Form 712A Non-Government Disclosure (Contractor, Academic, UARC etc.)

1. Check the Event Website for the form deadline at www.mors.org.
2. If using the MORS Presenter Center all forms must be uploaded in the Presenter Center.
3. For other events e-mail the completed form to liz.marriott@mors.org.
4. If the presentation classification and or Disclosure Statement below is different than those entered in the MORS Presenter Center, correct the Presenter Center entries to ensure proper labeling of the Presentation.

PART I

Author Request - The following author(s) request authority to disclose the following presentation at the MORS Event below with subsequent publication in the MORS Event Report and posting on the MORS website if applicable.

Principal Author: Thomas A. Donnelly

Other Author(s):

Principal Author's Organization and complete mailing address:
SAS Institute Inc.
27 Farmingdale Ln
Newark, DE 19711

Principle Author Phone: 302-489-9291

Principle Author FAX:

Principle Author E-mail: tom.donnelly@jmp.com

Principal Author's Signature X: Thomas A. Donnelly

Date: 3 June 2021

MORS Event: 89th MORSS

Event Date(s): 21-24 June 2021

Presentation Type ☐ Plenary ☐ Course ☒ Tutorial ☐ Special Session ☐ Poster ☐ Demonstration

☐ Working/Composite/Distributed or Focus Group List All ☐ Other

Title of Presentation: Modeling with Streamed Sensor Data – "What to do when your data is a curve."

Presentation ID (if assigned): 56941

Classification ☐ SECRET ☐ SECRET//REL TO FVEY ☐ CONFIDENTIAL ☐ CONFIDENTIAL//REL TO FVEY

☒ UNCLASSIFIED ☐ UNCLASSIFIED W/FOUO ☐ Other

Distribution Statement ☒ A (Publicly Releasable) ☐ B ☐ C ☐ D ☐ E (see side 2 for definitions)

A.	This work was performed in connection with a government contract.	☐ YES (Complete Parts I, II & III)
B.	This presentation is based on material developed by the author as part of company/organization approved research e.g. IR&D and was NOT done under a government contract.	☐ YES (Complete Parts I, II & III)
C.	This presentation was NOT done under a government contract, contains no government information, is my own work and is approved for public release.	☒ YES (Complete Part I only)

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

89th MORSS

Webcast Tutorial 56941

June 24, 2021

Tom Donnelly, PhD, CAP

JMP Defense & Aerospace Team

Principal System Engineer

tom.donnelly@jmp.com

302-489-9291

Statistical Discovery.™ From SAS.

Outline

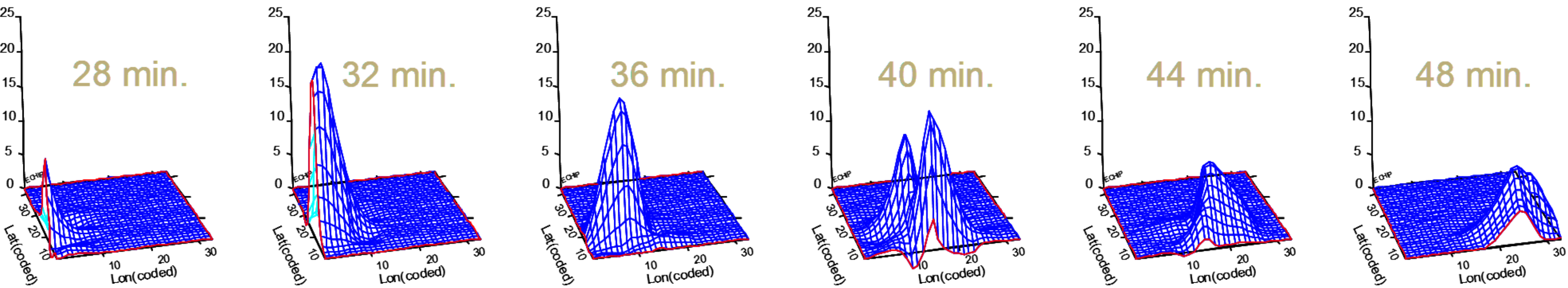
- Who is doing FDA?
- My old Army problem
- What are Examples of Functional Data?
- What is Functional Data Analysis (FDA)?
- How do we analyze functional data?
- How do we use Functional Principal Component (FPC) scores to model responses?
- **Simple case study with one FPC score**
- **More complex case study – predicting wafer condition from 5 sensor streams & 12 FPC scores**
- My old Army problem solved
- Summary
- Additional Resources

Who in DoD is Looking at or Already Using FDA?

- CCDC - Armaments Center
- Army Evaluation Center
- Eglin AFB
- Edwards AFB
- NAWCWD
- COTF
- NUWC
- MCOTEA
- IDA
- JHU/APL
- LMCO
- NGC

First ran into Functional Data 15 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

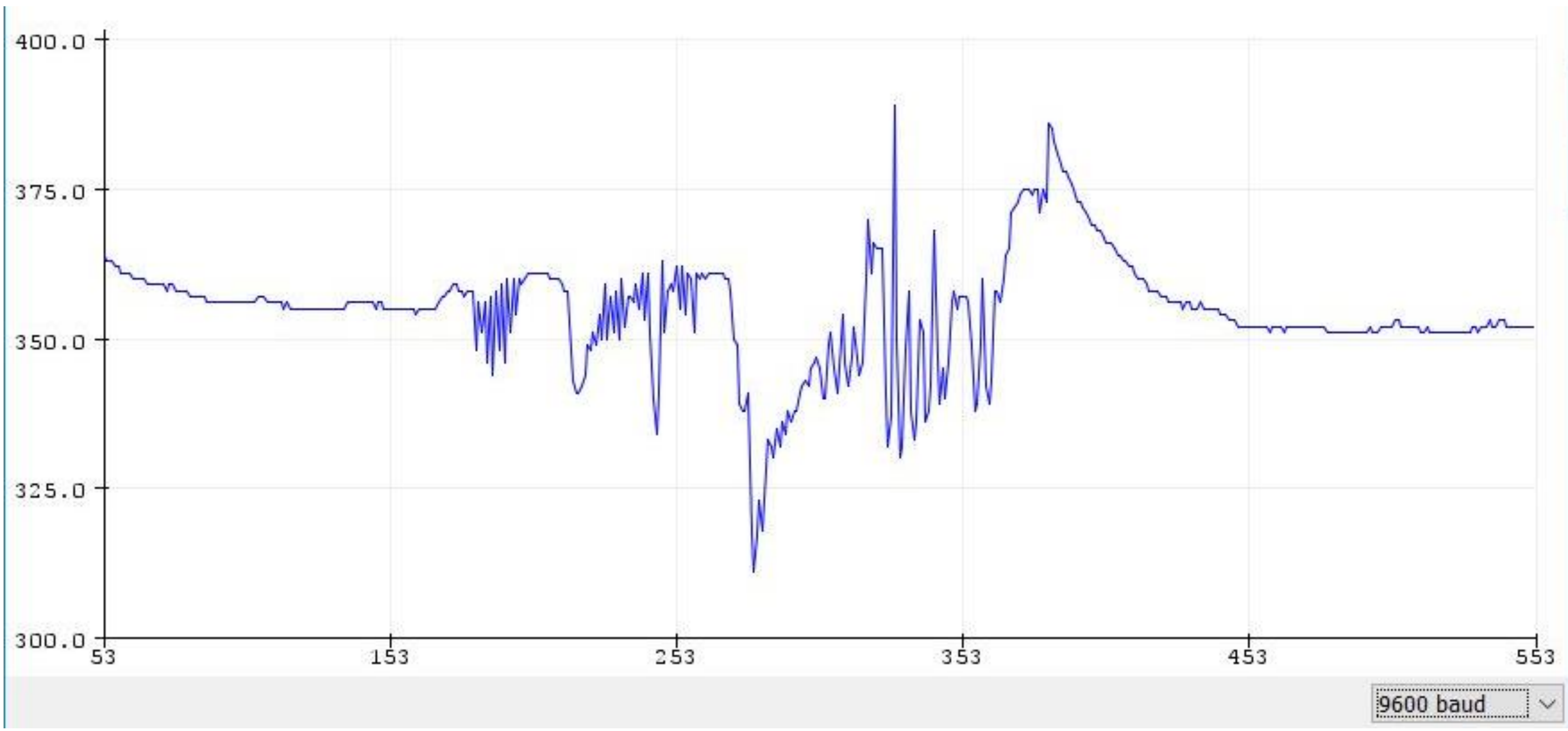
- Able to model Concentration ***at a particular time,***
- or Dosage ***at end of time,***
- but ***NOT*** Concentration ***shape over time***
- Profs. Jeff Wu & Roshan Joseph from Georgia Tech ISyE suggested using Functional Data Analysis

Examples of Functional Data

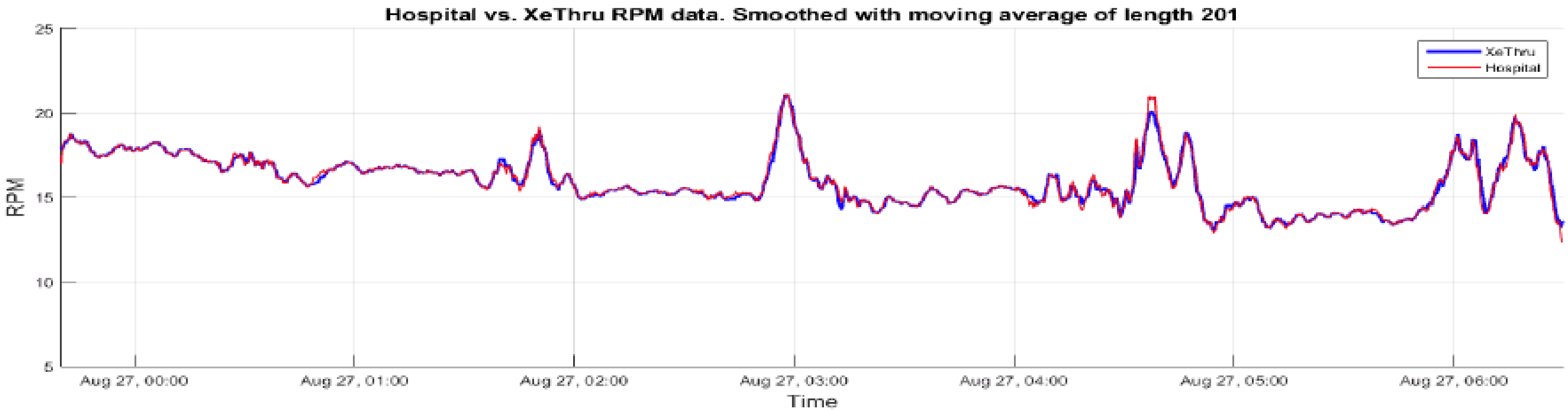
- Sensor streams
- Measurements taken over a range
- Vibration signals
- Spectral data
- Tool wear
- Gun barrel degradation
- Radar/sonar signatures
- Trajectories of flights between cities
- Tracking of surgeon hand movement
- Electrocardiograms (EKGs)

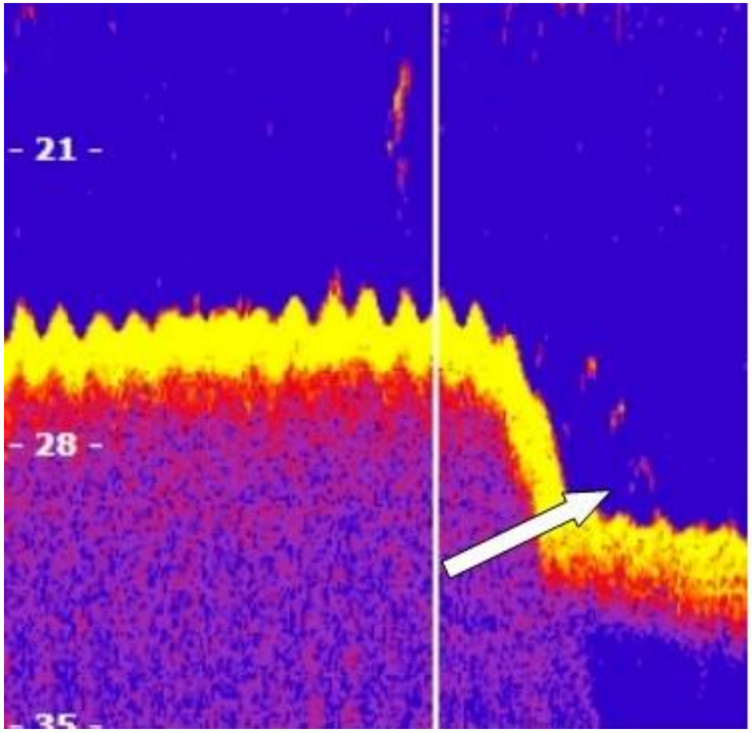
- *Almost any response in a longitudinal order*

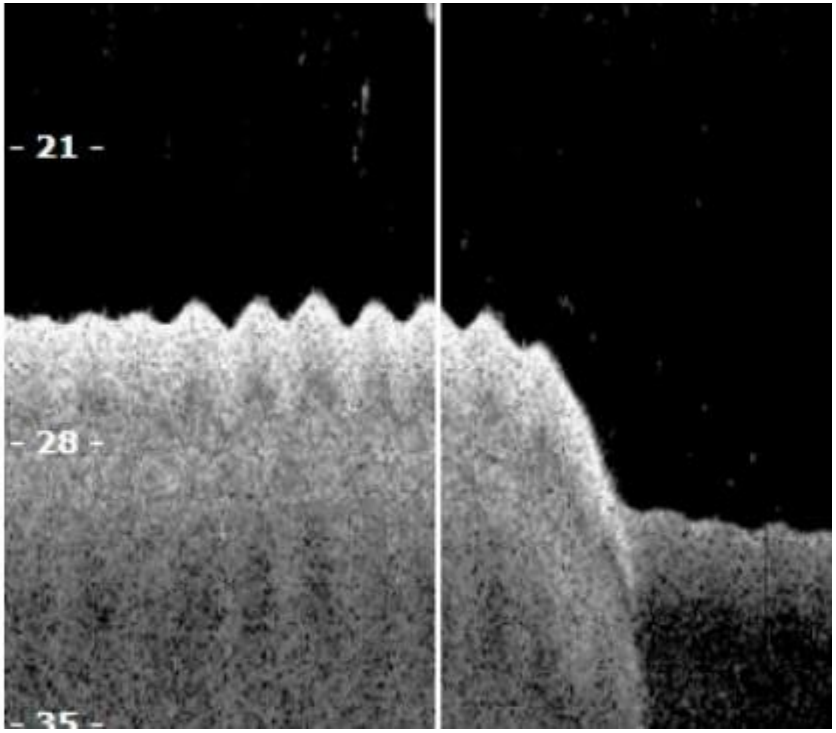
Vibration Sensor



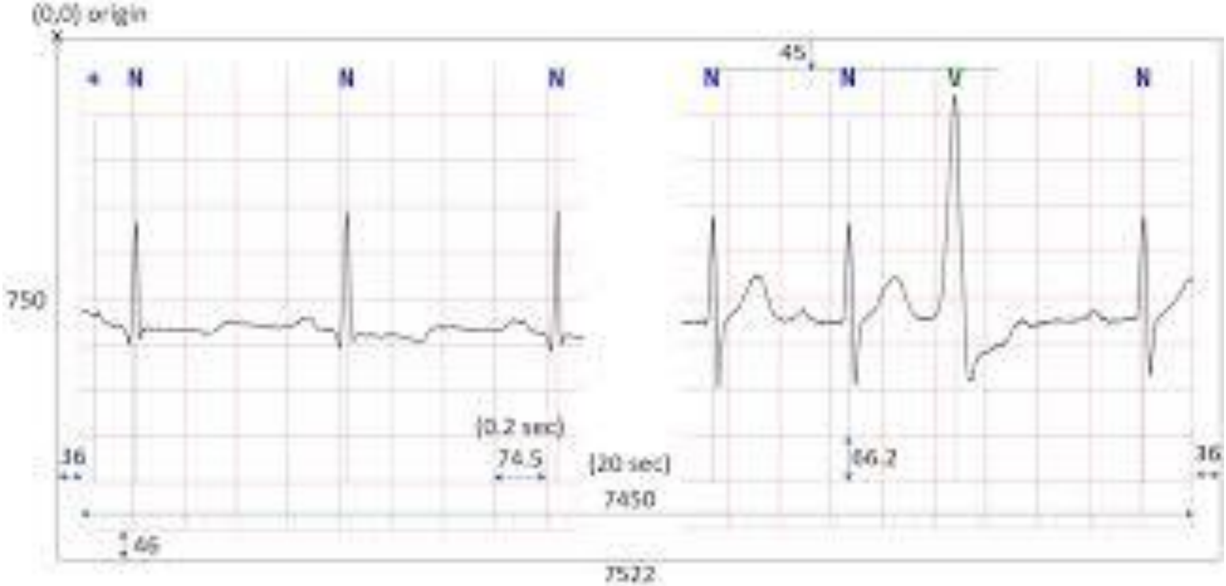
Radar and Sonar Data

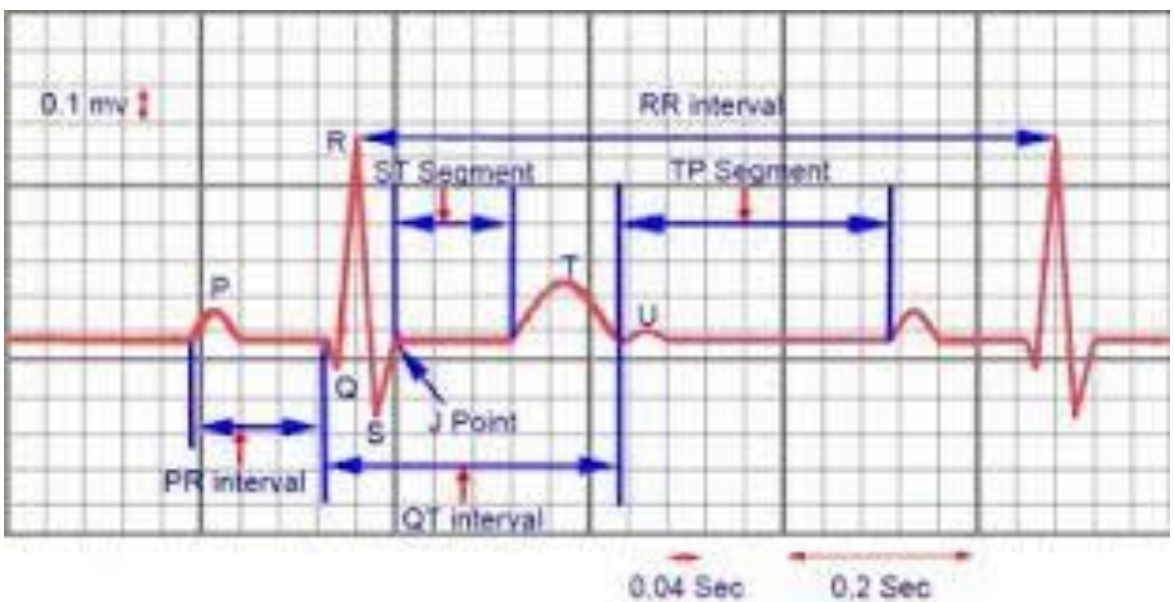






Electrocardiograms





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.

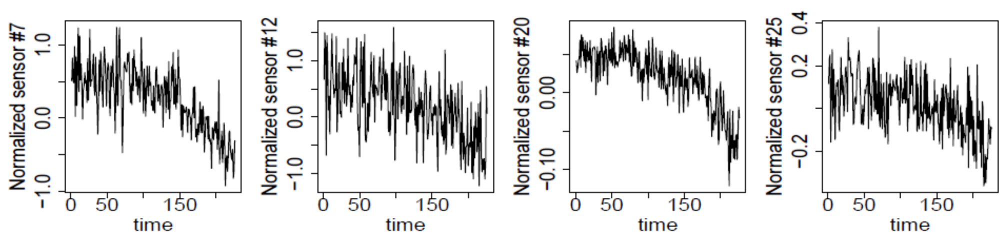


Fig. 3: Removing the effect of operating conditions on sensor data

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

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

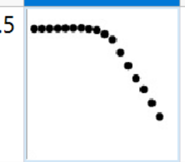
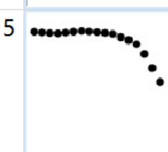
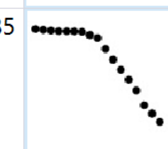
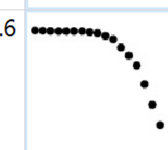
What is Functional Data Analysis?

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.

