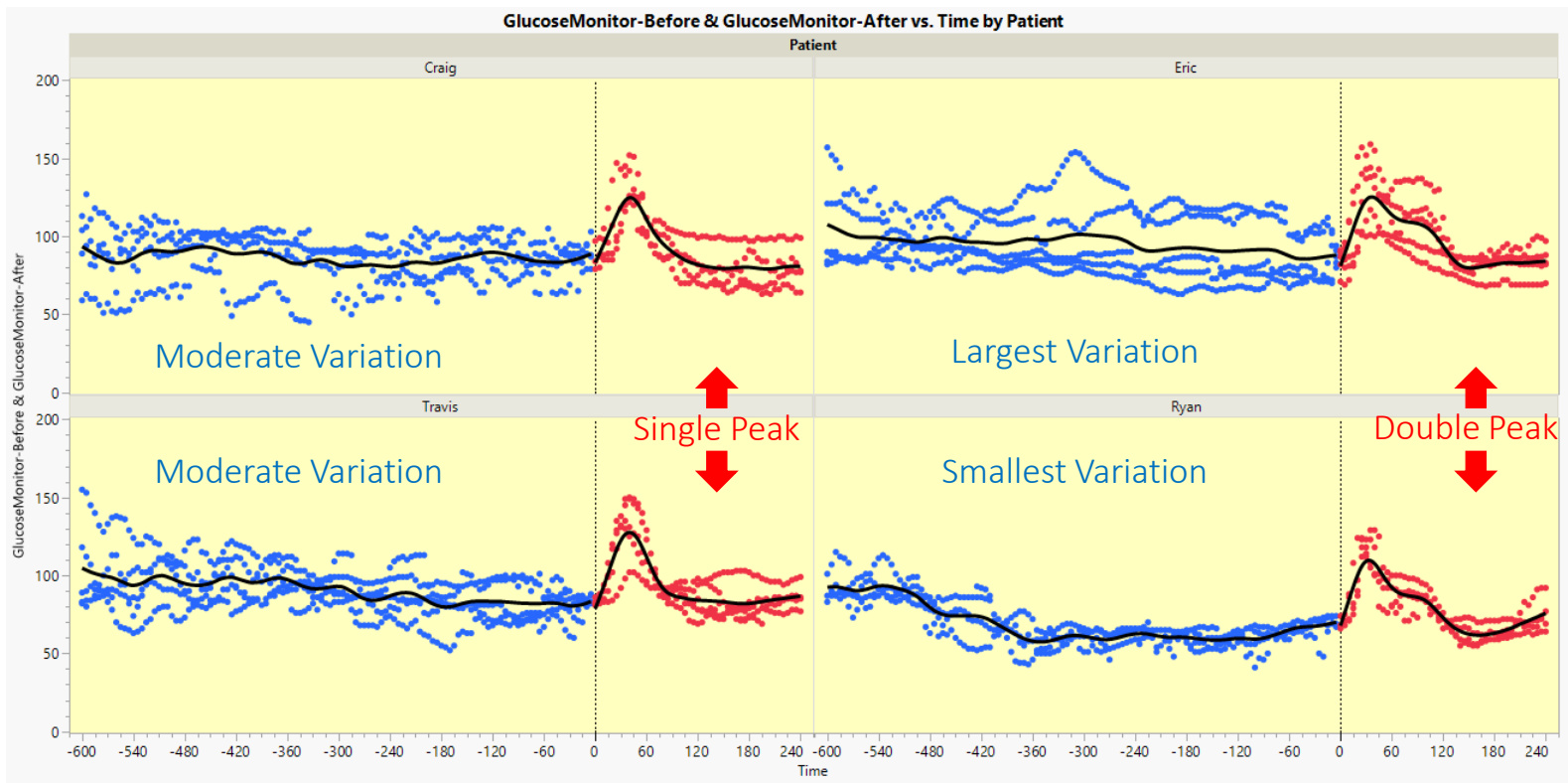
# Smoothed Glycemic Response across 91 Patient-Snack Combinations



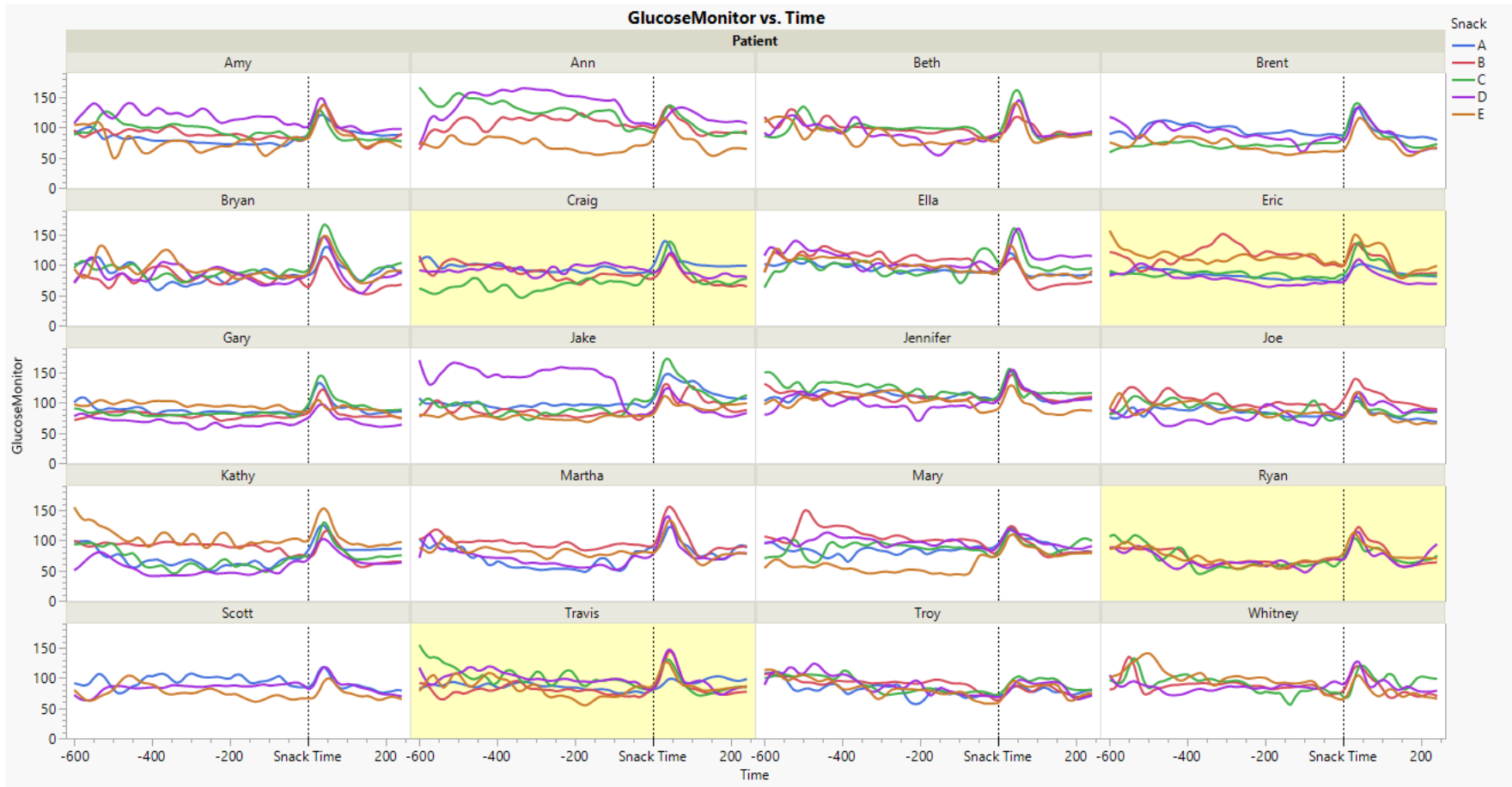
# Glycemic Response for 20 Patients, Each Eating 3, 4, or 5 Snacks