Analysis Method Overview: Data Landmarks

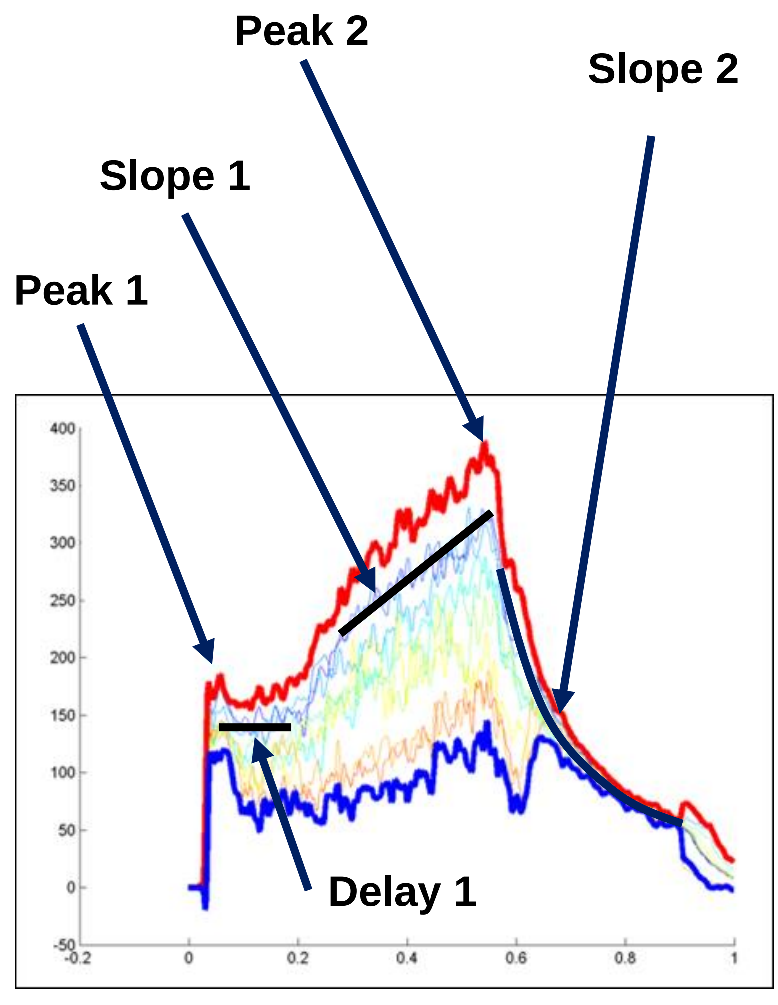
Curve was split into sections and key points and slopes were used as separate results

Standard statistical methods compared each landmark value

Landmarks from new tests were compared to previous runs

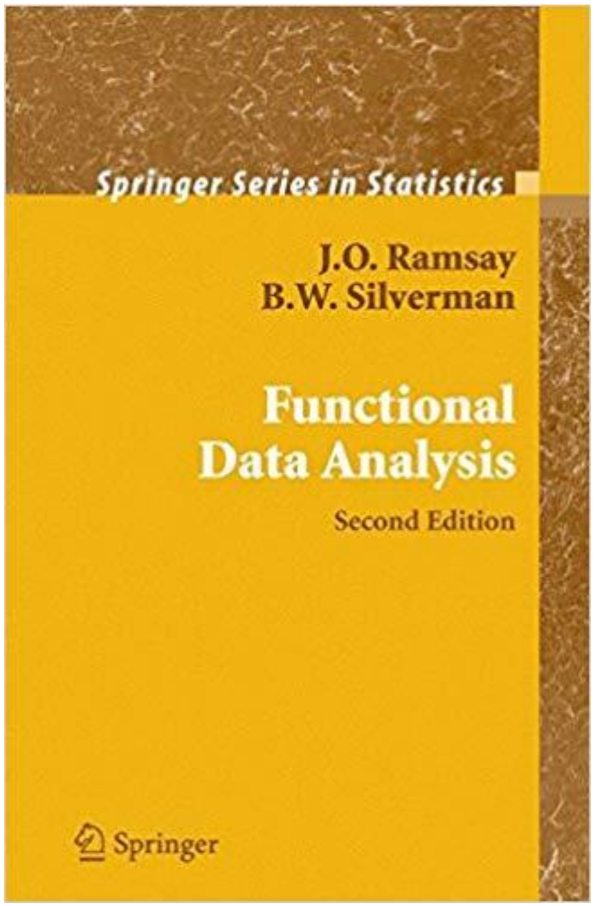
Most effective non-FDA option

Must perform statistical analysis on each landmark

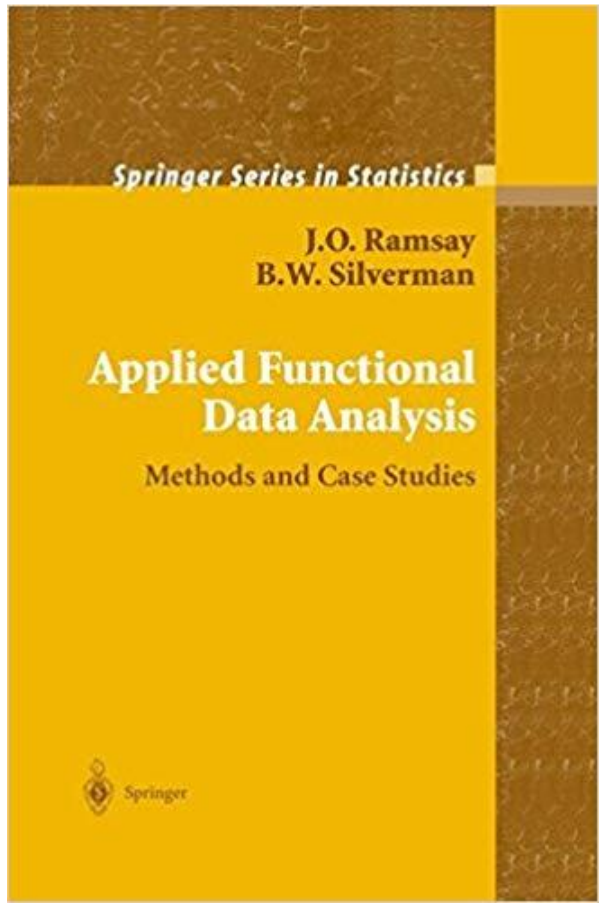


Based on slide from David Harrison of Lockheed Martin Corporation

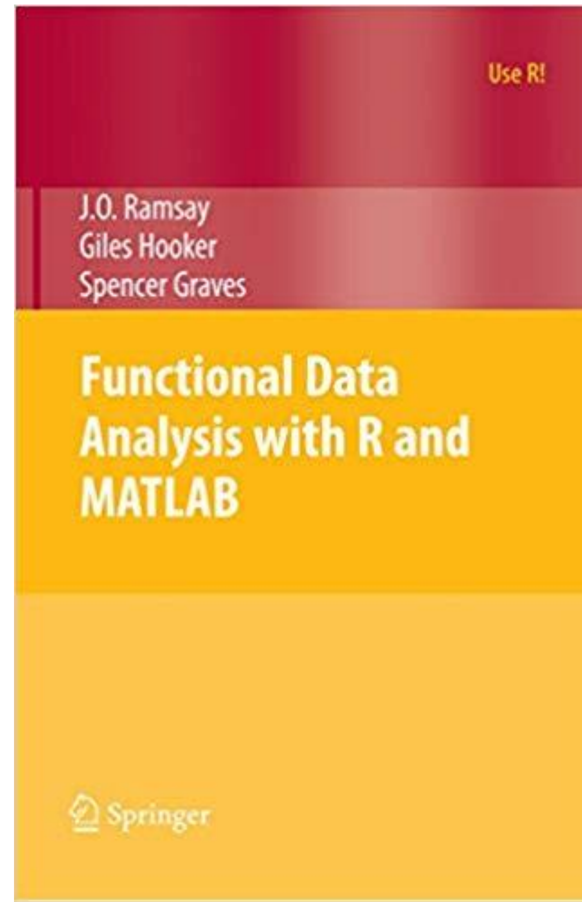
Functional Data Analysis seminal work by James O. Ramsay and Bernard W. Silverman



2005
(1e 1997)



2005



2009

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

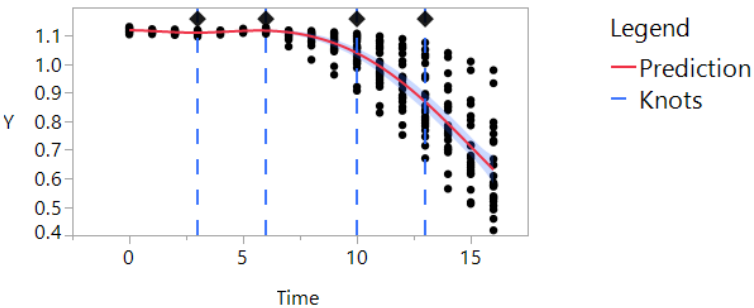
Functional Data Analysis

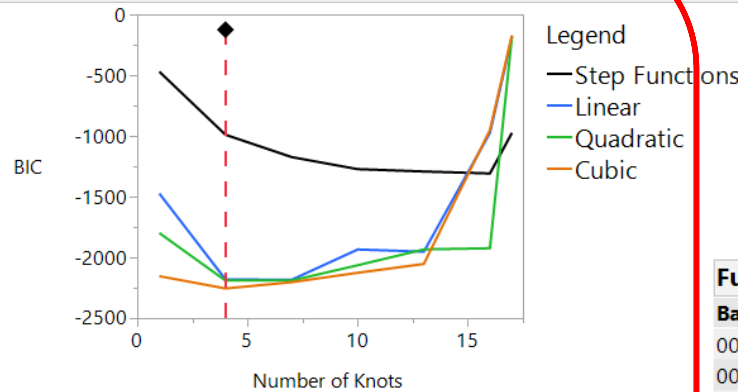
- F-DOE & F-ML use functional principal components analysis (F-PCA)
- F-PCA breaks the data into ***FPC Scores*** and ***Eigenfunctions*** in a dimension reduction that is closely analogous to classical PCA
- FPC Scores are scalars that explain ***function-to-function variation***
- Eigenfunctions explain 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

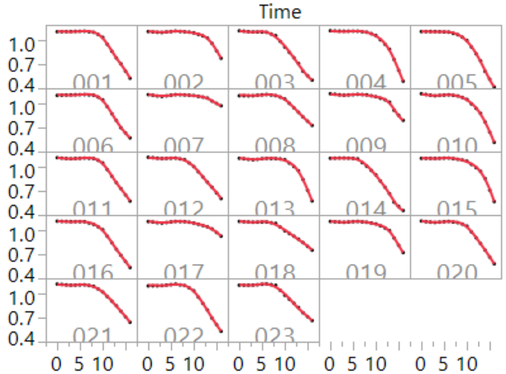
How do we analyze Functional data?

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





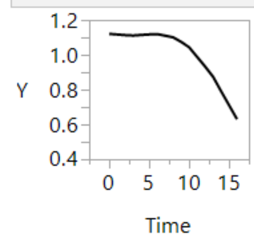


Summaries

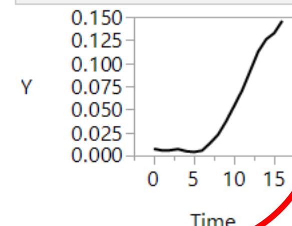
Overall

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

Mean



Standard Deviation

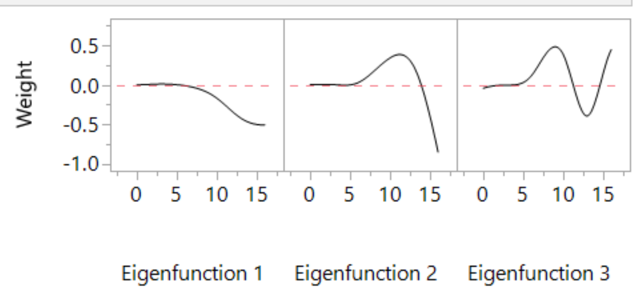


Function Summaries

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%

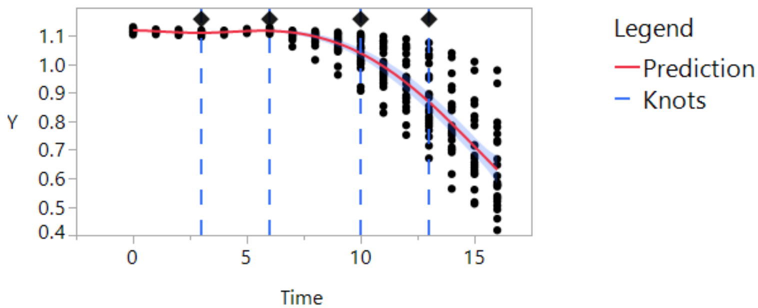


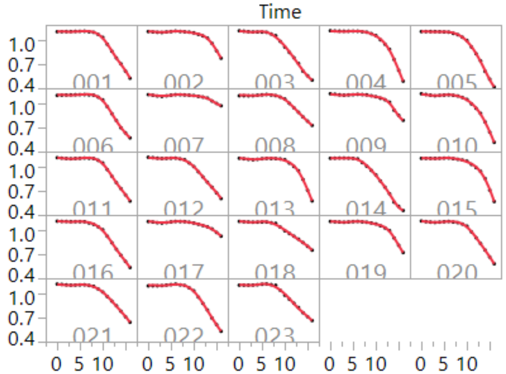
jmp Statistical Discovery. From SAS.

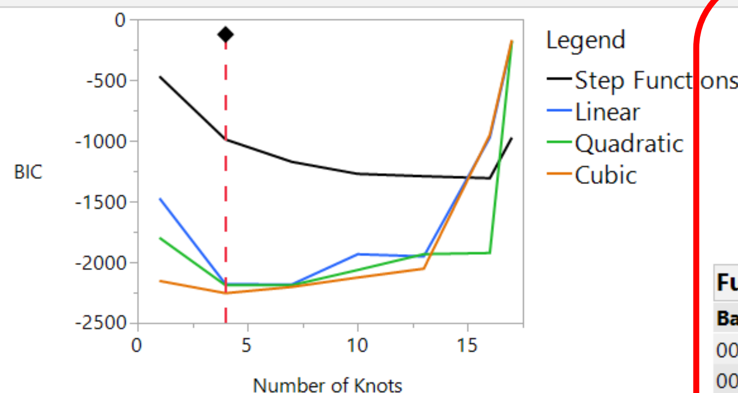
How do we analyze Functional data?

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





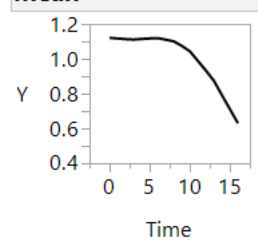


Summaries

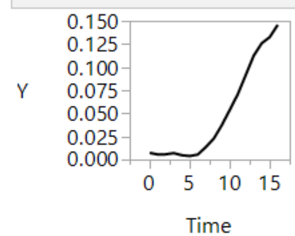
Overall

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

Mean

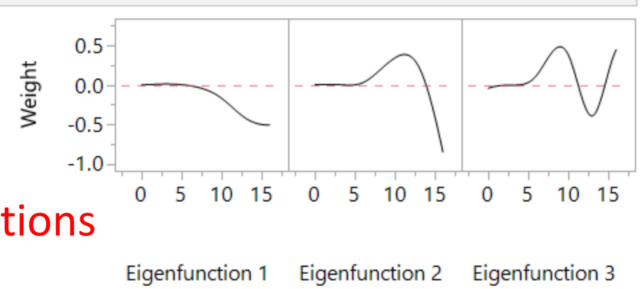


Standard Deviation



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

Functional Principal Component Scores

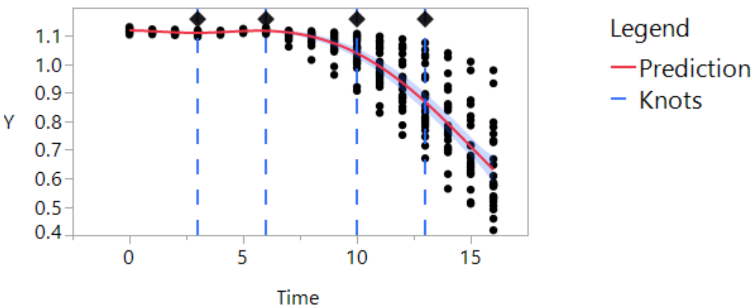
Function Summaries

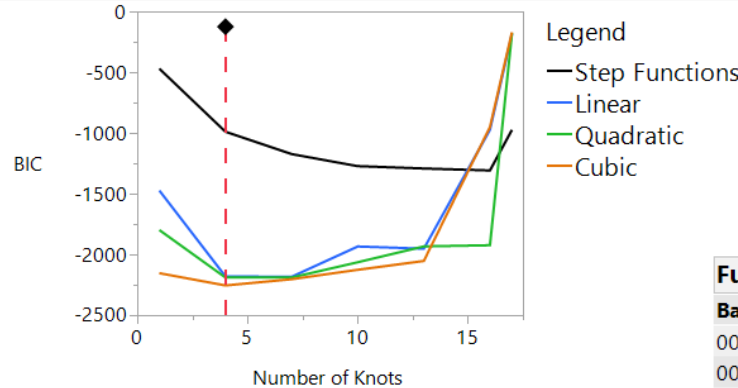
Batch	FPC 1	FPC 2	FPC 3
001	0.1762753	0.0249696	0.0391766
002	-0.353504	0.0180562	-0.019895
003	0.3748338	-0.091577	-0.032289
004	0.0587479	0.1194643	-0.022659
005	0.3493758	0.0517594	-0.011833
006	0.1462727	0.0217465	0.056155
007	-0.534657	-0.078567	0.0063818
008	-0.11289	-0.033664	0.0282277
009	-0.316959	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

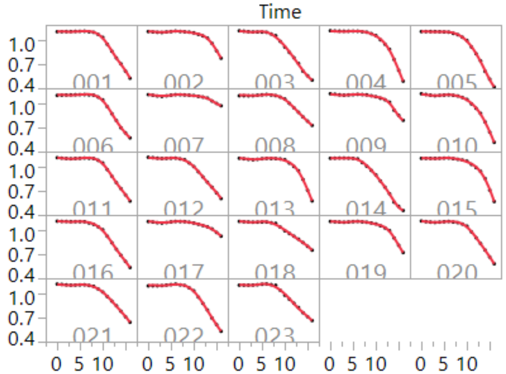
How do we analyze Functional data?

3. Eigenfunctions explain the longitudinal variation.
4. Function Summaries (FPC scores) explain function-to-function variation.

Model Selection







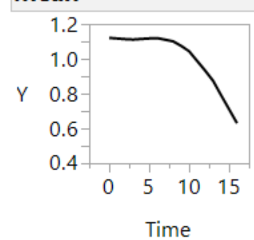
Explains function-to-function variation

Summaries

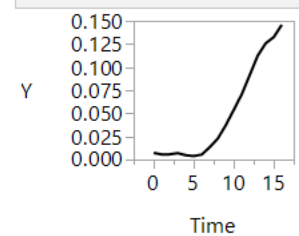
Overall

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

Mean



Standard Deviation

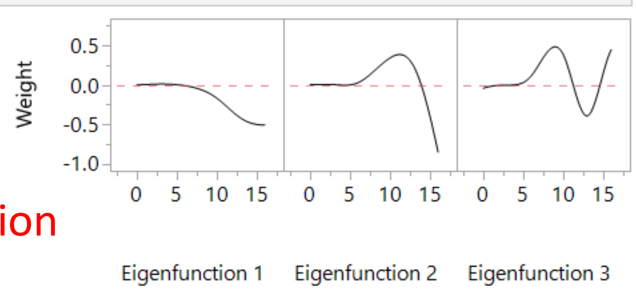


Function Summaries

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%

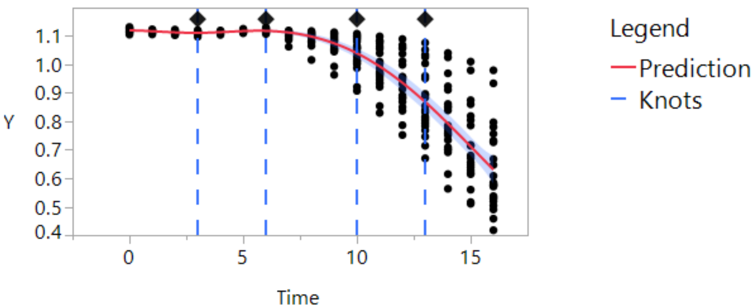


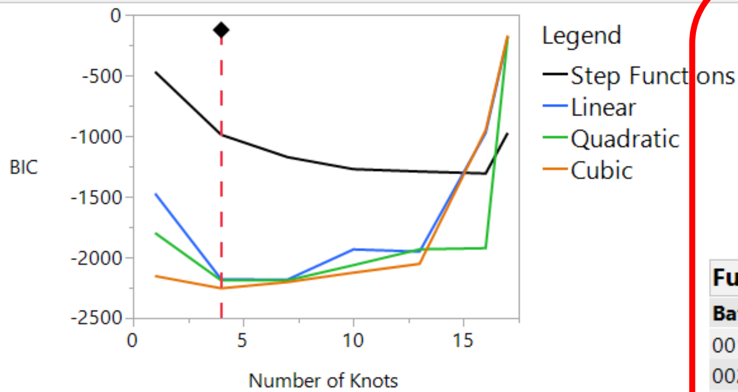
Explains longitudinal variation

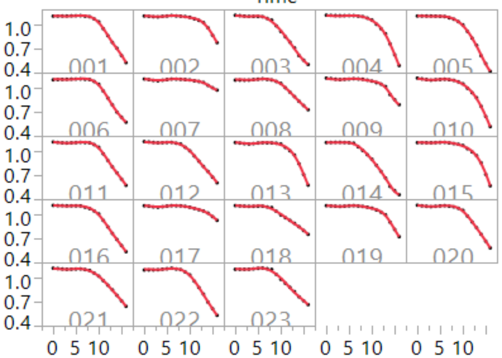
How do we analyze Functional data?

3. Eigenfunctions explain the longitudinal variation.
4. Function Summaries (FPC scores) explain function-to-function variation.

Model Selection





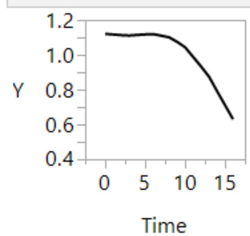


Summaries

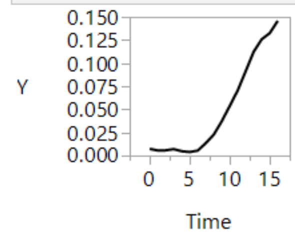
Overall

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

Mean

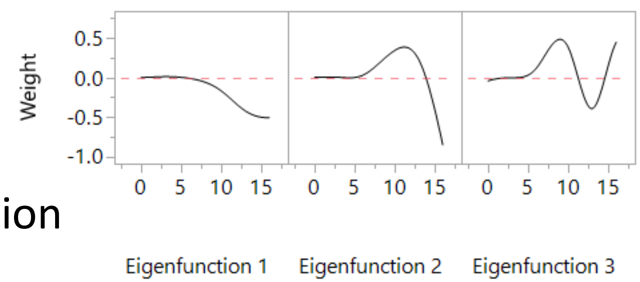


Standard Deviation



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

Explains function-to-function variation

Function Summaries

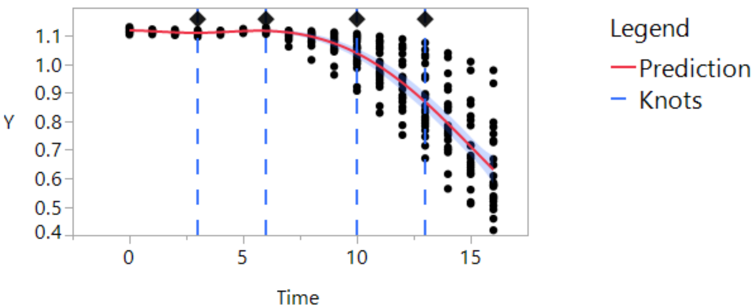
Batch	FPC 1	FPC 2	FPC 3
001	0.1762753	0.0249696	0.0391766
002	-0.353504	0.0180562	-0.019895
003	0.3748338	-0.091577	-0.032289
004	0.0587479	0.1194643	-0.022659
005	0.3493758	0.0517594	-0.011833
006	0.1462727	0.0217465	0.056155
007	-0.534657	-0.078567	0.0063818
008	-0.11289	-0.033664	0.0282277
009	-0.316959	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

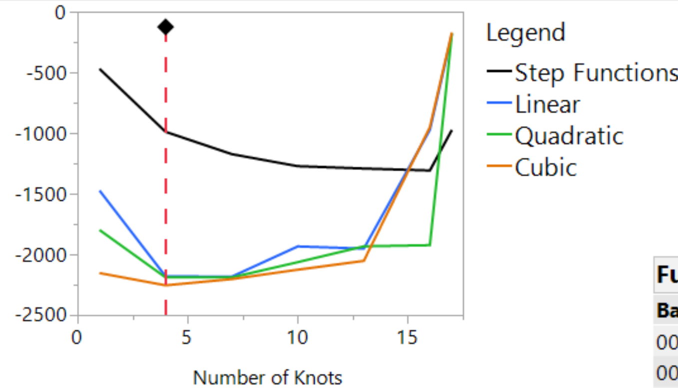
jmp Statistical Discovery.™ From SAS.