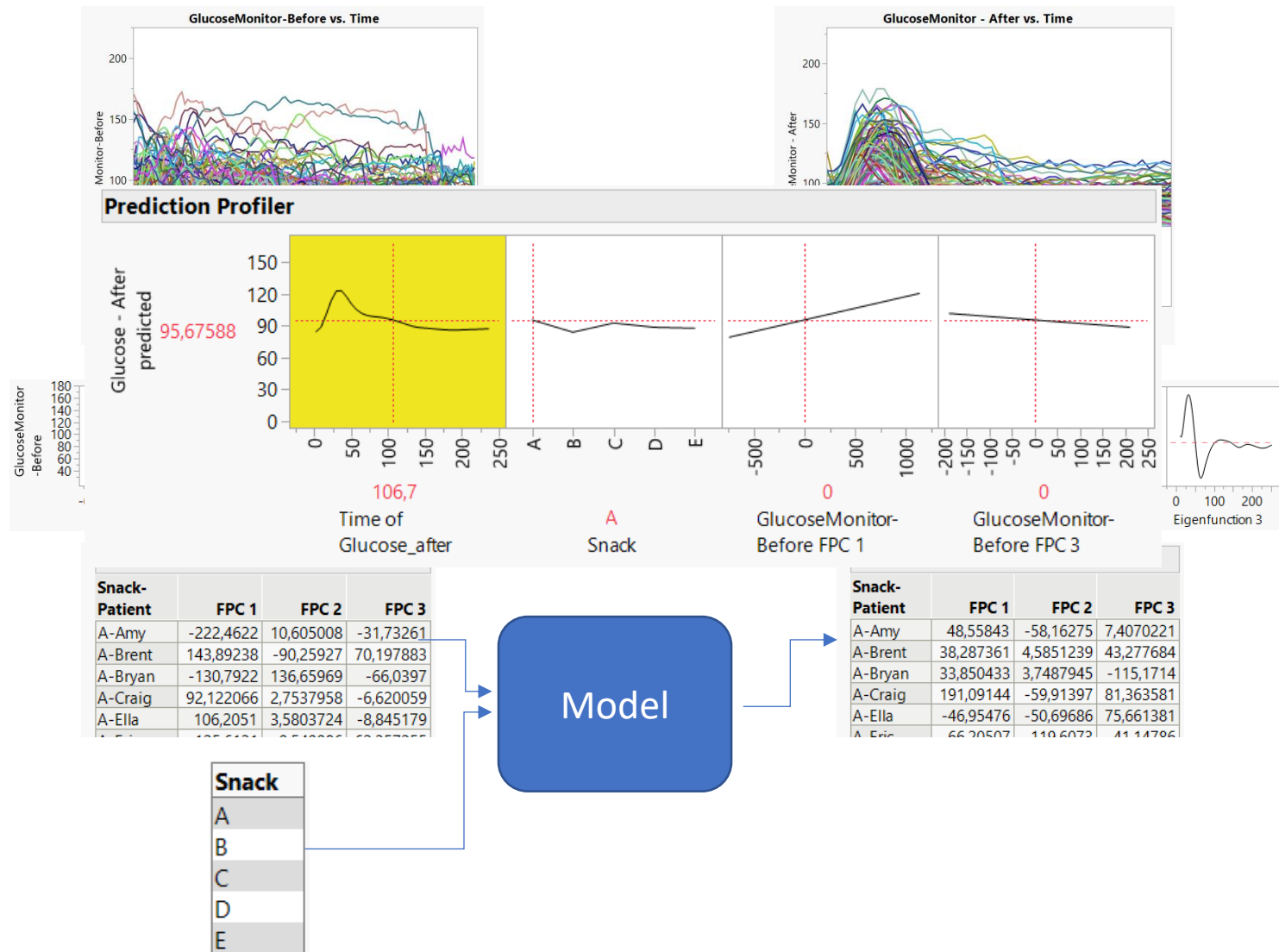
# Data for these Four Patients NOT Used in Fitting Model, BUT Used to Test Model Predictions



# Smoothed Data for Snacks by Patient

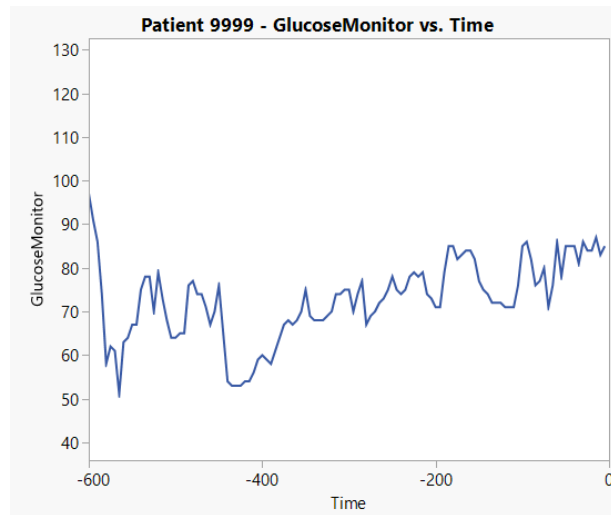


# Creating a Model to Predict Glucose Response:





# Applying that Model to New Patient Data

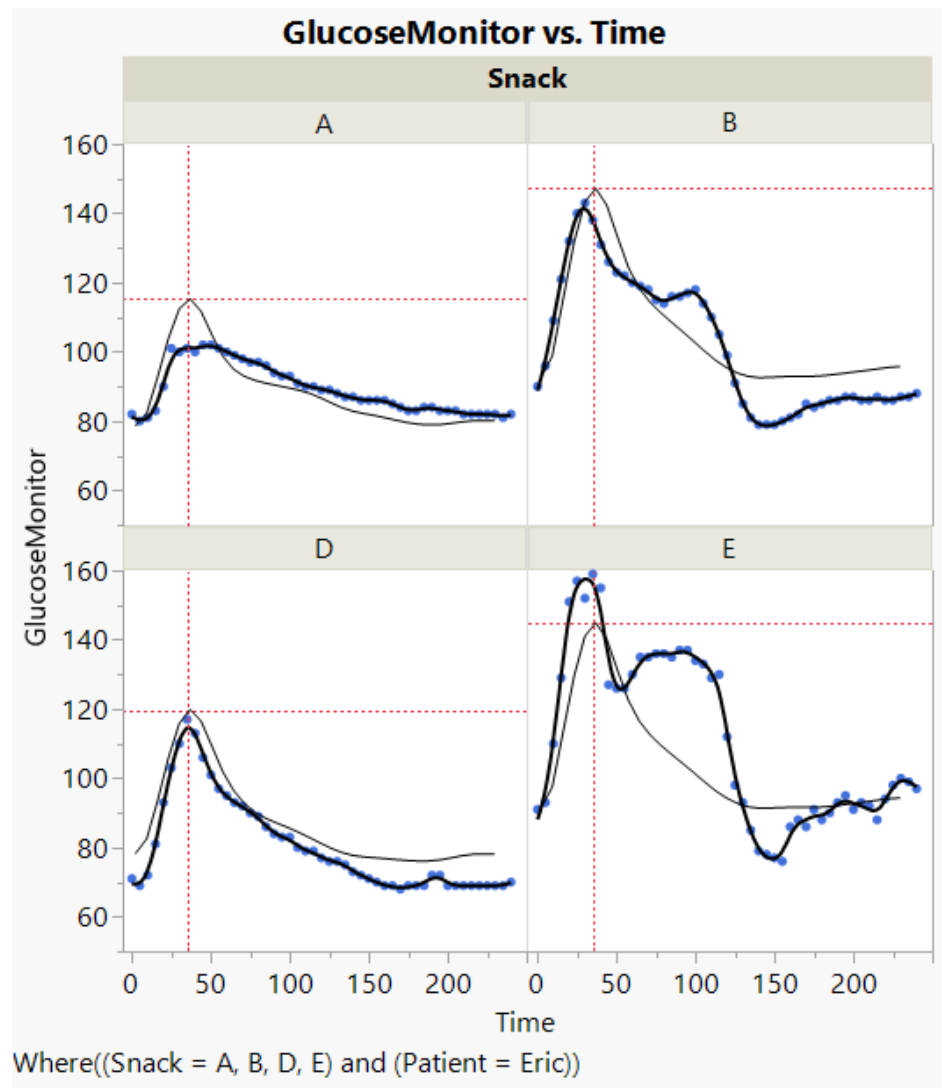


| Function Summaries |            |           |           |           |
|--------------------|------------|-----------|-----------|-----------|
| Snack-Patient      | Validation | FPC 1     | FPC 2     | FPC 3     |
| E-Martha           | Training   | -160,6882 | 31,287303 | 20,383365 |
| E-Mary             | Training   | -736,9591 | -55,79129 | 12,4177   |
| E-Ryan             | Training   | -443,8172 | 110,69514 | -27,43718 |
| E-Scott            | Training   | -243,6854 | -98,48814 | 210,66843 |
| E-Travis           | Training   | -165,9502 | 113,06902 | 94,797109 |
| E-Troy             | Training   | -130,2186 | 98,642639 | 45,688343 |
| E-Whitney          | Training   | 204,38139 | 144,27509 | -12,36567 |
| 9999               | Validation | -358,5495 | -189,2069 | -86,08579 |

| Snack |
|-------|
| A     |
| B     |
| C     |
| D     |
| E     |

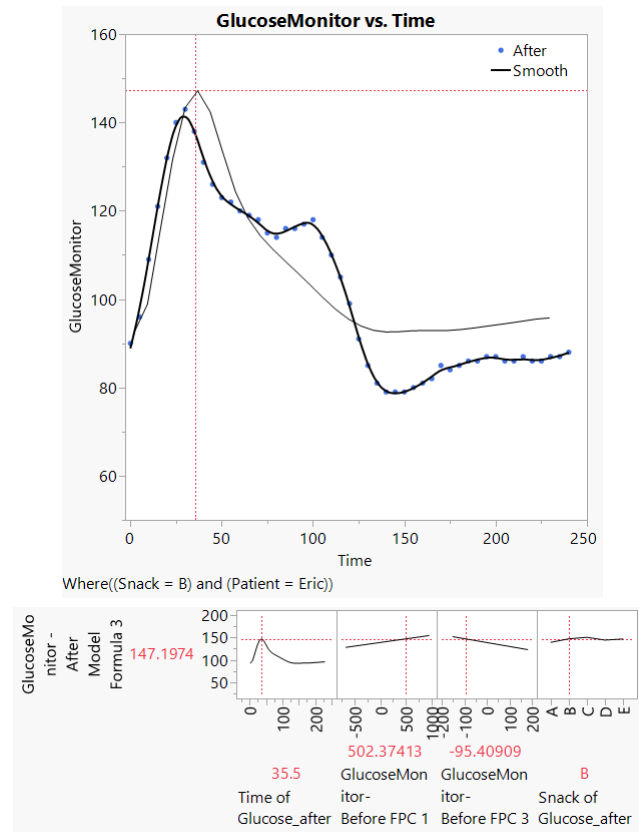
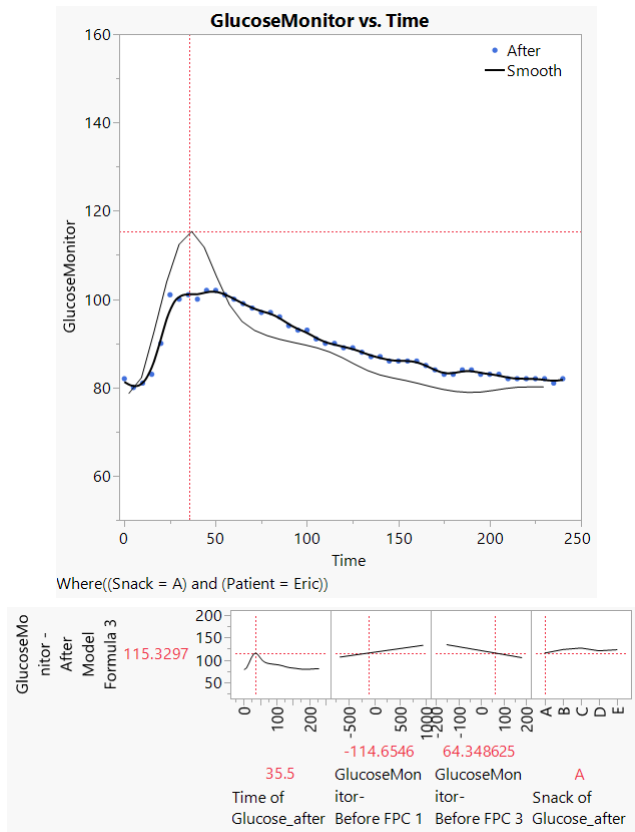