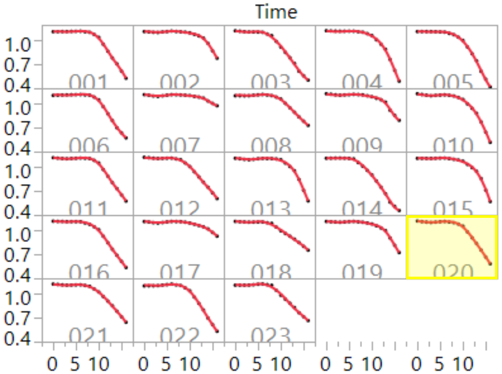
How do we analyze Functional data?

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

Model Selection





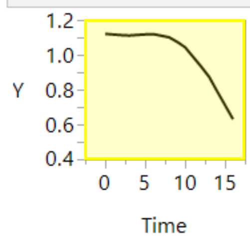


Summaries

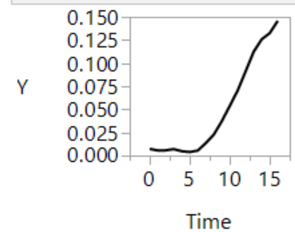
Overall

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

Mean



Standard Deviation



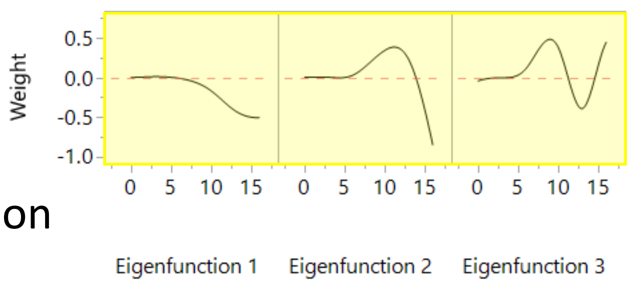
Explains function-to-function variation

Function Summaries

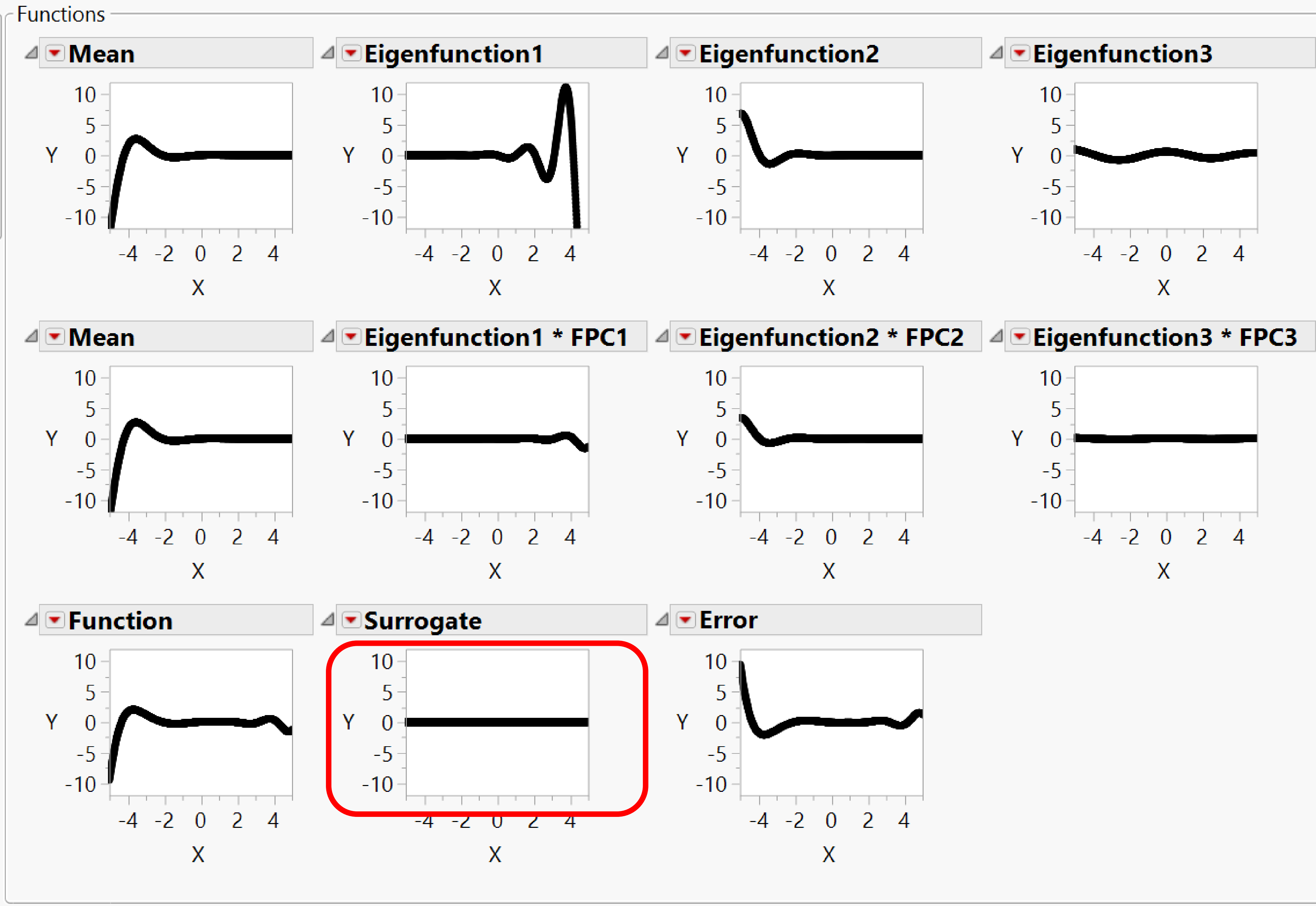
Batch	FPC 1	FPC 2	FPC 3
001	0.1762753	0.0249696	0.0391766
002	-0.353504	0.0180562	-0.019895
003	0.3748338	-0.091577	-0.032289
004	0.0587479	0.1194643	-0.022659
005	0.3493758	0.0517594	-0.011833
006	0.1462727	0.0217465	0.056155
007	-0.534657	-0.078567	0.0063818
008	-0.11289	-0.033664	0.0282277
009	-0.316959	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

Functional PCA

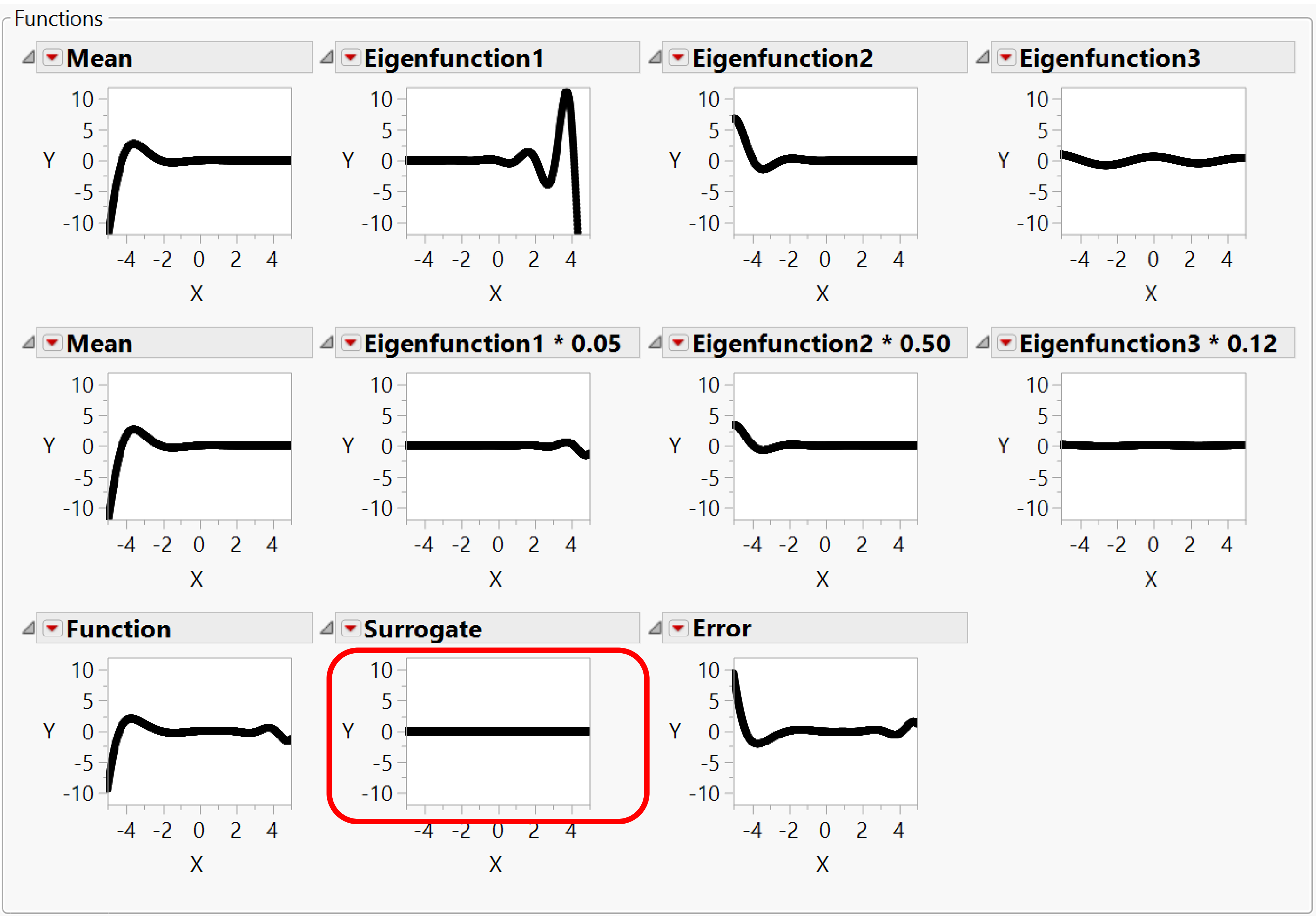
FPC	Eigenvalue	20 40 60 80	Percent	Cumulative
1	0.06855		92.2%	92.2%
2	0.00480		6.46%	98.7%
3	0.00087		1.17%	99.8%



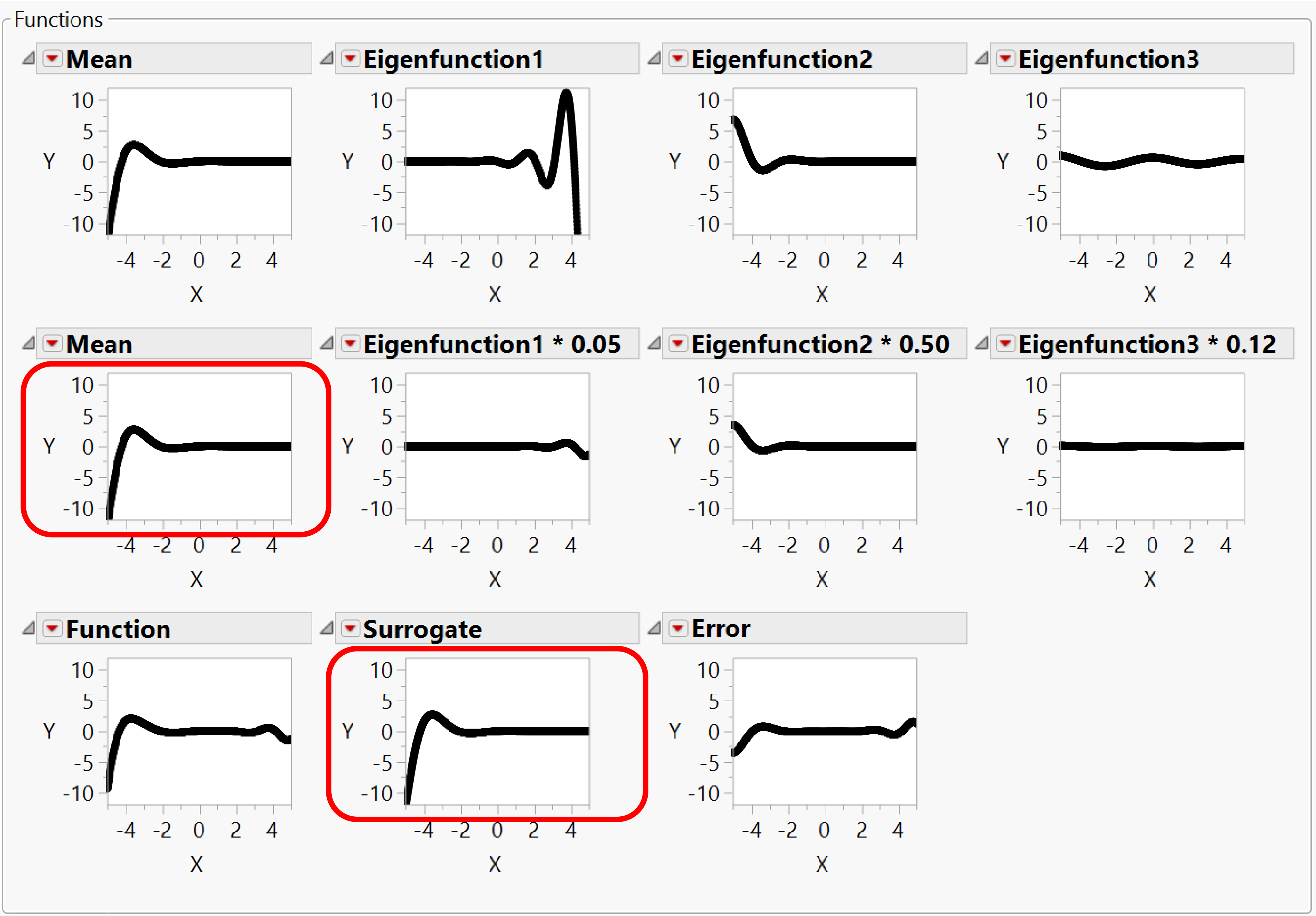
Explains longitudinal variation



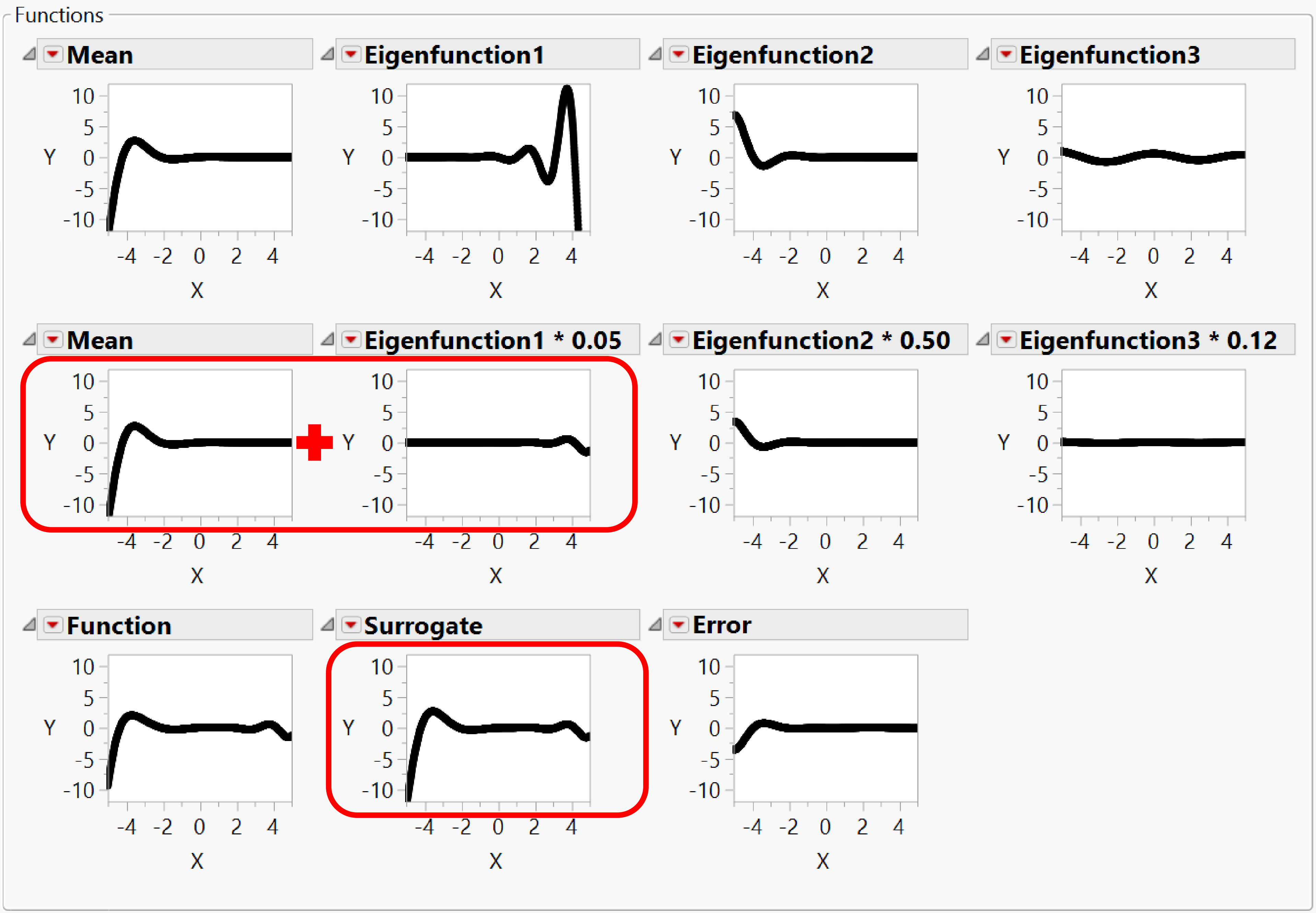
$$Y(X) = Surrogate$$



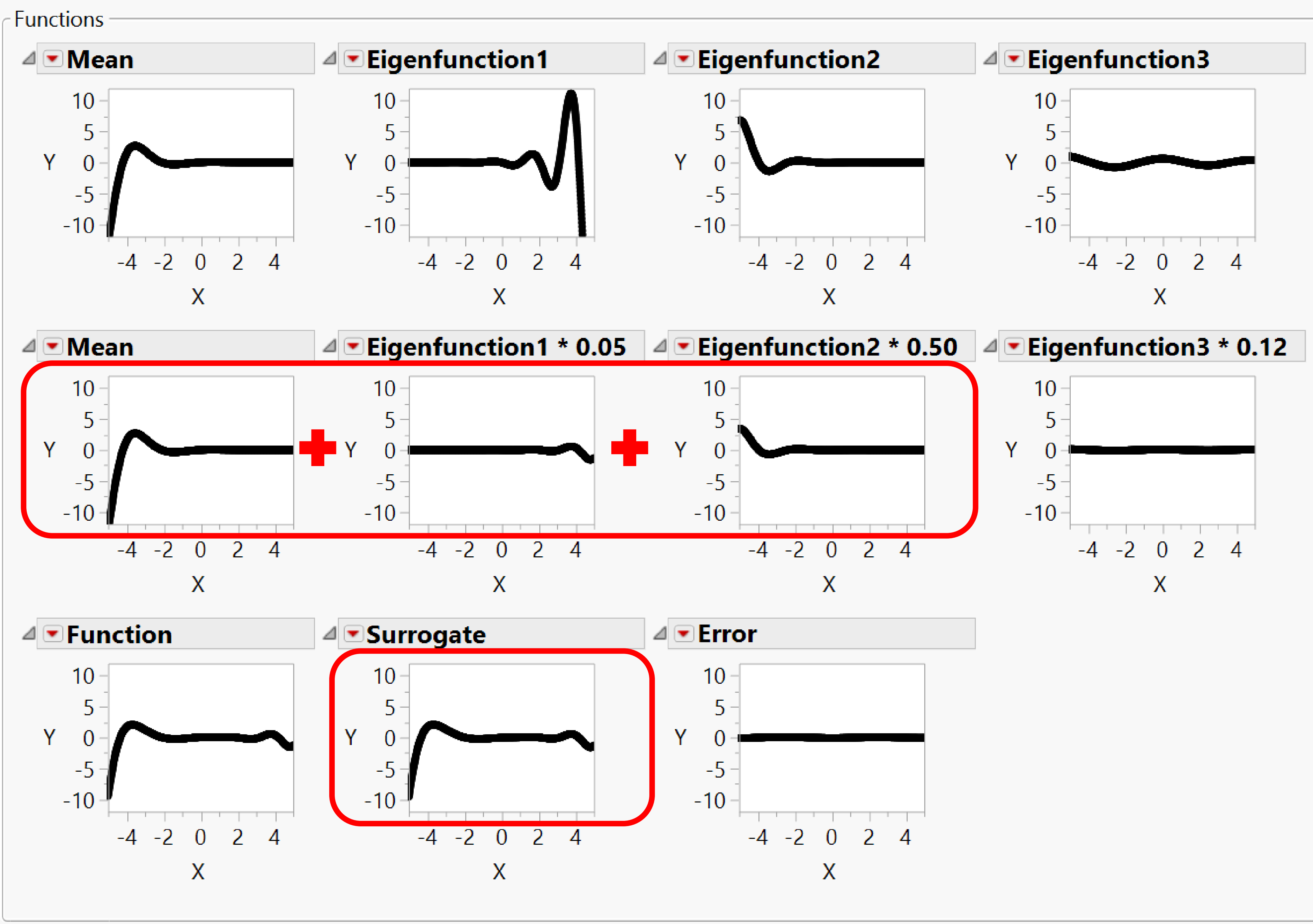
$$Y(X) = Surrogate$$


Functions
Mean
Eigenfunction1
Eigenfunction2
Eigenfunction3
Mean
Eigenfunction1 * 0.05
Eigenfunction2 * 0.50
Eigenfunction3 * 0.12
Function
Surrogate
Error
Y
X