Overlay of predictions (thin curves) on top of actual observations (blue dots and thick curve) for held out patient Eric for snacks A, B, D, & E.

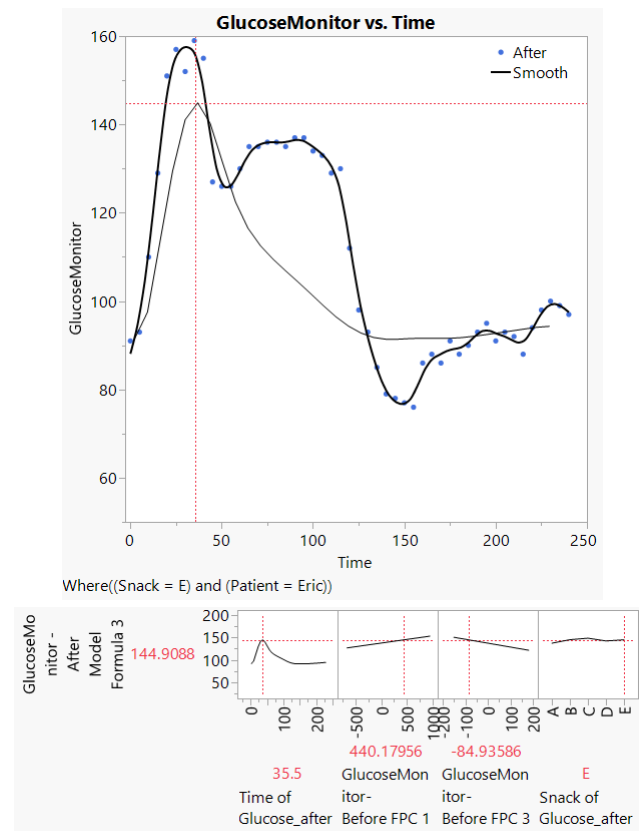
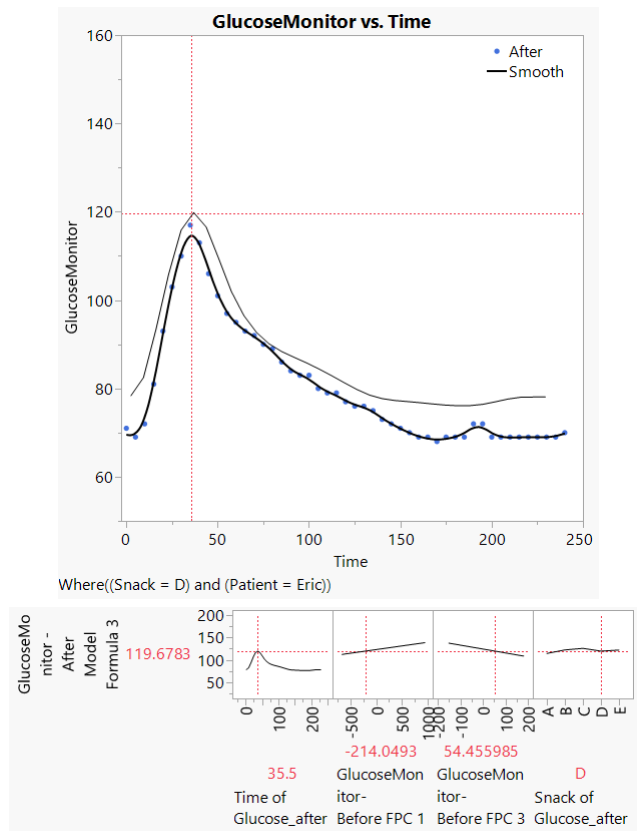




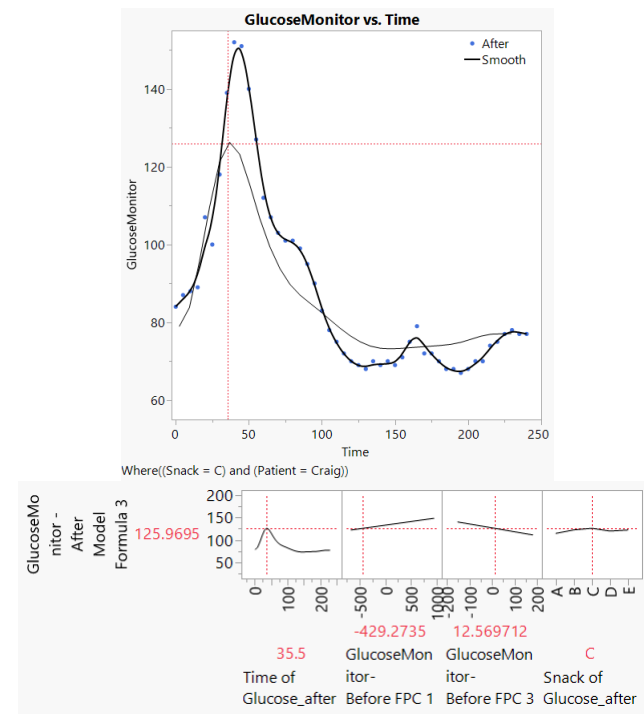
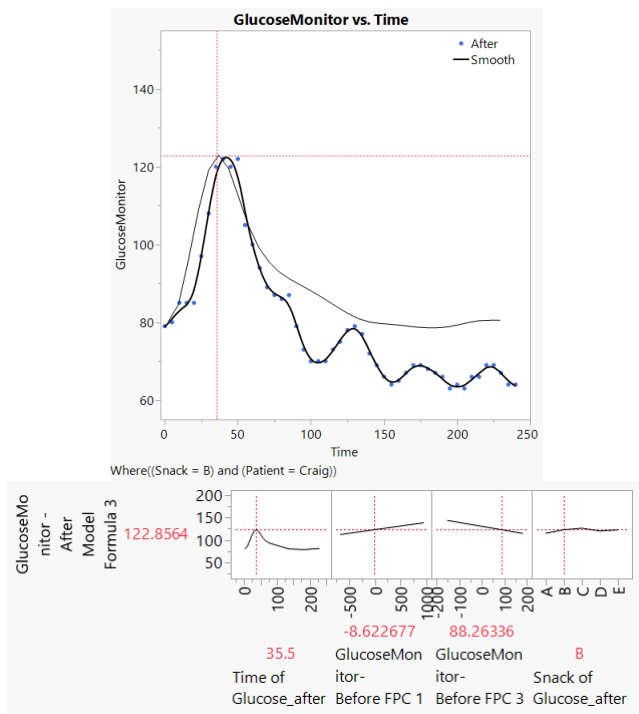
# Overlay of Predictions for Held out Patient Eric