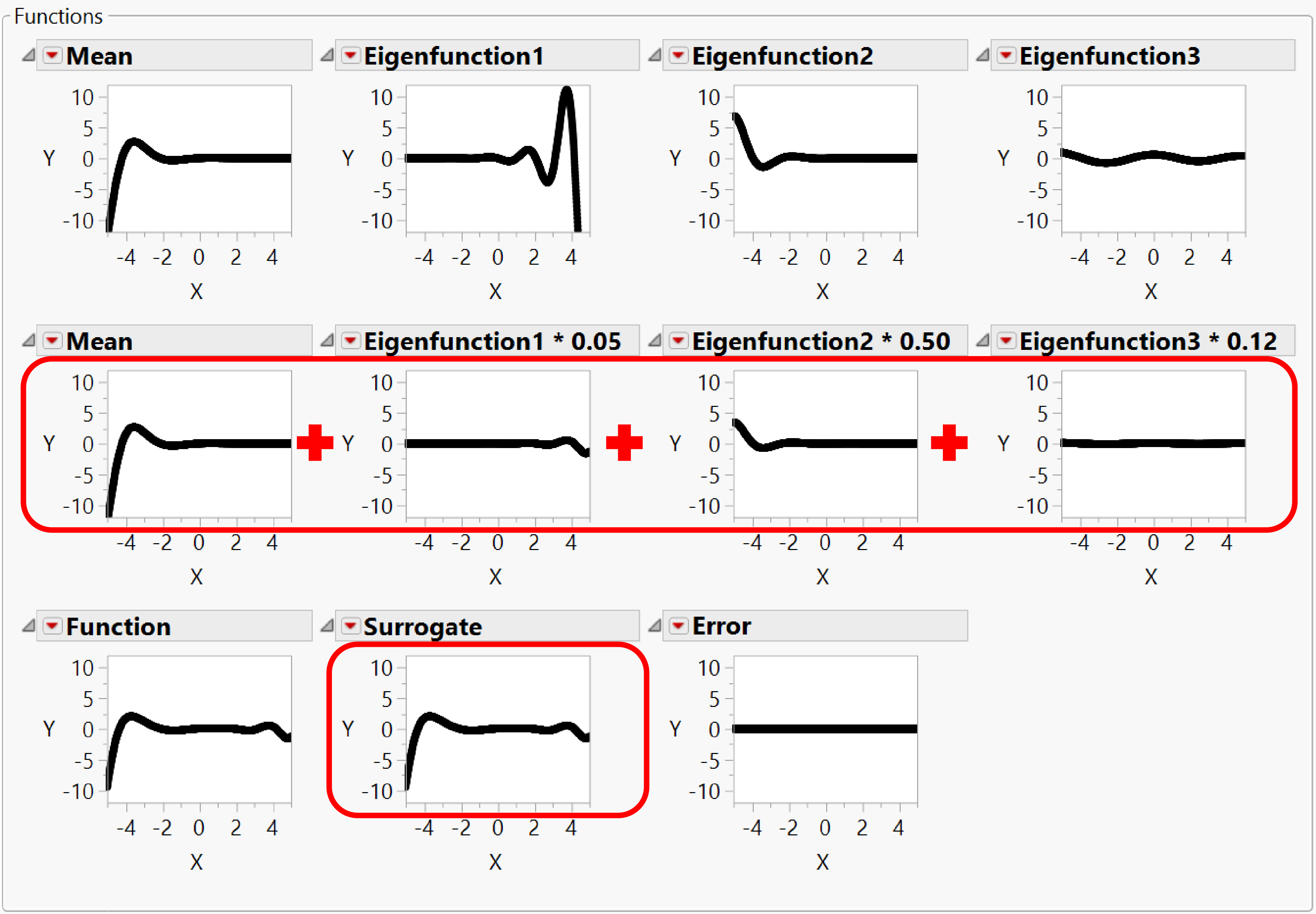
$$Y(X) = \mu(X)$$



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



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

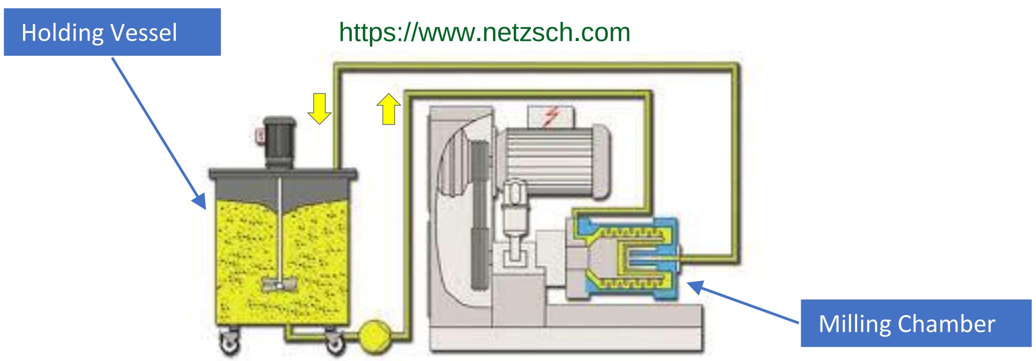


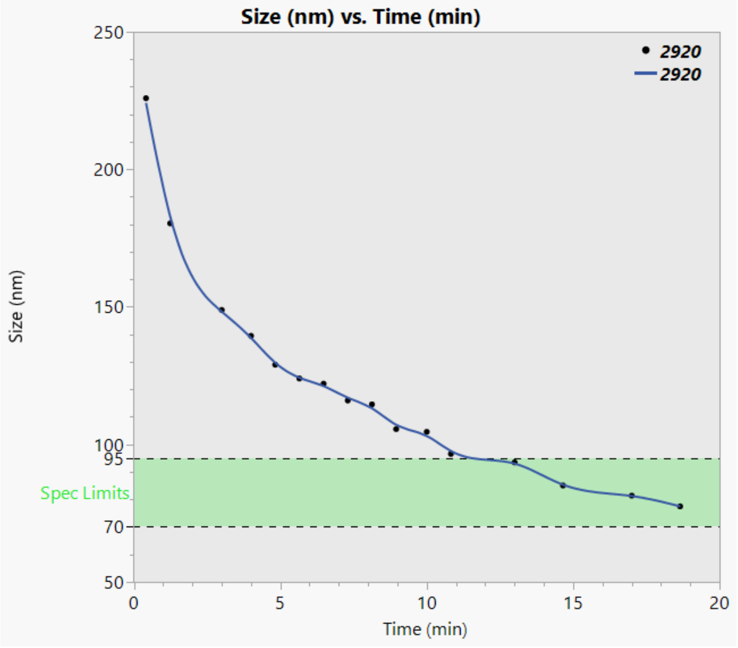
$$Y(X) = \mu(X) + 0.05 \cdot E_1(X) + 0.50 \cdot E_2(X) + 0.12 \cdot E_3(X)$$

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

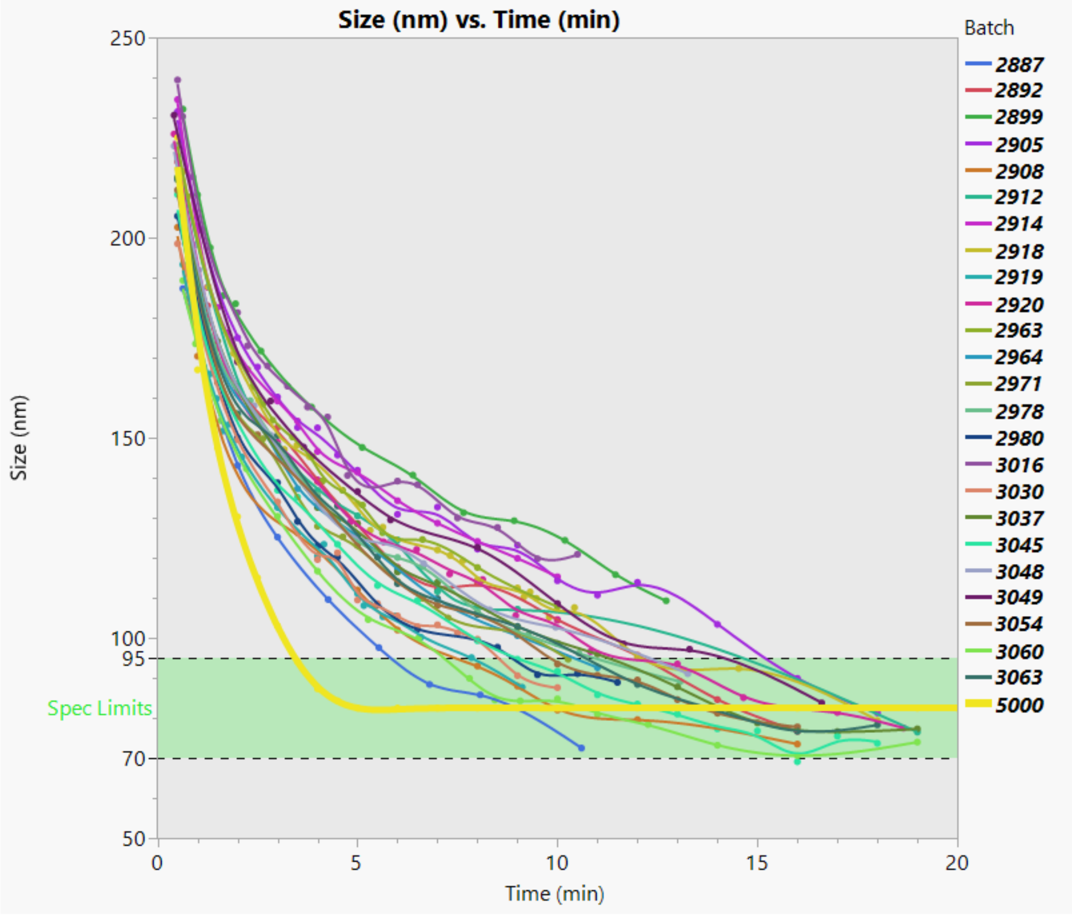




Example DoE Response

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

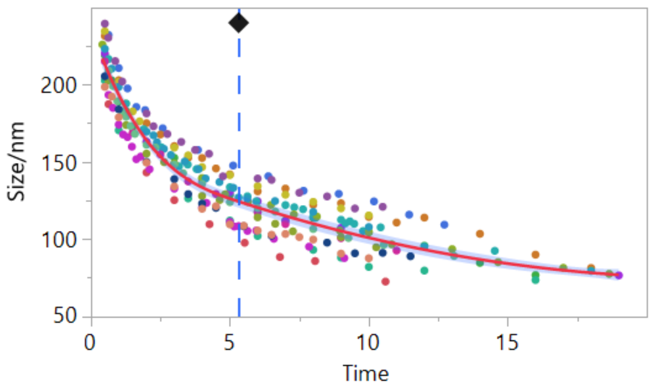


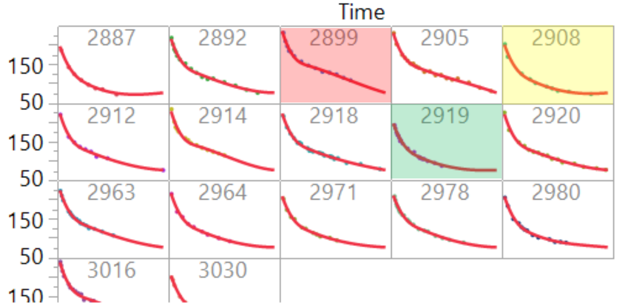
Single Eigenfunction and Associated FPC Scores for each Batch

B-Spline on Initial data

Model Controls

Model Selection





Functional PCA

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

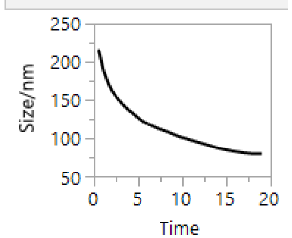
Function Summaries

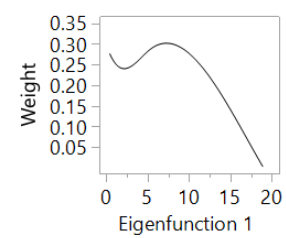
Batch of Mill DOE responses	FPC 1
2887	-85.9675
2892	2.2051353
2899	79.337968
2905	58.272695
2908	-60.35218
2912	7.7564941
2914	49.652658
2918	22.510929
2919	-56.34453
2920	9.1386688
2963	29.363581
2964	-13.79295
2971	-21.19822
2978	-6.822422
2980	-38.32591
3016	71.771035
3030	-47.20546