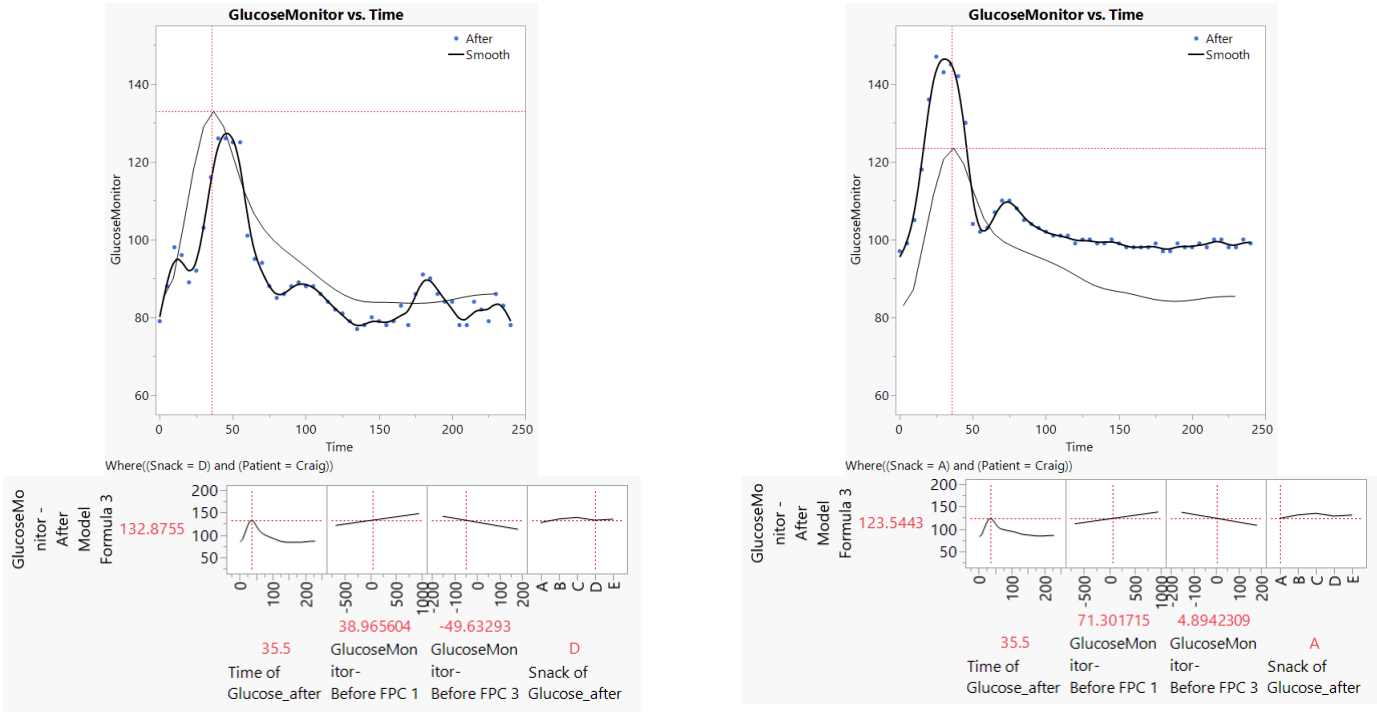
# Overlay of Predictions for Held out Patient Eric



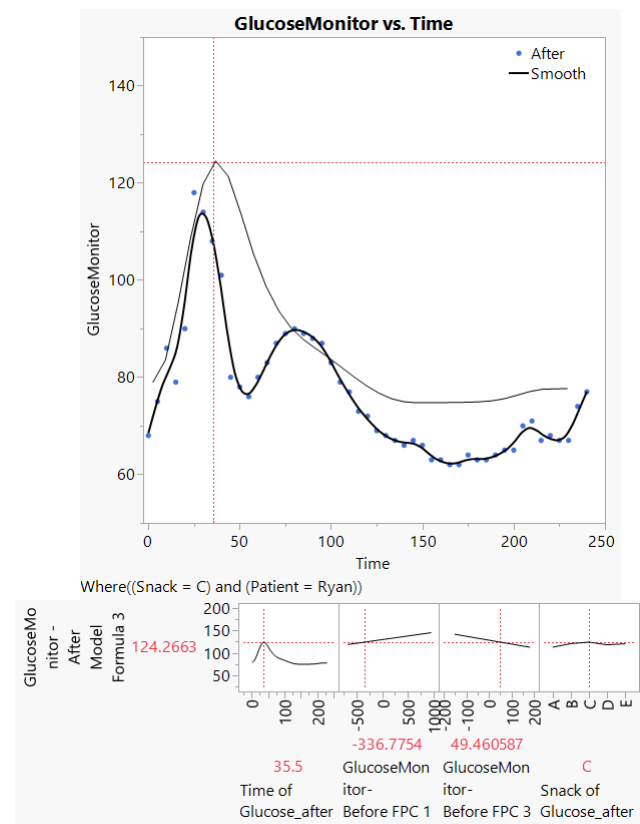
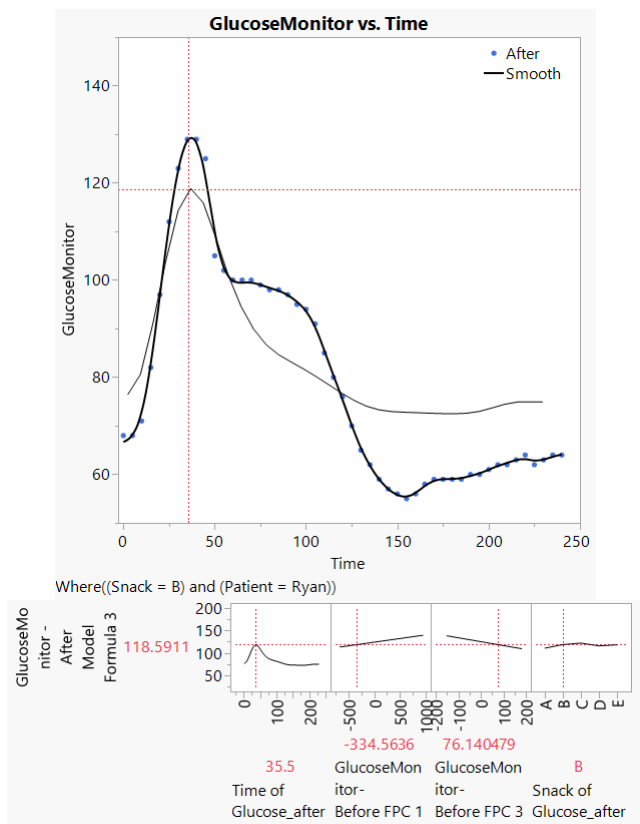
# Overlay of Predictions for Held out Patient Craig