More Different: +79

Similar: -58 ± 2

Mean

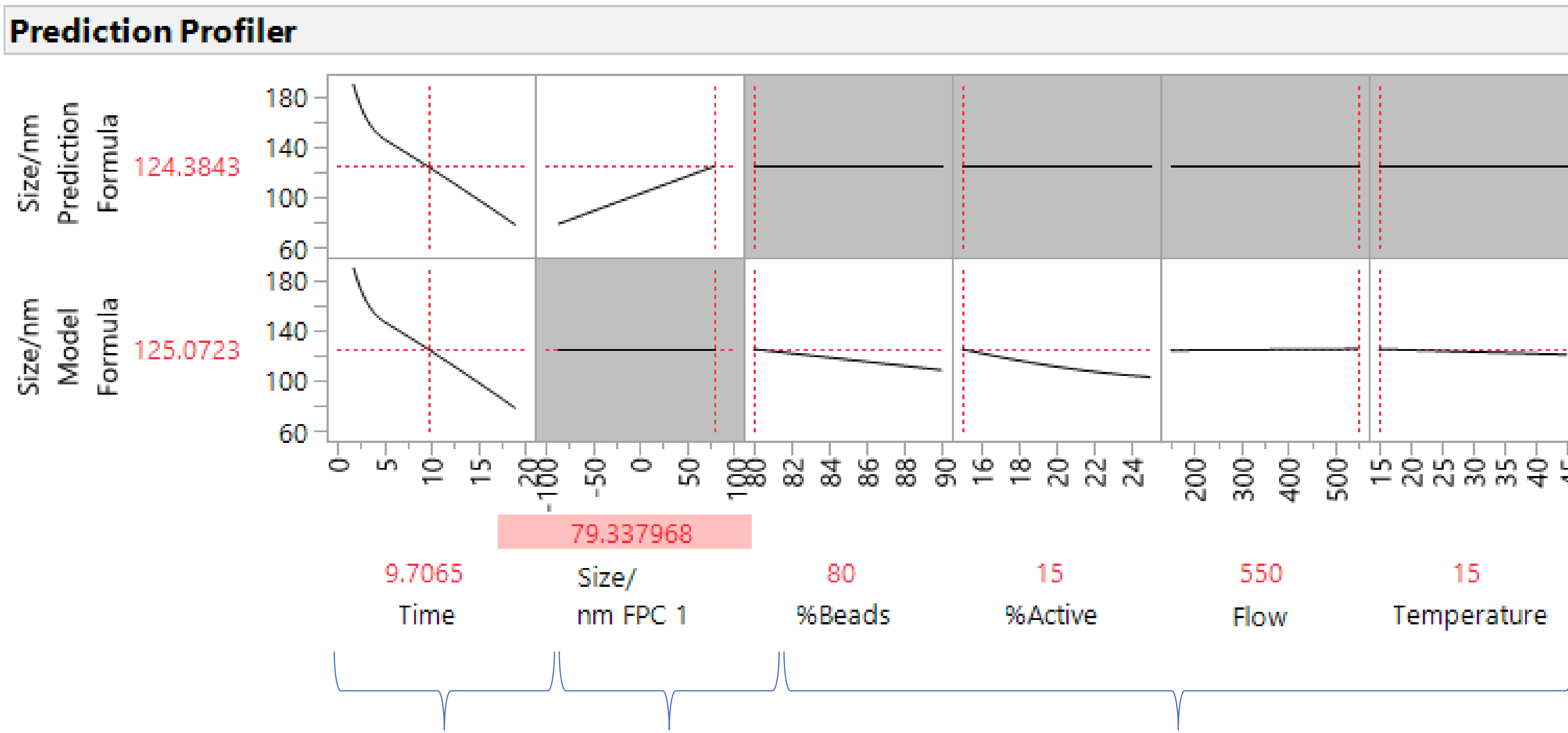




$$Y(X) = \mu(X) + \mathrm{FPC1} \cdot E_1(X)$$

Model the FPC Scores as functions of the DOE factors

Batch 2899



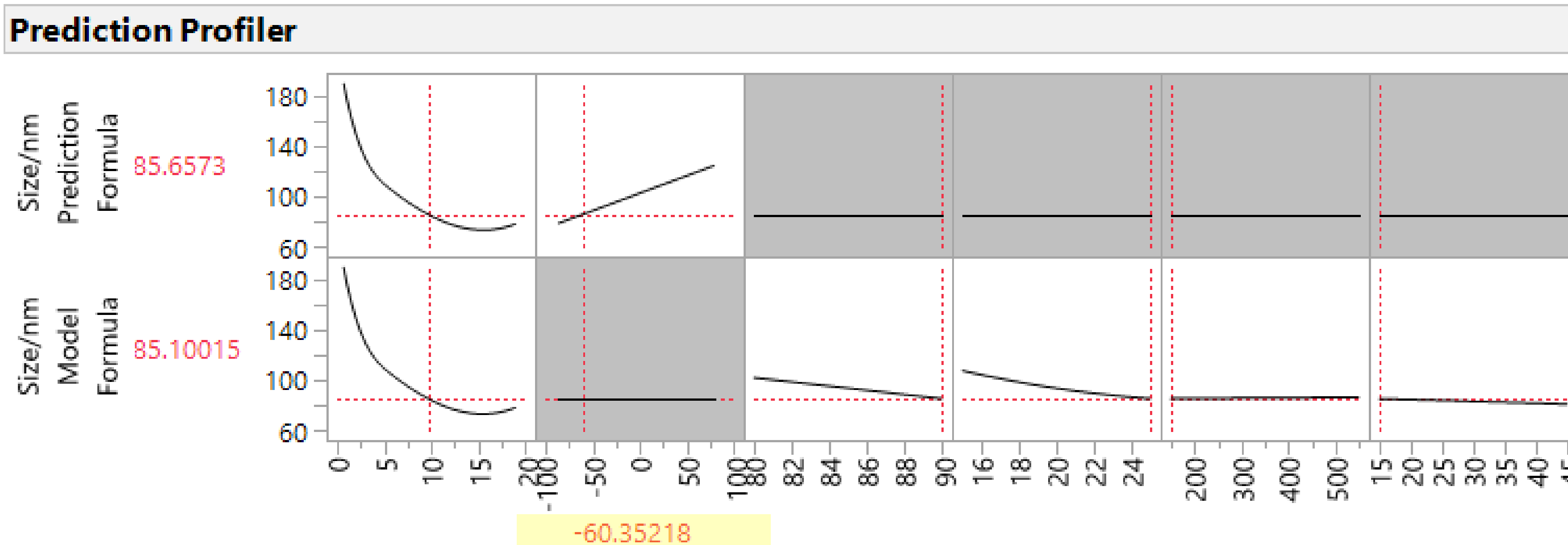
Functional Data ***Curve***

FPC Score

DoE Factors

Model the FPC Scores as functions of the DOE factors

Batch 2908



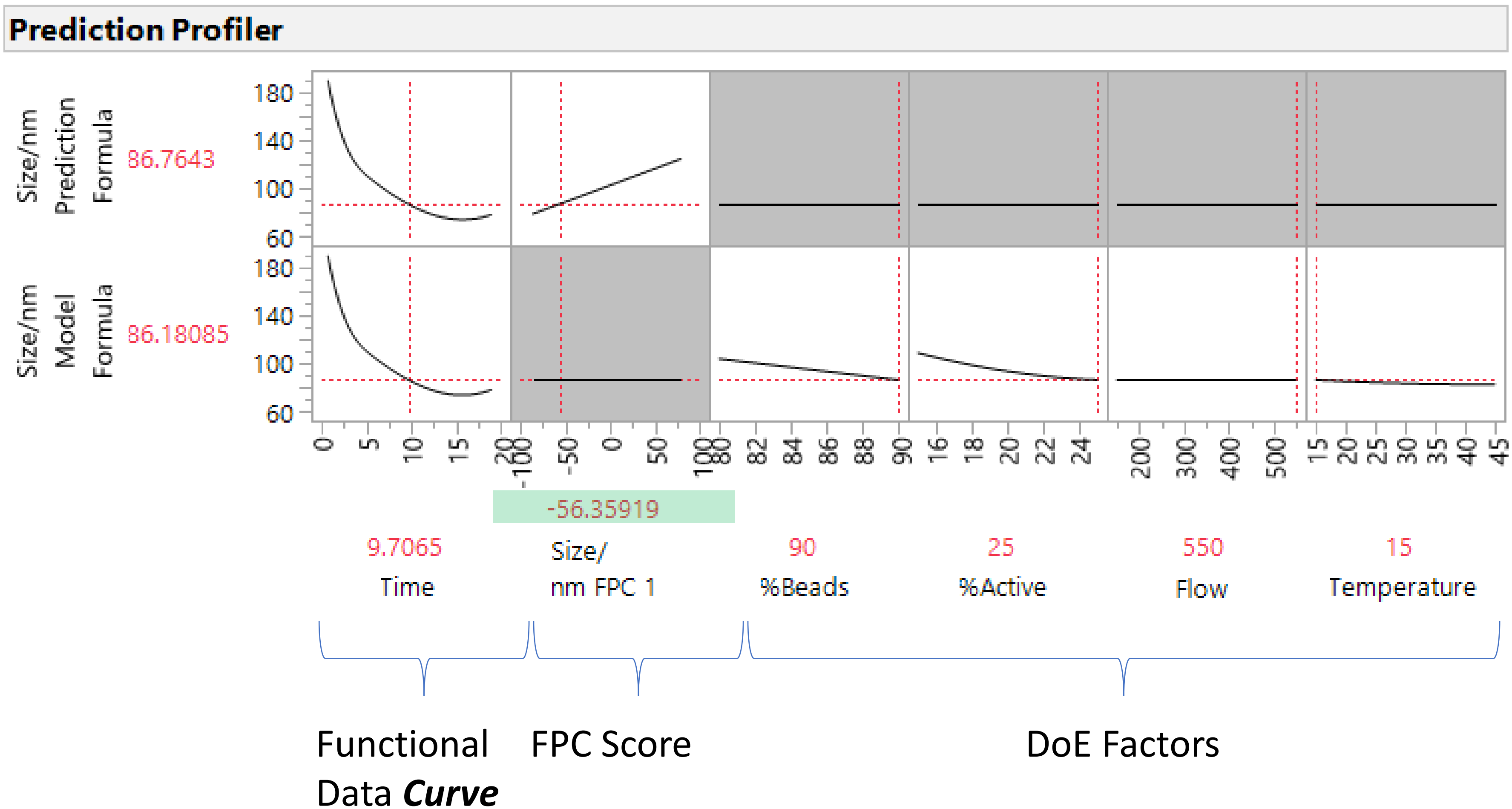
Functional Data ***Curve***

FPC Score

DoE Factors

Model the FPC Scores as functions of the DOE factors

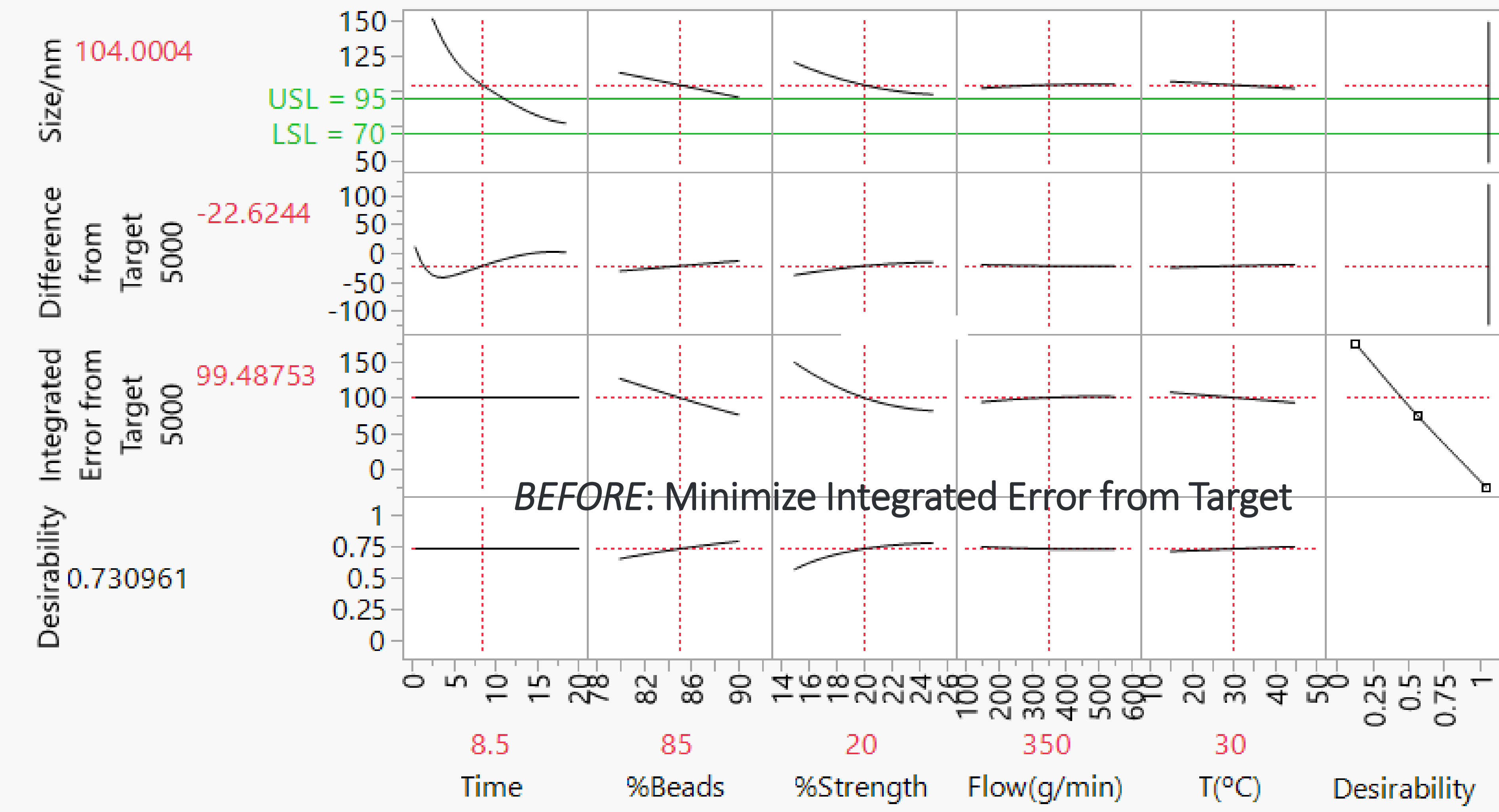
Batch 2919



jmp Statistical Discovery.™ From SAS.

FDOE Profiler

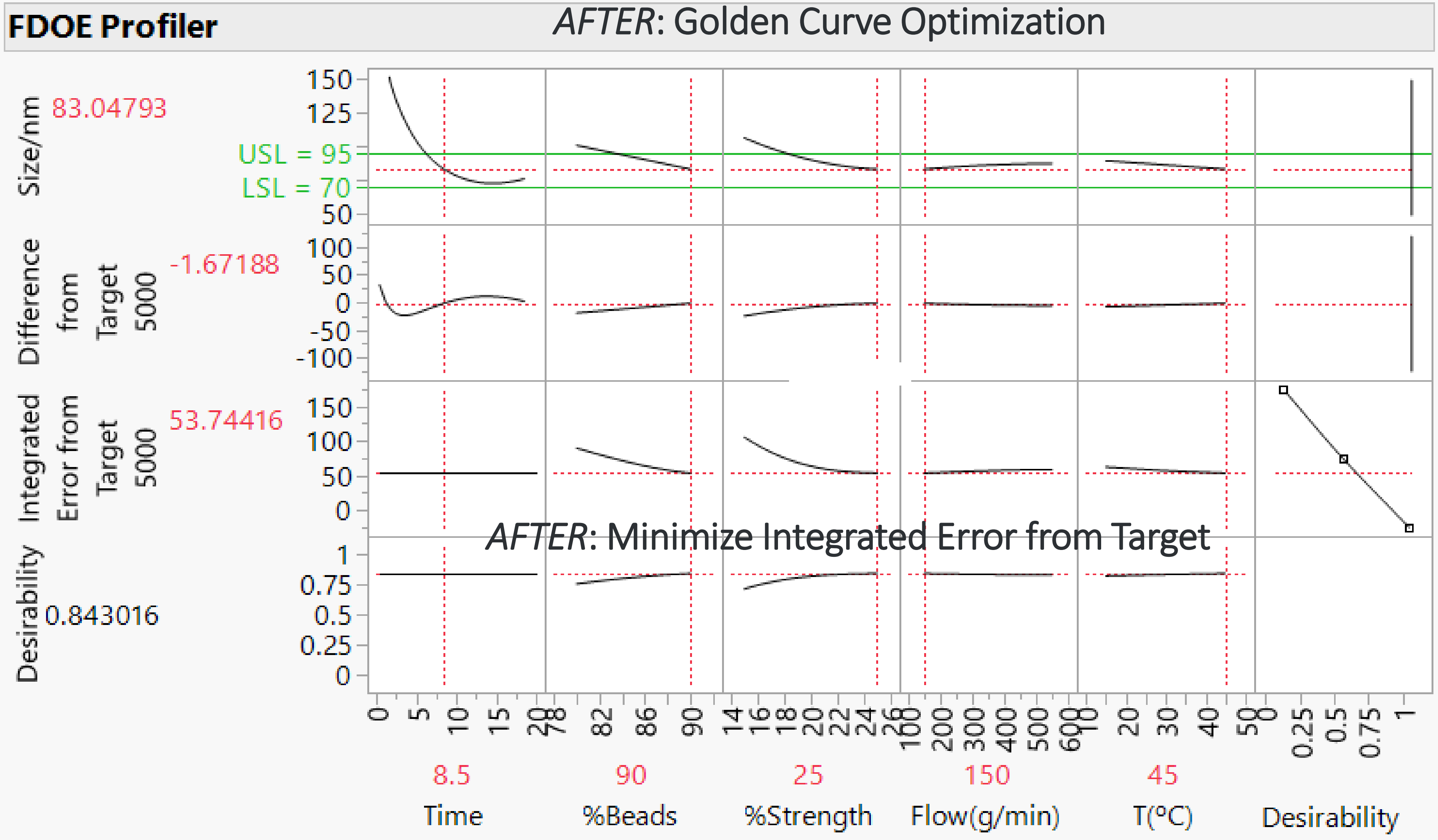
BEFORE: Golden Curve Optimization



Functional Response,
Difference from Target,
Integrated Error

DoE Factors

Desirability
(0=Bad, 1=Good)

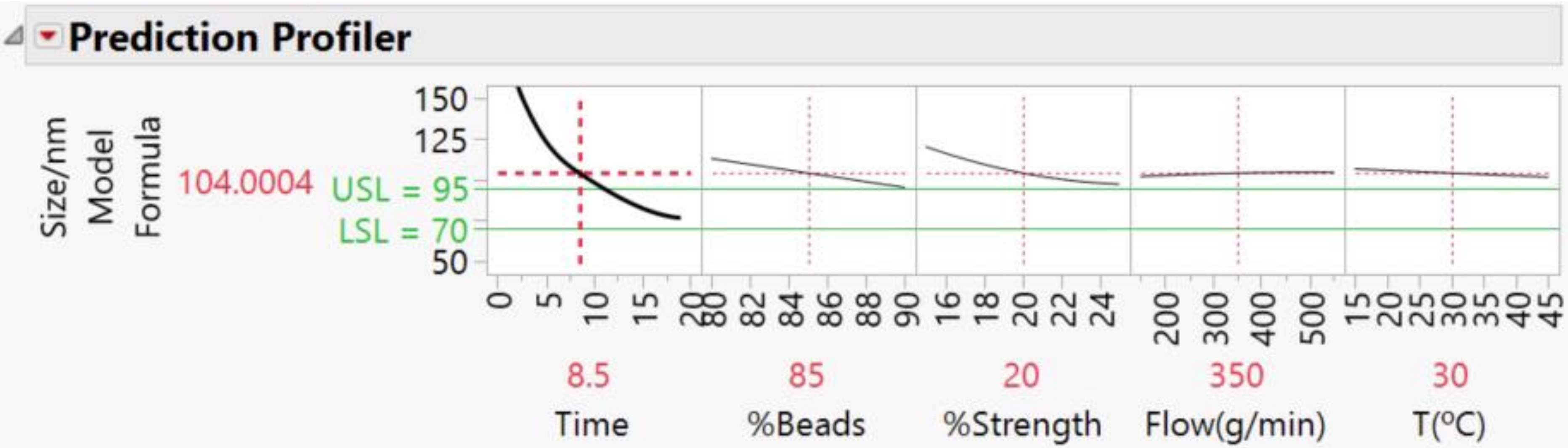


Functional Response,
Difference from Target,
Integrated Error

DoE Factors

Desirability
(0=Bad, 1=Good)

Final Prediction Model



Functional
Data ***Curve***

DoE Factors

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***

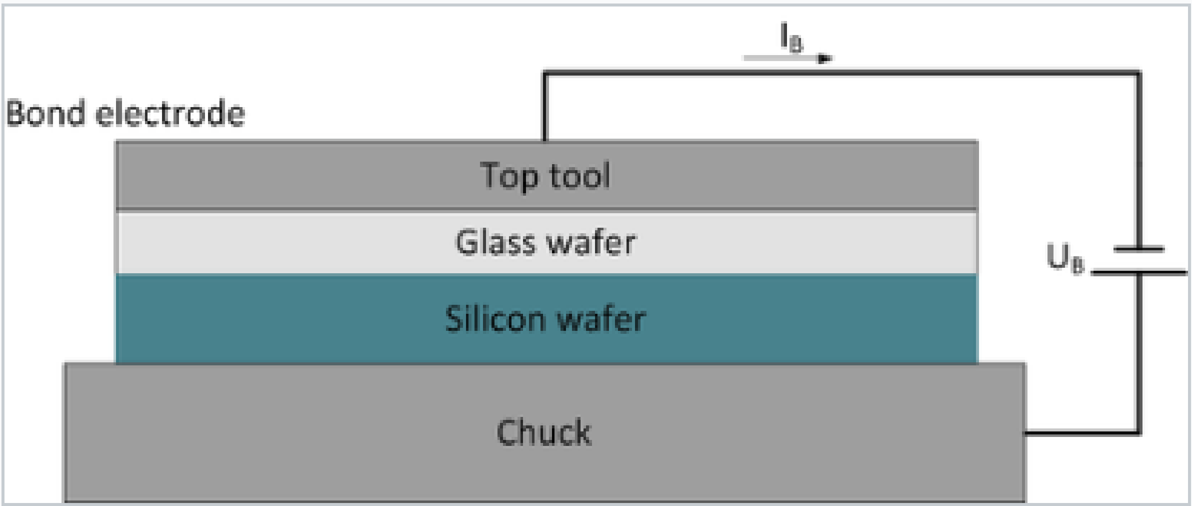
2. **Functional Response Machine Learning (F-ML)**: Goal is to use the functional data – *i.e. the **curve(s)*** – to predict something
   a) yield of a batch
   b) probability of detection / failure / hit

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



Scheme of anodic bonding procedure. The top tool works as a cathode and the chuck as anode.

Picture from *Wikipedia…*

Glass Bonded to Silicon Wafer
ISSUE: 12% of Wafers become Defective
BUT won't Know for Weeks which have Failed!

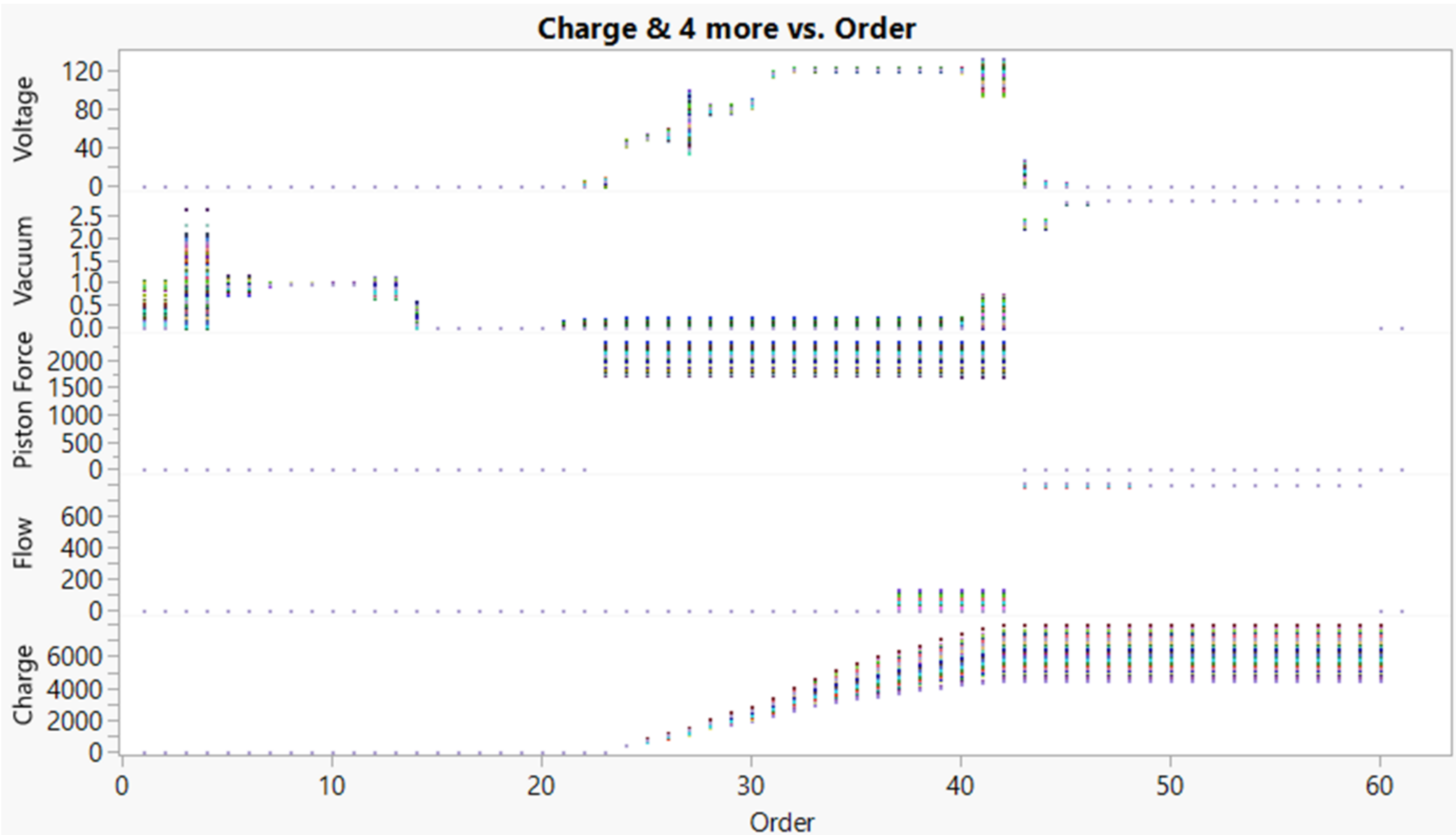
Anodic Bond Data:
2000 Wafers X 61 Time Steps = 122,000 Rows

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

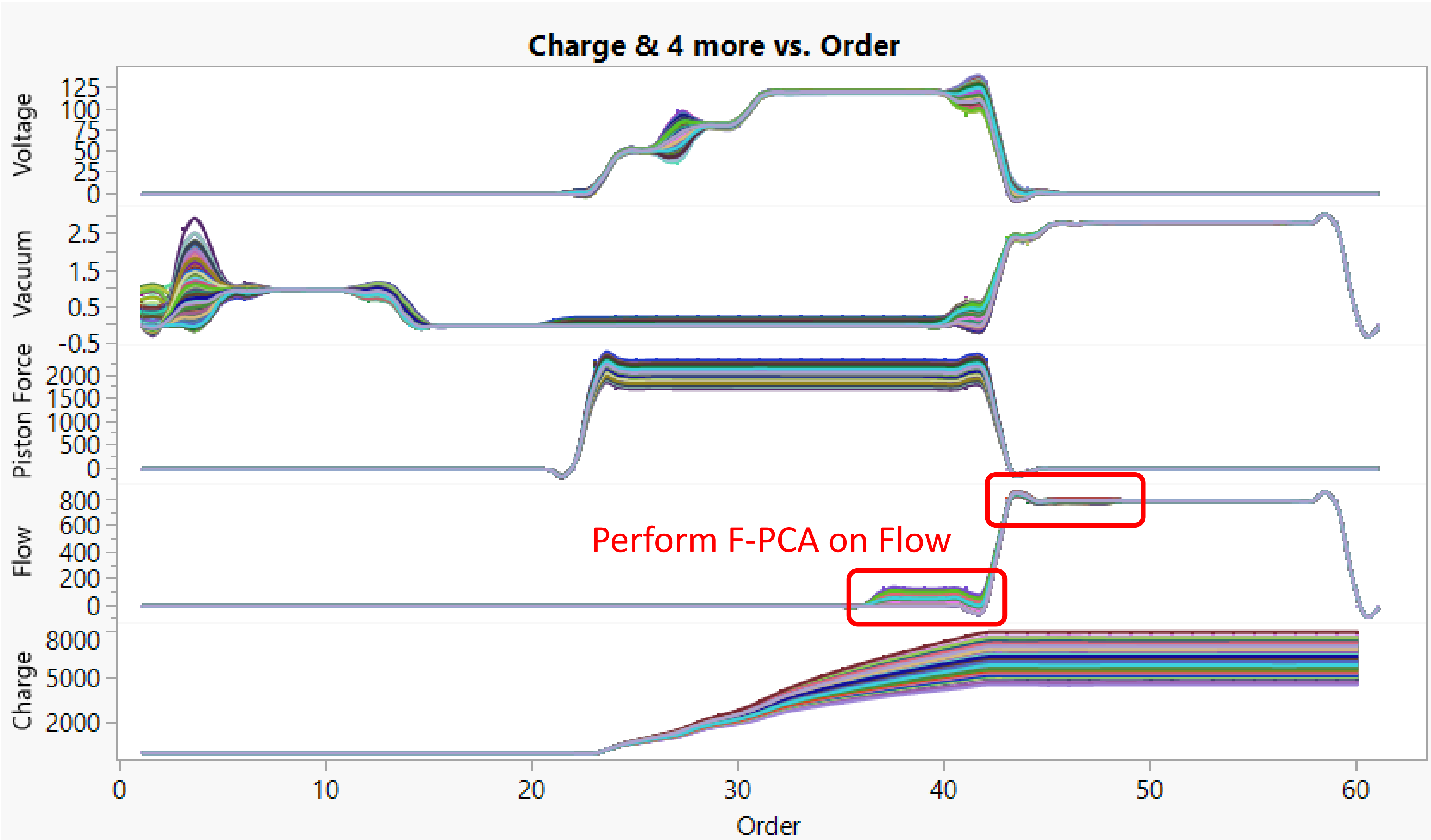
	Wafer Id	Condition	Validation	Order	Charge	Flow	Piston Force	Vacuum	Voltage
1	1	GOOD	Training	1	0.00	0.3013	0	0.00	0.00
2	1	GOOD	Training	2	0.00	0.3013	0	0.00	0.00
3	1	GOOD	Training	3	0.00	0.3013	0	0.50	0.00
4	1	GOOD	Training	4	0.00	0.3013	0	0.50	0.00
5	1	GOOD	Training	5	0.00	0.3013	0	0.94	0.00
6	1	GOOD	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



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



What is Functional Data Analysis?

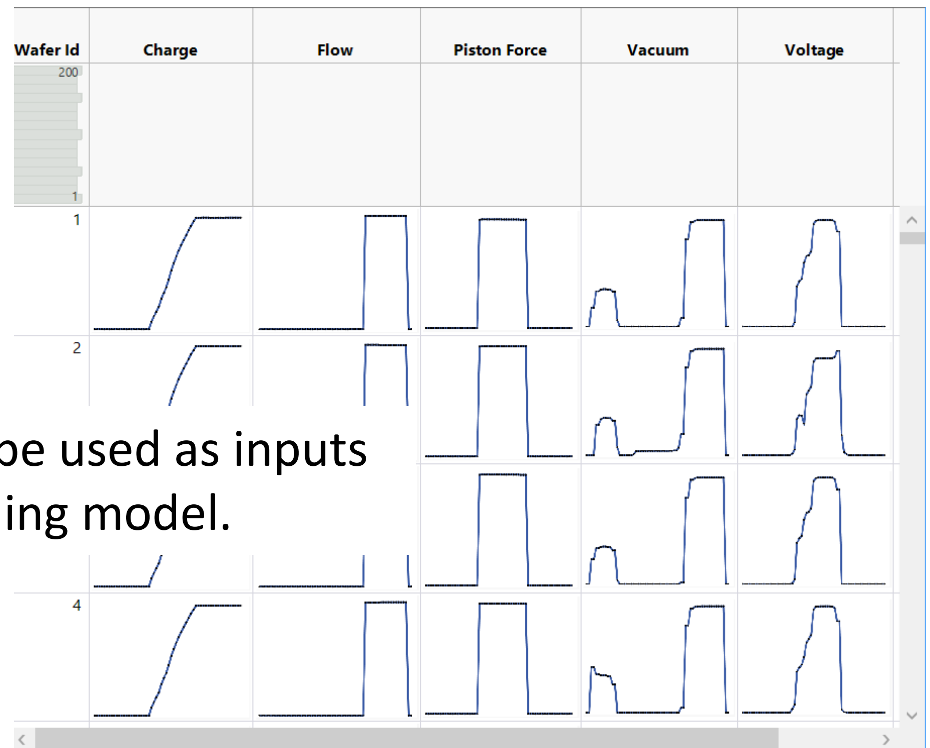
Functional data analysis (FDA) is a branch of statistics that analyzes data providing information about **curves**, **surfaces** or anything else **varying over a continuum**. In its most general form, under an FDA framework each sample element is considered to be a **function**.

2000 wafers X 61 rows/wafer

	Wafer Id	Order	Charge	Flow	Piston Force	Vacuum	Voltage
1	1	1	0.00	0.3013	0	0.00	0.00
2	1	2	0.00	0.3013	0	0.00	0.00
3	1	3	0.00	0.3013	0	0.50	0.00
4	1	4	0.00	0.3013	0	0.50	0.00
5	1	5	0.00	0.3013	0	0.94	0.00
6	1	6	0.00	0.3013	0	0.94	0.00
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	0.00
13	1	13	0.00	2.3e-6	0	0.93	0.00
14	1	14	0.00	2.3e-6	0	0.16	0.00
15	1	15	0.00	2.3e-6	0	0.00	0.00
16	1	16	0.00	2.3e-6	0	0.00	0.00
17							

122,000 rows of data

2000 Functional Data Streams

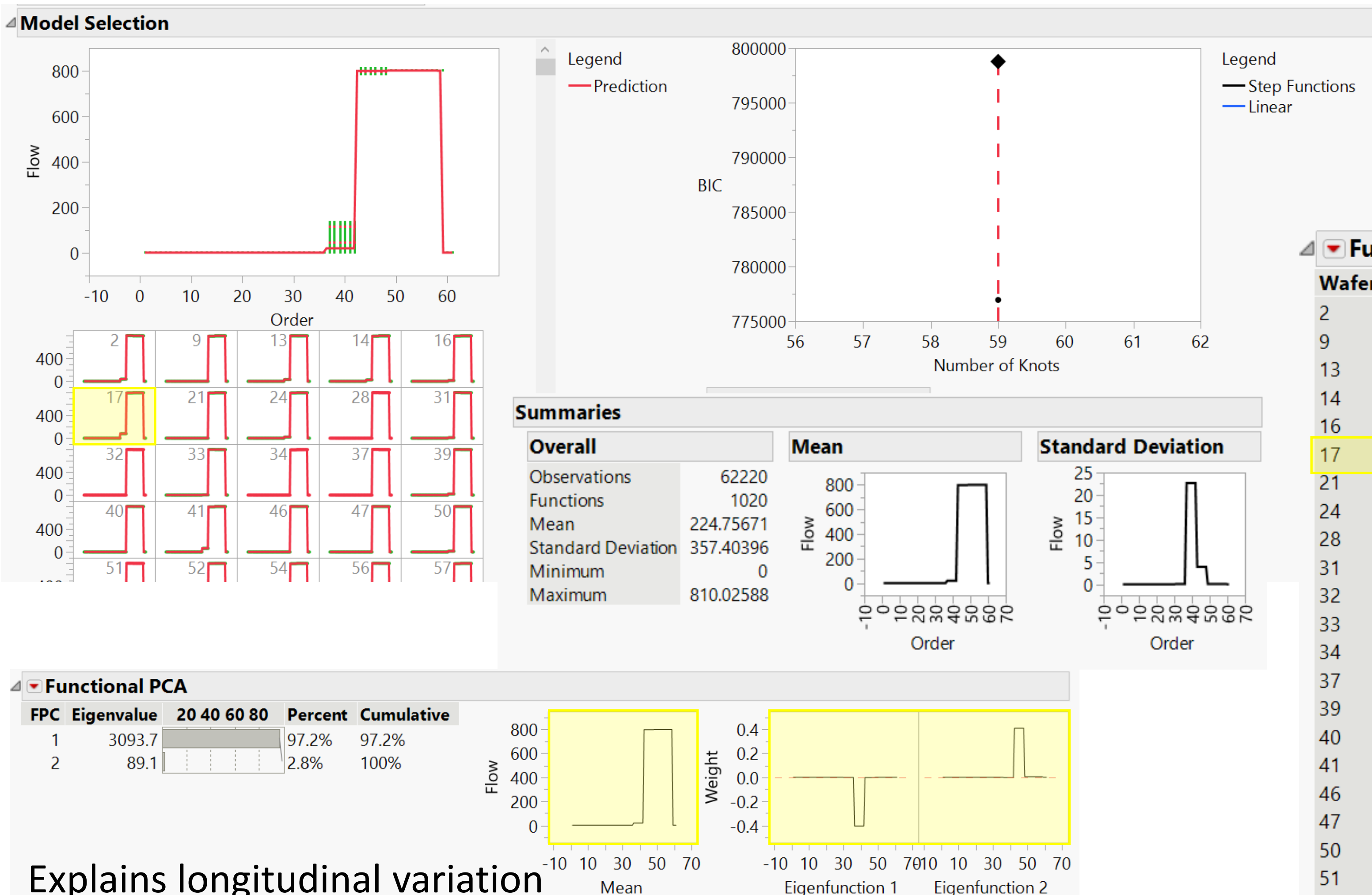


Curve is the fundamental unit of observation

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

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 (Flow) curves.



Summaries – Overall

Observations	62220
Functions	1020
Mean	224.75671
Standard Deviation	357.40396
Minimum	0
Maximum	810.02588

Functional PCA

FPC	Eigenvalue	20 40 60 80	Percent	Cumulative
1	3093.7		97.2%	97.2%
2	89.1		2.8%	100%

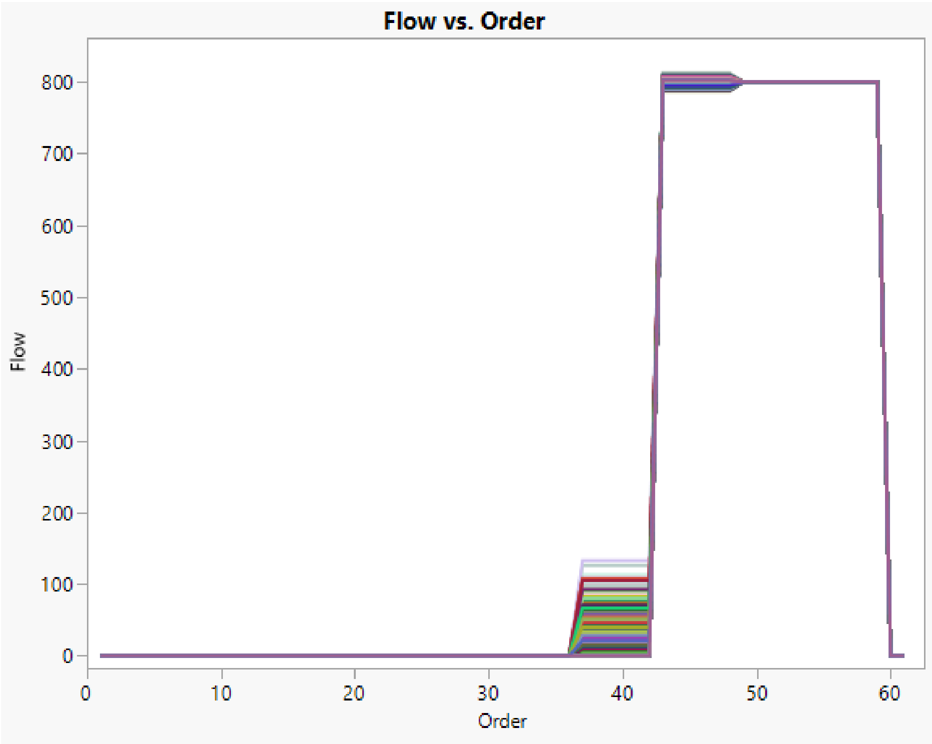
Explains longitudinal variation

Explains function-to-function variation

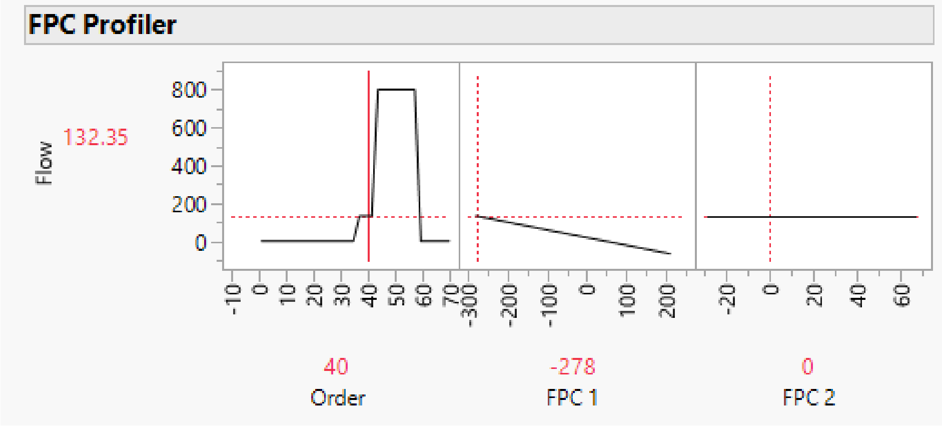
Function Summaries

Wafer Id	FPC 1	FPC 2
2	-46.10211	11.874549
9	47.459441	-10.16256
13	-32.28389	4.9754367
14	-17.78627	-7.657502
16	11.790375	-3.702705
17	-150.7928	-2.945669
21	47.429885	-3.366023
24	-5.124357	-2.054579
28	47.373917	9.5421021
31	21.356508	14.547288
32	47.373123	9.6952779
33	38.728718	11.029717
34	47.430385	-3.453213
37	47.353598	14.193705
39	-3.770084	9.608758
40	47.381348	7.8100951
41	-109.9012	-9.799899
46	47.387051	6.4879083
47	47.389894	5.8384965
50	10.406475	4.6709762
51	47.377736	8.641319
52	47.401717	3.1086259

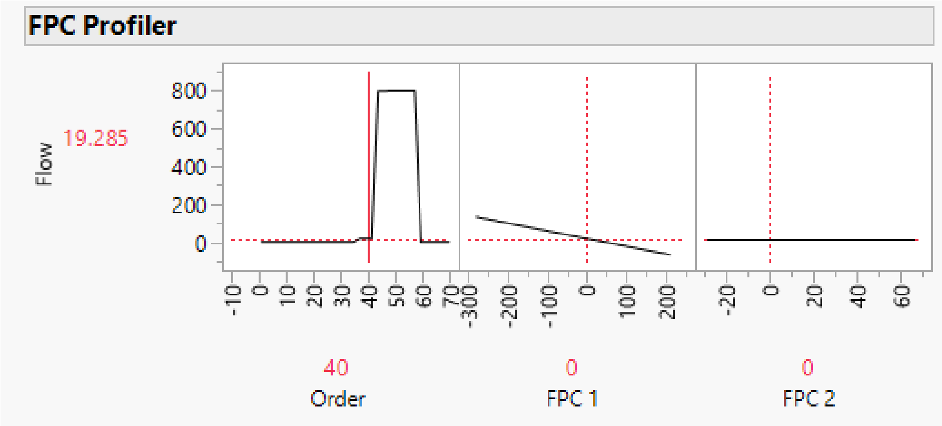
Flow FPC1



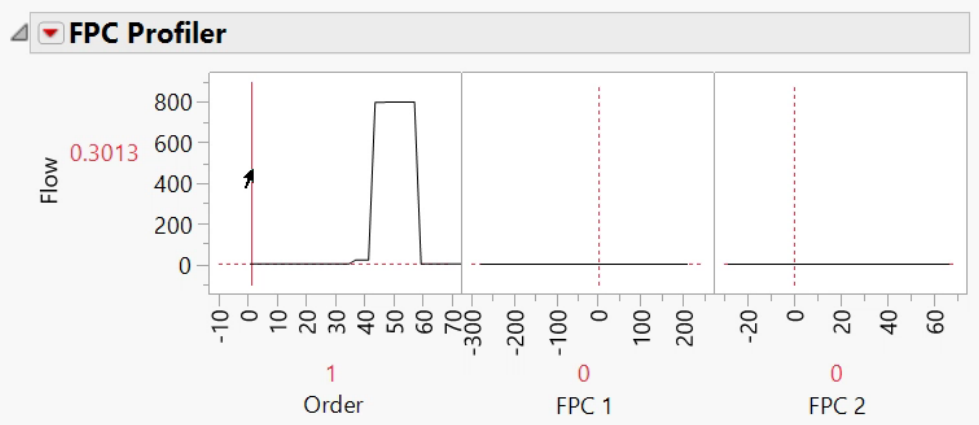
Flow vs. Order
800
700
600
500
400
300
200
100
0
Flow
0
10
20
30
40
50
60
Order



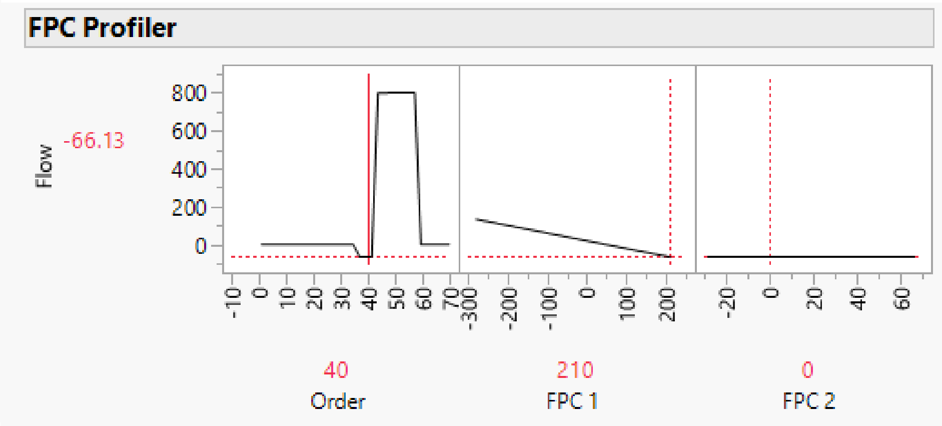
FPC Profiler
Flow
132.35
40
Order
-278
FPC 1
0
FPC 2



FPC Profiler
Flow
19.285
40
Order
0
FPC 1
0
FPC 2



FPC Profiler
Flow
0.3013
1
Order
0
FPC 1
0
FPC 2



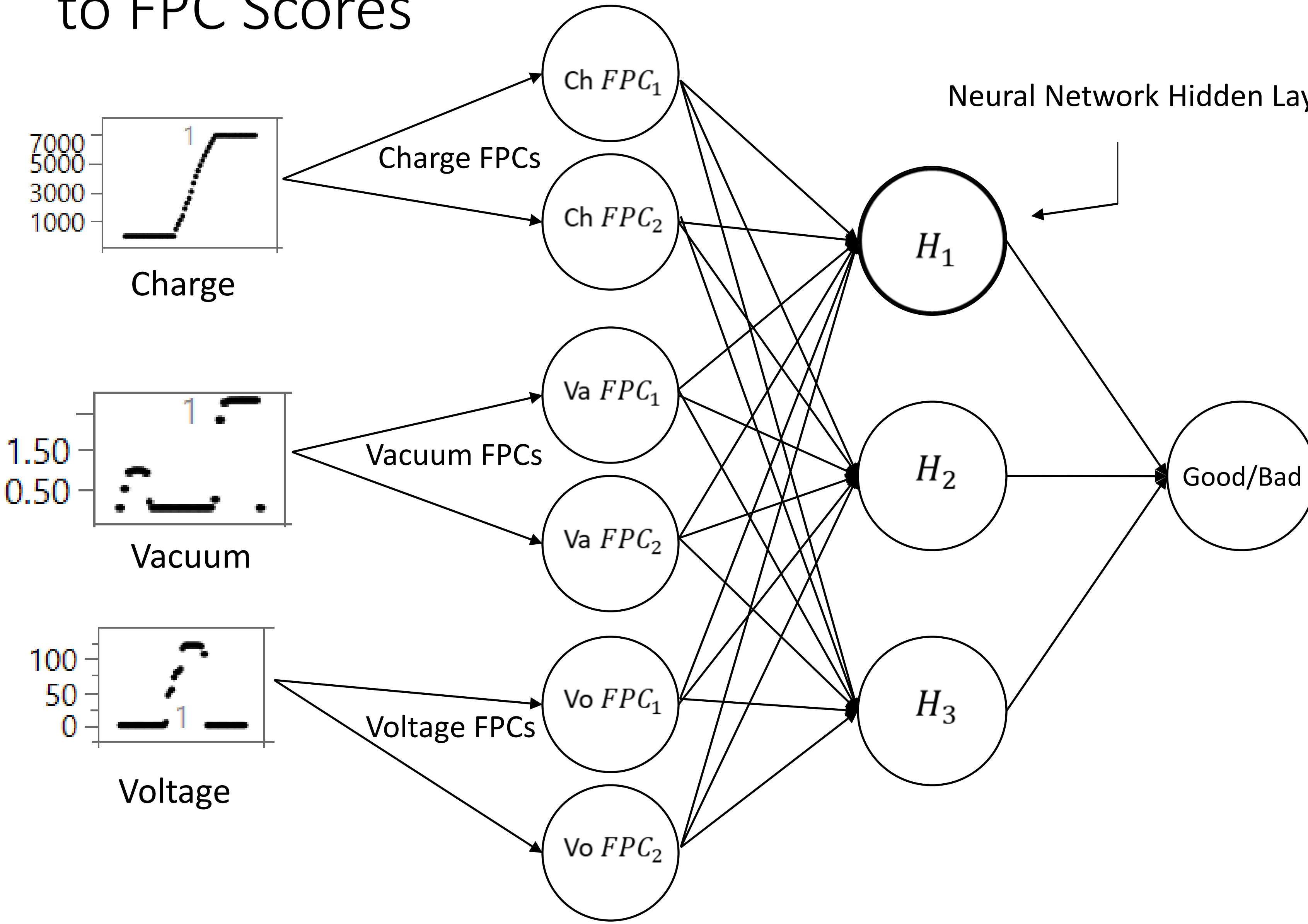
FPC Profiler
Flow
-66.13
40
Order
210
FPC 1
0
FPC 2