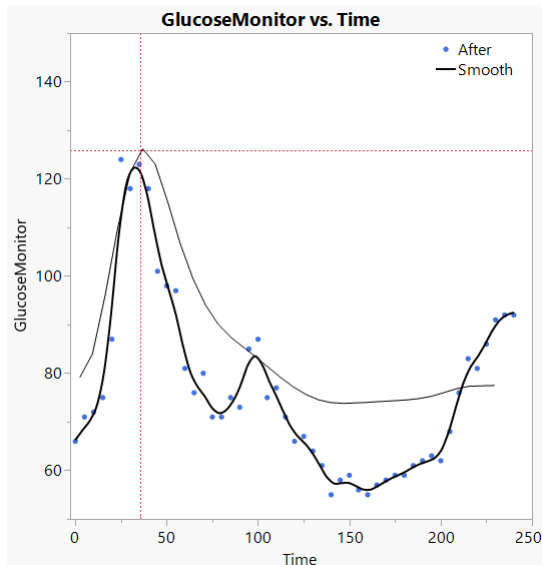
# Overlay of Predictions for Held out Patient Craig



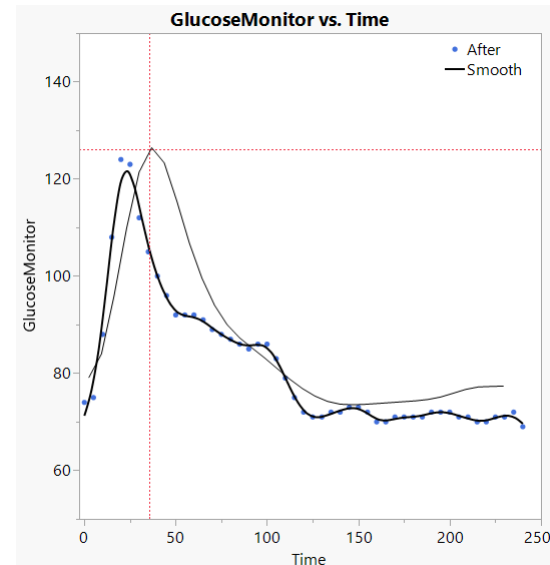
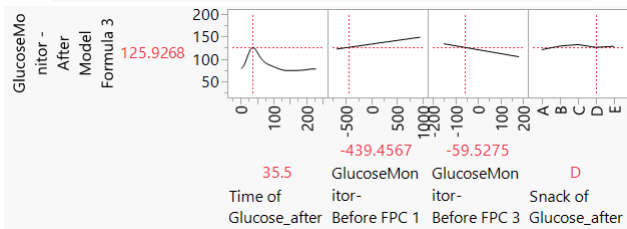
# Overlay of Predictions for Held out Patient Ryan



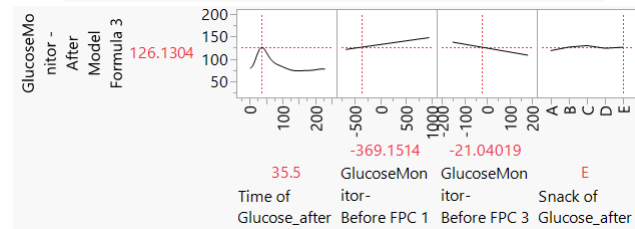
# Overlay of Predictions for Held out Patient Ryan



Where((Snack = D) and (Patient = Ryan))



Where((Snack = E) and (Patient = Ryan))



# Summary

- Glucose monitor data from a panel of patients were used to show how one can use functional data as both input to and output from the same model.
- The glycemic response over time of patients could be predicted using their baseline (fasting) curve and the type of snack consumed as inputs to the model.
- Functional data shows up in many forms such as sensor data, or spectral data, virtually any response in a longitudinal order.
- These data are often summarized to allow for “landmark” analysis. This approach does not take advantage of all the data that has been collected and can lead to missing out on effects of the shape of data.

# Thank You. Questions?

Webcast recordings at  
[www.jmp.com/fedgov](http://www.jmp.com/fedgov)

Thanks to my JMP colleagues  
upon whose work much of  
this presentation is based:

Chris Gotwalt  
Ryan Parker  
Brady Brady  
Pete Hersh  
Phil Kay

**Tom Donnelly, PhD, CAP**  
*Principal System Engineer*  
[tom.donnelly@jmp.com](mailto:tom.donnelly@jmp.com)  
302-489-9291