Table of 12* FPC Scores used to Model Condition

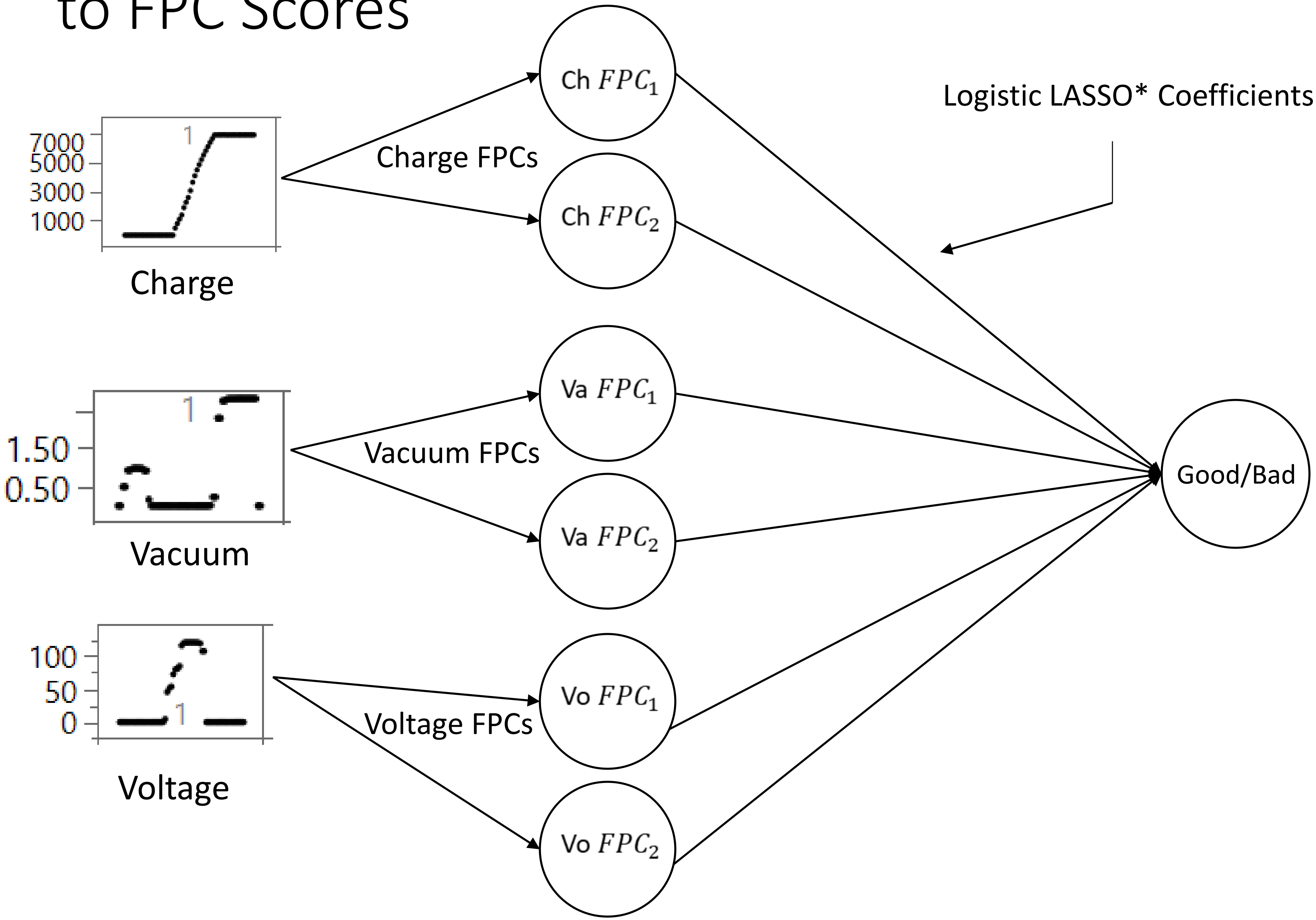
Wafer Id	Condition	Validation	Charge FPC 1	Flow FPC 1	Flow FPC 2	Piston Force FPC 1	Vacuum FPC 1	Vacuum FPC 2	Vacuum FPC 3
1 2 3 4 5 1,995 others	GOOD BAD	Training Validation Test	8295 -8903	45.7 -281	34.7 -28.6	1311 -1442	2.97 -1.04	0.58 -0.72	0.3 -0.59
2	GOOD	Training	1939.8213944	-47.99467507	11.835188042	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 BAD	Training Validation Test	1 0.06	GOOD BAD
o	1564	GOOD	Test	0.9961988532	GOOD
o	1565	GOOD	Test	0.9415687312	GOOD
o	1566	GOOD	Test	0.9670452594	GOOD
o	1567	GOOD	Test	0.5550716285	GOOD
+	1568	BAD	Test	0.4930775442	BAD
+	1569	BAD	Test	0.5335959152	GOOD
o	1570	GOOD	Test	0.6131324314	GOOD
+	1571	BAD	Test	0.8668675093	GOOD
+	1572	BAD	Test	0.2583111765	BAD
o	1573	GOOD	Test	0.8699599618	GOOD
+	1574	BAD	Test	0.7201964039	GOOD
o	1575	GOOD	Test	0.99613207	GOOD
+	1576	BAD	Test	0.5647955597	GOOD
o	1577	GOOD	Test	0.9891191688	GOOD
o	1578	GOOD	Test	0.987960478	GOOD
o	1579	GOOD	Test	0.9750283598	GOOD

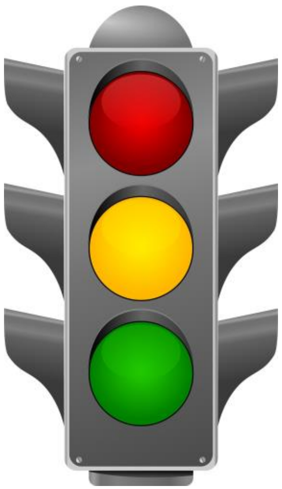
Model these probabilities to develop decision tree “stoplight.”

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

Level	Rate	Prob	Count
GOOD	0.8830	0.8830	883
BAD	0.1170	0.1170	117

Probability(Condition=GOOD)<0.7541564548

Level	Rate	Prob	Count
GOOD	0.4596	0.4622	74
BAD	0.5404	0.5378	87

Probability(Condition=GOOD)>=0.7541564548

Level	Rate	Prob	Count
GOOD	0.9642	0.9641	809
BAD	0.0358	0.0359	30

Candidates

96%/4%

Probability(Condition=GOOD)<0.3950198207

Level	Rate	Prob	Count
GOOD	0.0476	0.0661	2
BAD	0.9524	0.9339	40

Candidates

5%/95%

Probability(Condition=GOOD)>=0.3950198207

Level	Rate	Prob	Count
GOOD	0.6050	0.6070	72
BAD	0.3950	0.3930	47

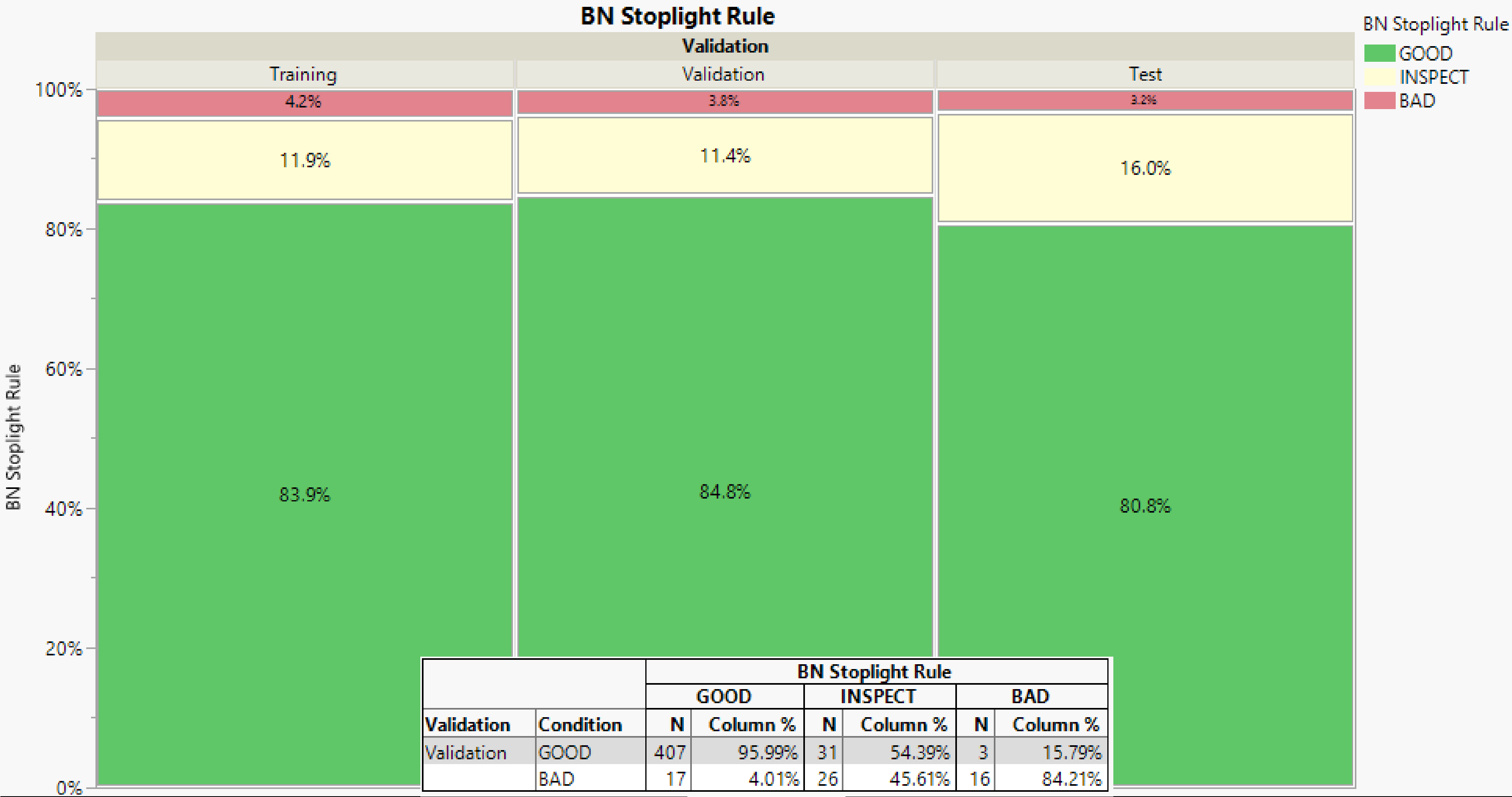
Candidates

60%/40%

Table of Model Predictions and Stoplight Rule

	Condition	Validation	Probability(Condition=GOOD)	BN 1-1-1(16) Most Likely Condition	BN Stoplight Rule
18/0 Cols	GOOD BAD	Training Validation Test	1 0.06	GOOD BAD	GOOD INSPECT BAD
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



		BN Stoplight Rule					
		GOOD		INSPECT		BAD	
Validation	**Condition**	**N**	**Column %**	**N**	**Column %**	**N**	**Column %**
Validation	GOOD	407	95.99%	31	54.39%	3	15.79%
	BAD	17	4.01%	26	45.61%	16	84.21%

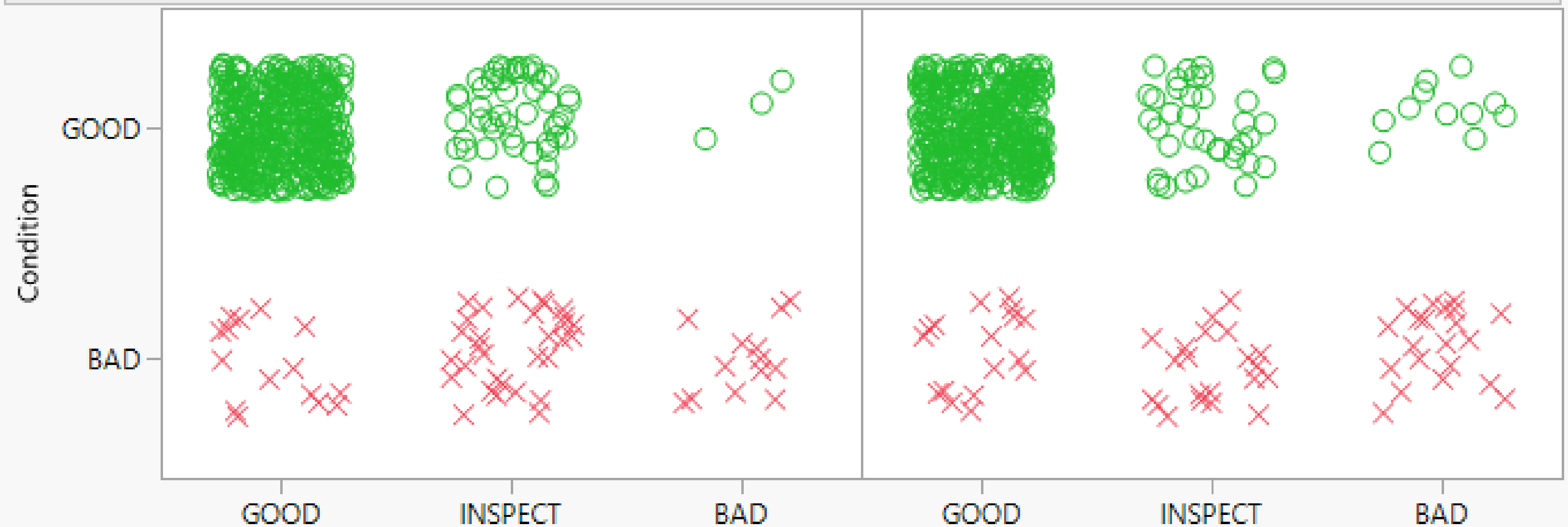
		BN Stoplight Rule					
		GOOD		INSPECT		BAD	
Validation	**Condition**	**N**	**Column %**	**N**	**Column %**	**N**	**Column %**
Training	GOOD	809	96.42%	72	60.50%	2	4.76%
	BAD	30	3.58%	47	39.50%	40	95.24%

		BN Stoplight Rule					
		GOOD		INSPECT		BAD	
Validation	**Condition**	**N**	**Column %**	**N**	**Column %**	**N**	**Column %**
Test	GOOD	389	96.29%	49	61.25%	3	18.75%
	BAD	15	3.71%	31	38.75%	13	81.25%

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

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^{th} or 20^{th} 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.

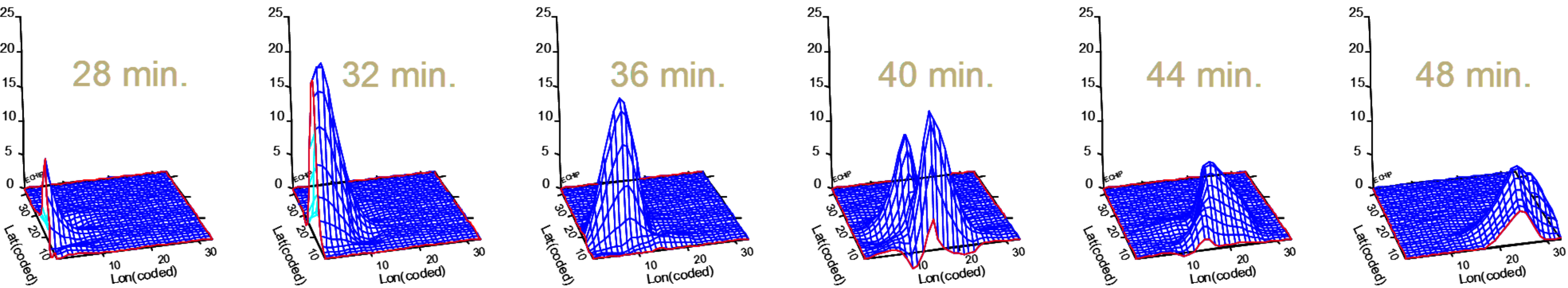
- When Functional Data Analysis of a response is combined with Design of Experiments one can model the shape of the data stream as a function of the design factors.

- 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 15 Years ago at the Army’s *Edgewood Chemical Biological Center*

Now…CCDC CBC

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



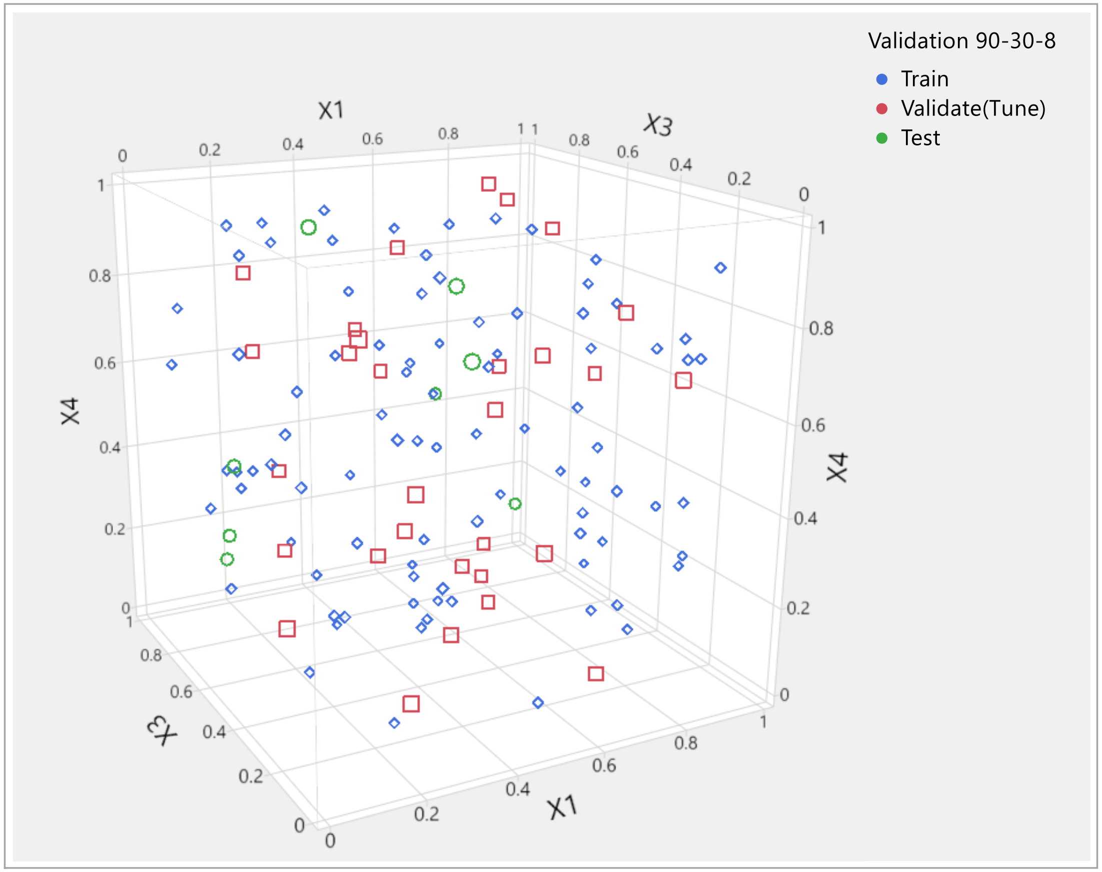
10-factor Agent Transport & Dispersion Simulation

- 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

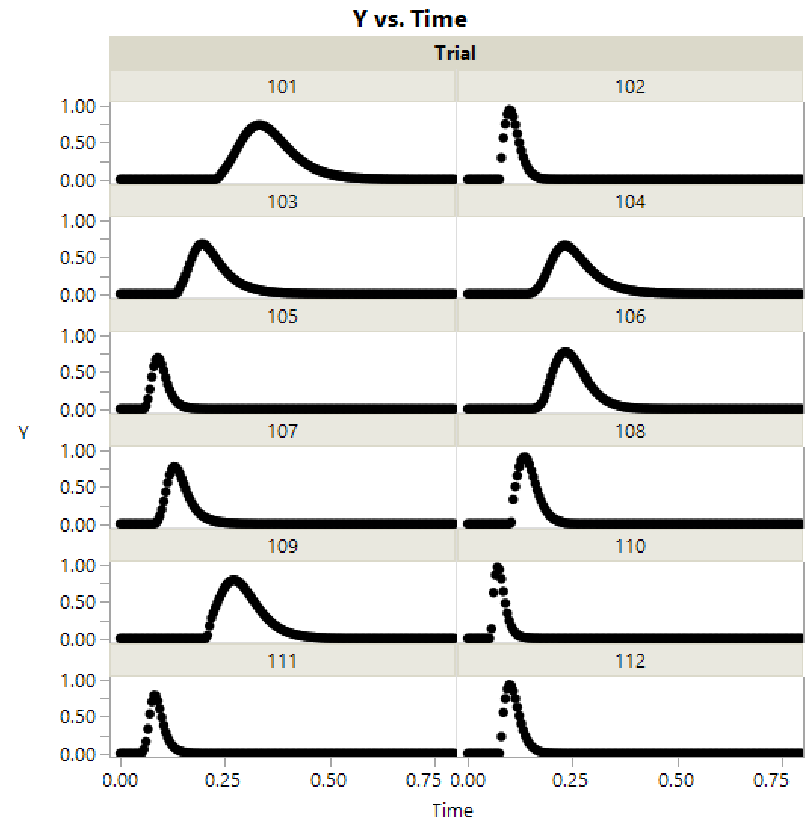
128 Computer Simulations Split into 3 Subsets:
90 Training, 30 Validation(Tune), and 8 Test



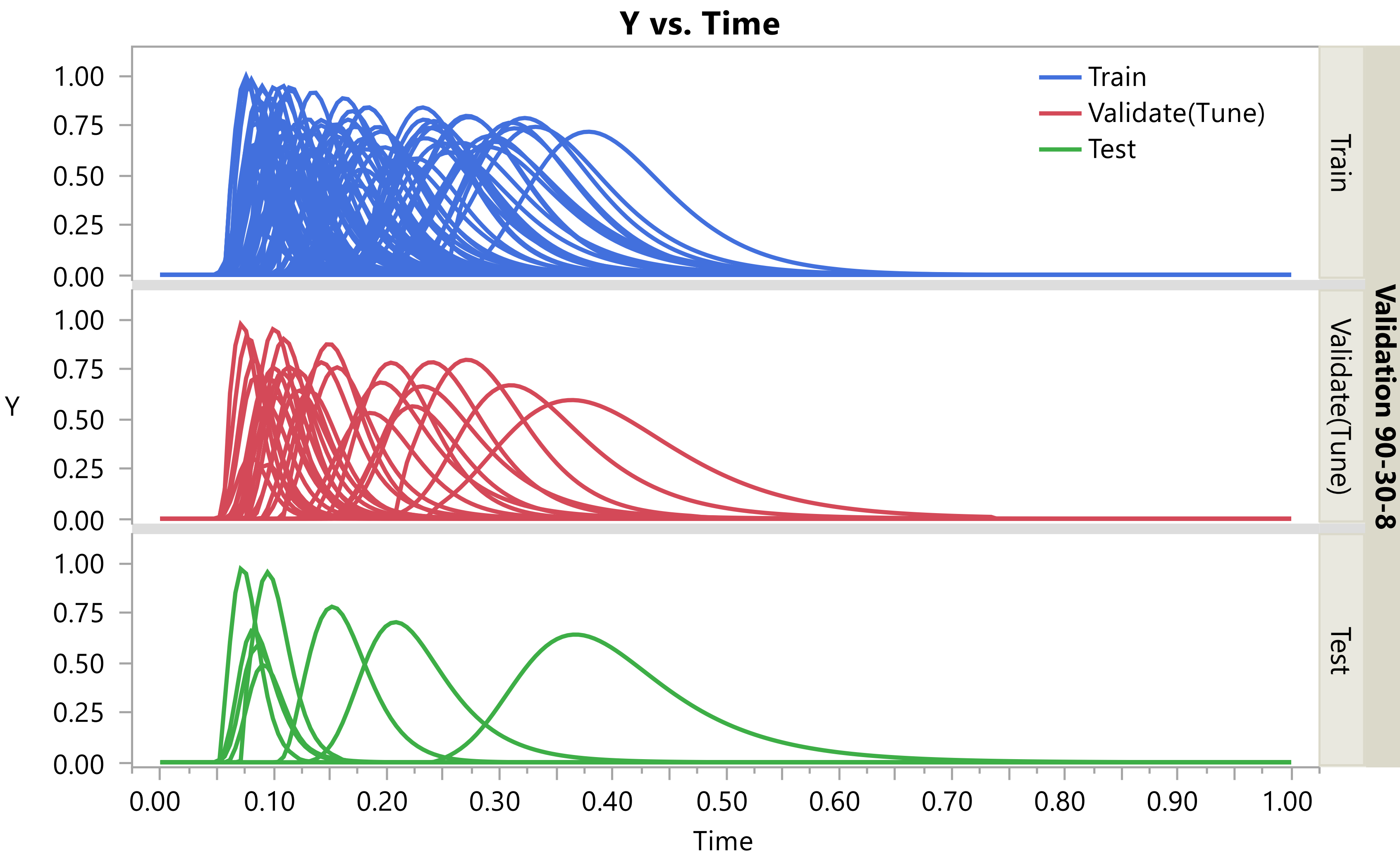
128 Unique-Trial Space-Filling Design of Experiments

	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

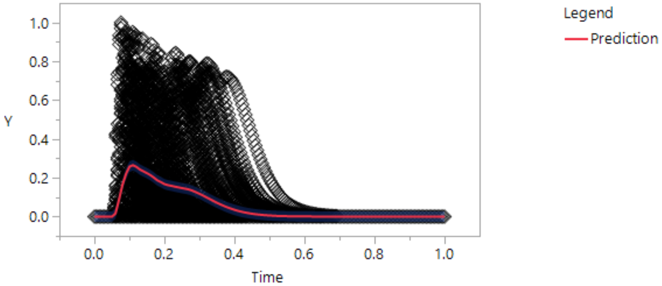
Y vs Time Data for Each Trial

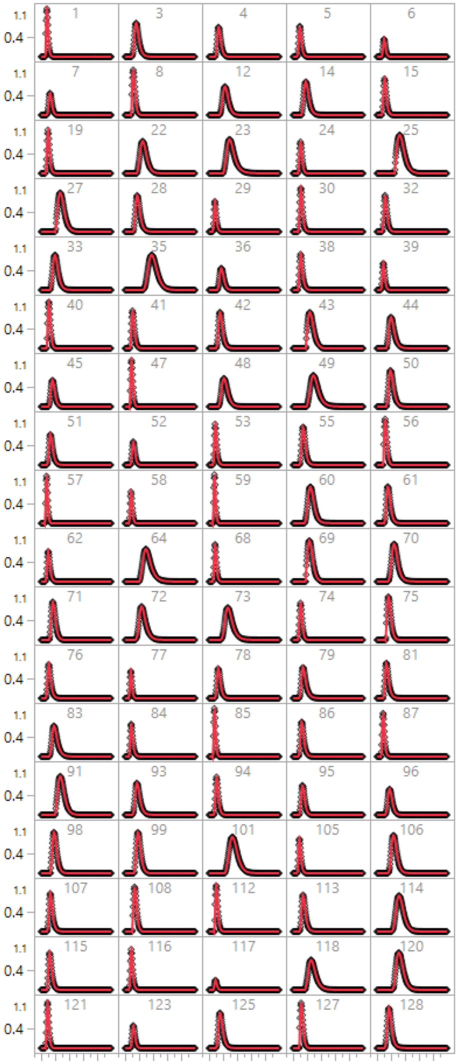


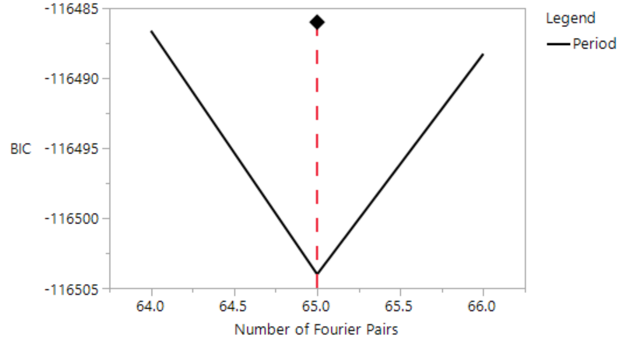
128 Simulations Split into 3 Subsets: 90 Training, 30 Validation(Tune), and 8 Test



Model Selection







Fourier Basis Model on Initial Data

90 Training, 30 Validation, 8 Test

Fit Statistics

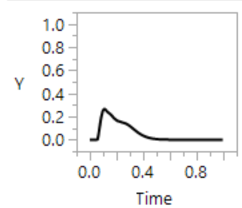
Pairs	65
-2 Log Likelihood	-119076.5
AICc	-118547.3
BIC	-116504
GCV	0.0002135
Y Std Dev	0.003181

Summaries

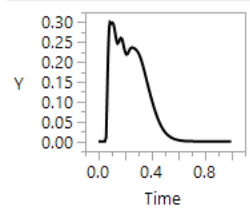
Overall

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

Mean



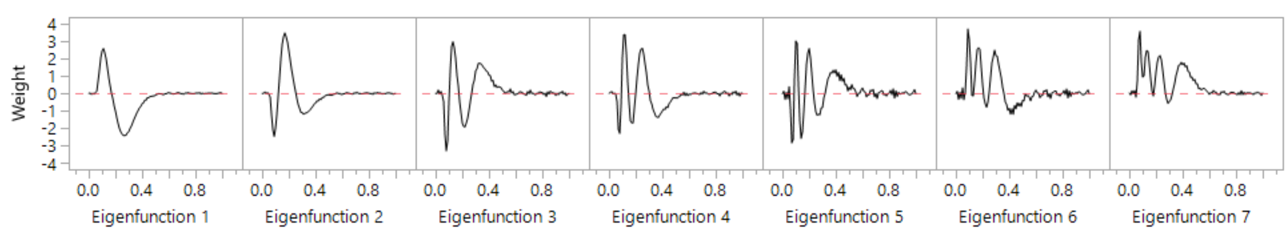
Standard Deviation



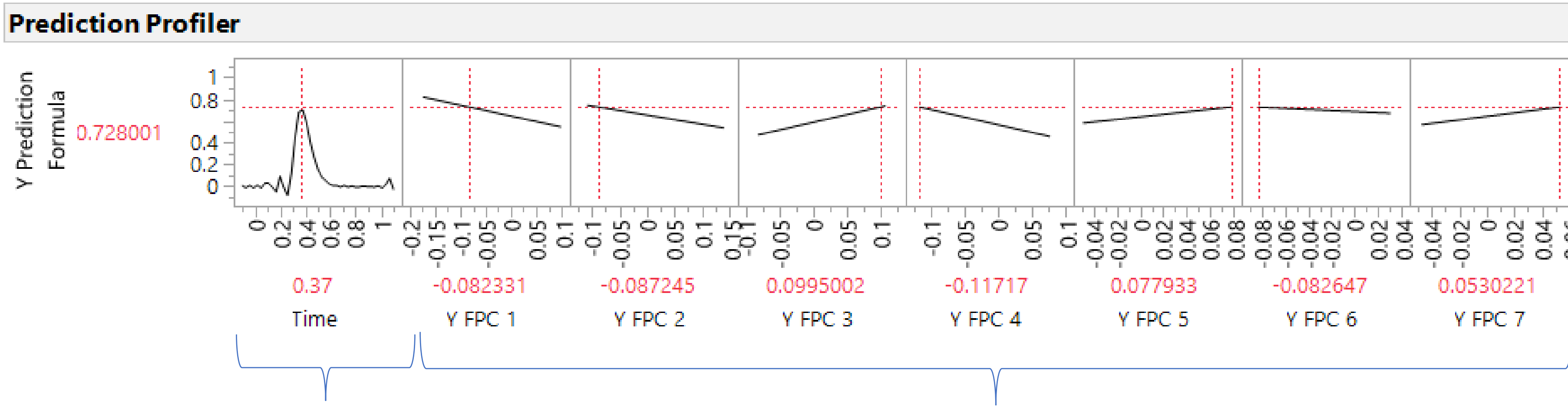
FPC	Eigenvalue	20 40 60 80	Percent	Cumulative
1	0.00771		42.8%	42.8%
2	0.00514		28.5%	71.3%
3	0.00240		13.3%	84.6%
4	0.00155		8.61%	93.2%
5	0.00054		3.02%	96.2%
6	0.00032		1.78%	98%
7	0.00019		1.06%	99.1%

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

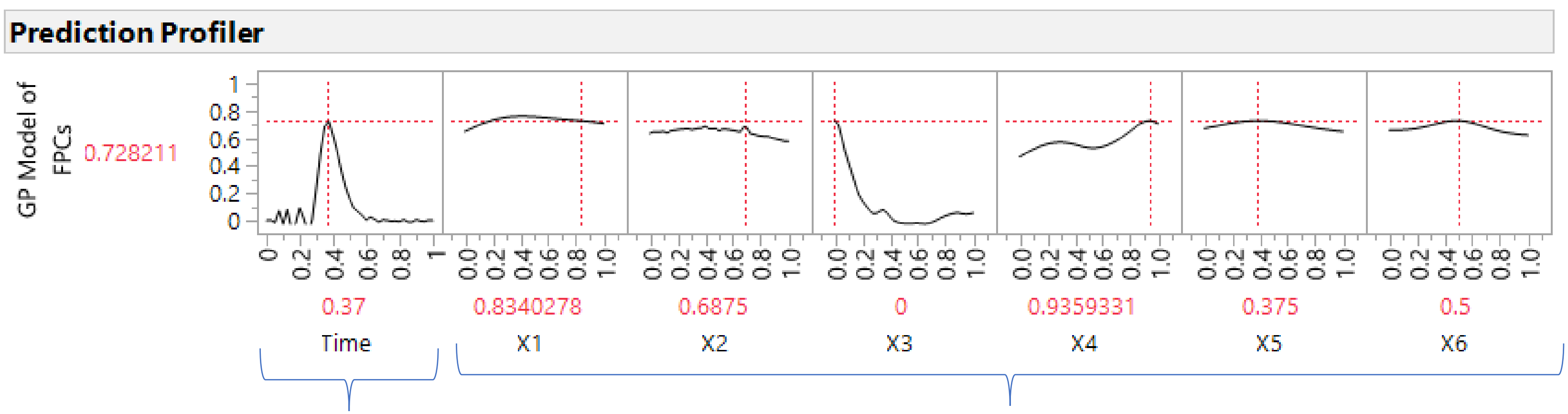


FPC Scores fit as function of DOE factors using Gaussian Process Model



Functional Data ***Curve***

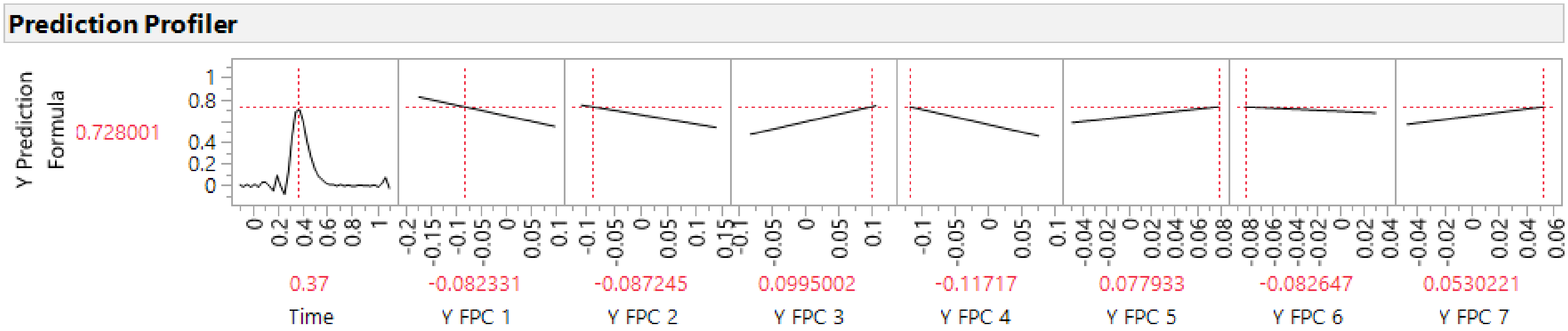
FPC Scores



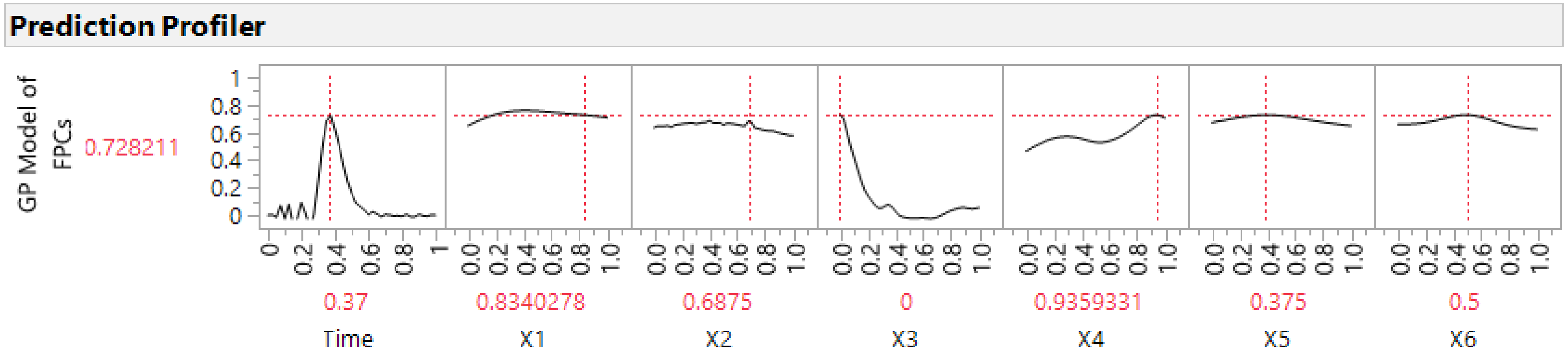
Functional Data ***Curve***

DoE Factors

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

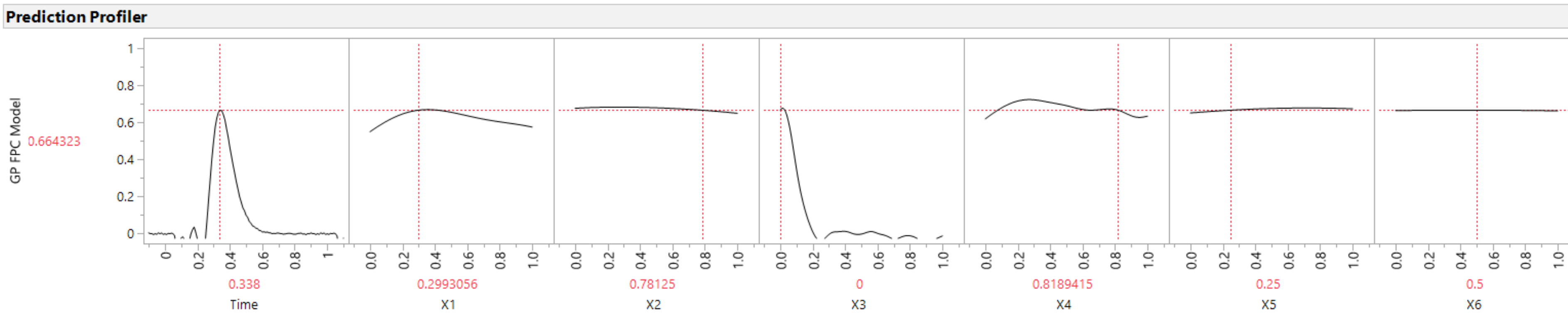


Functional Data ***Curve*** = Σ(“Y_i FPC Score" * “Y_i Eigenfunction“) + “Y Mean Formula"



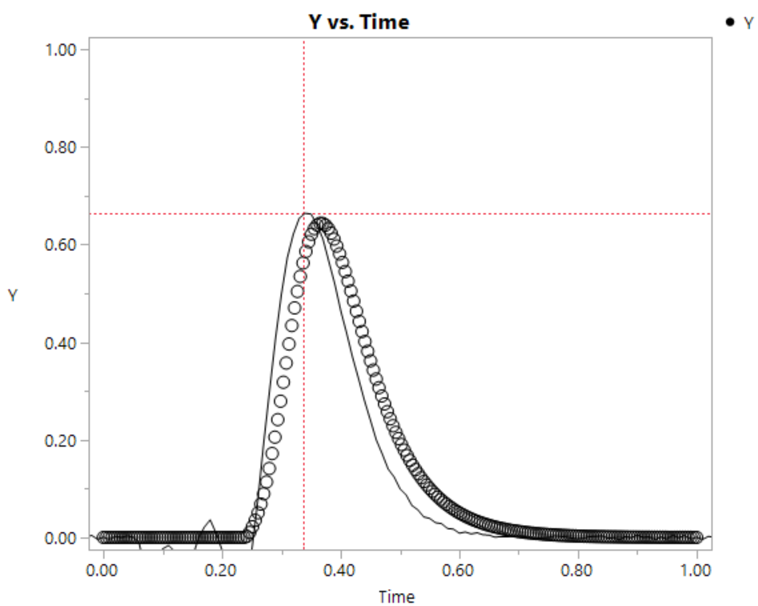
Functional Data ***Curve*** = Σ(“Y_i FPC Score Prediction Formula" * “Y_i Eigenfunction“) + “Y Mean Formula"

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



Functional Data ***Curve***

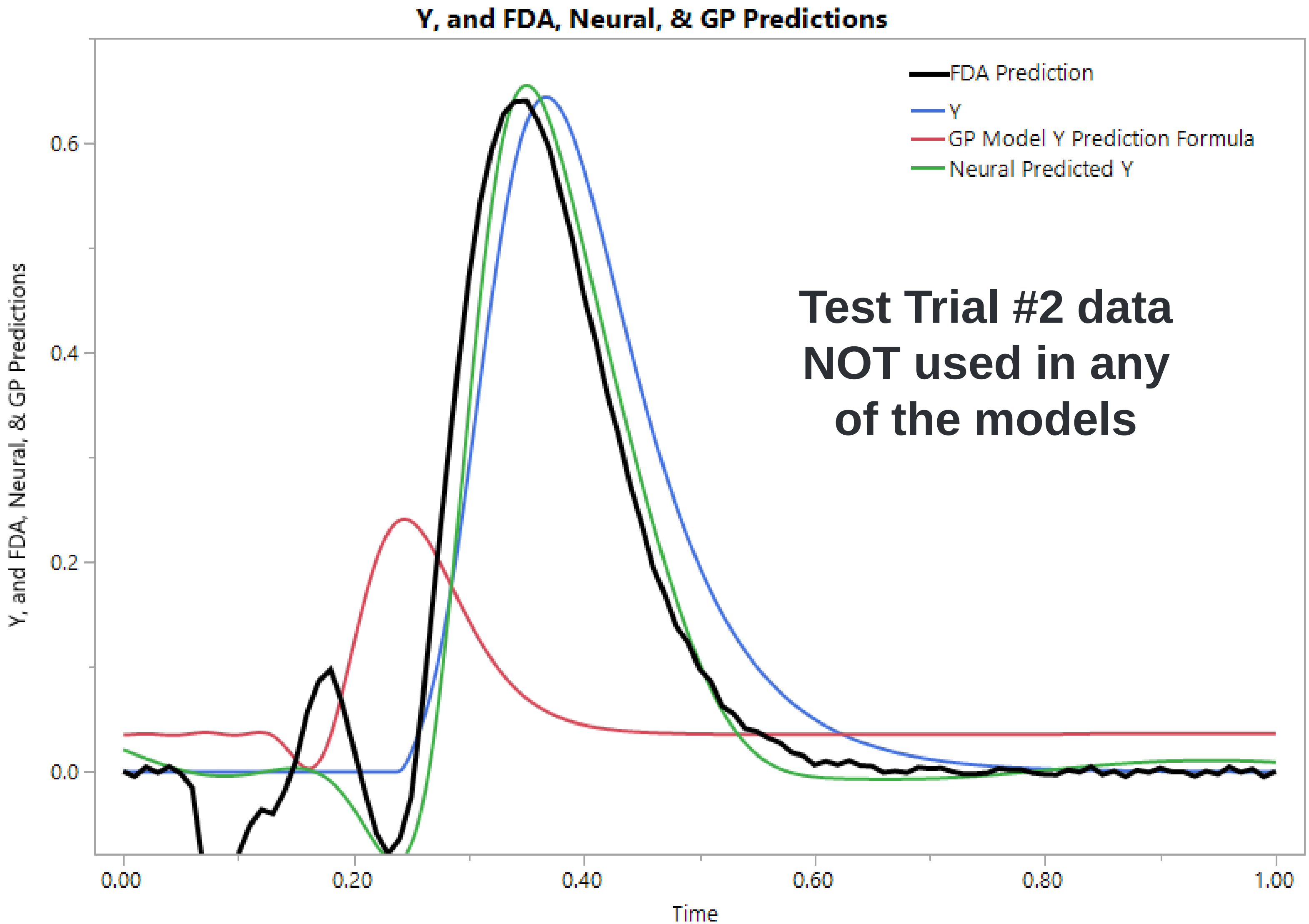
DoE Factors



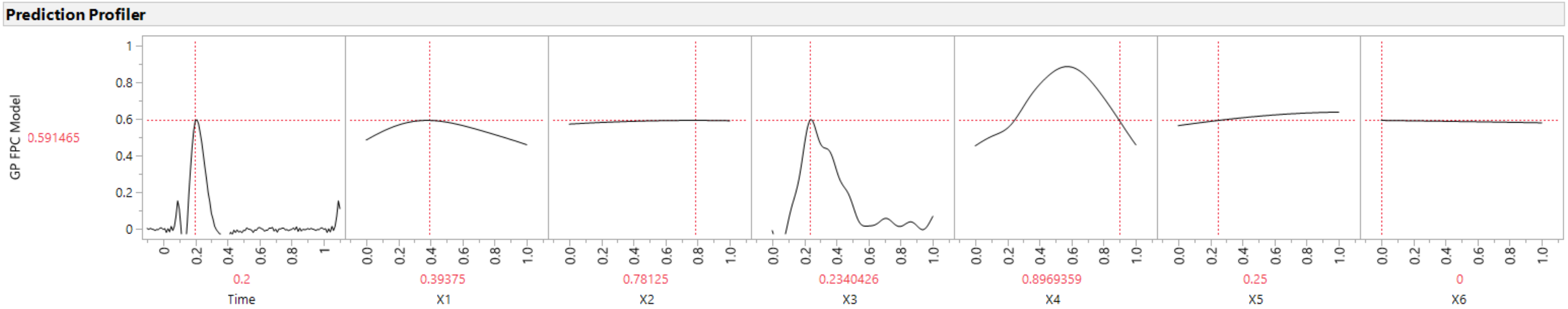
Test Trial #2

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

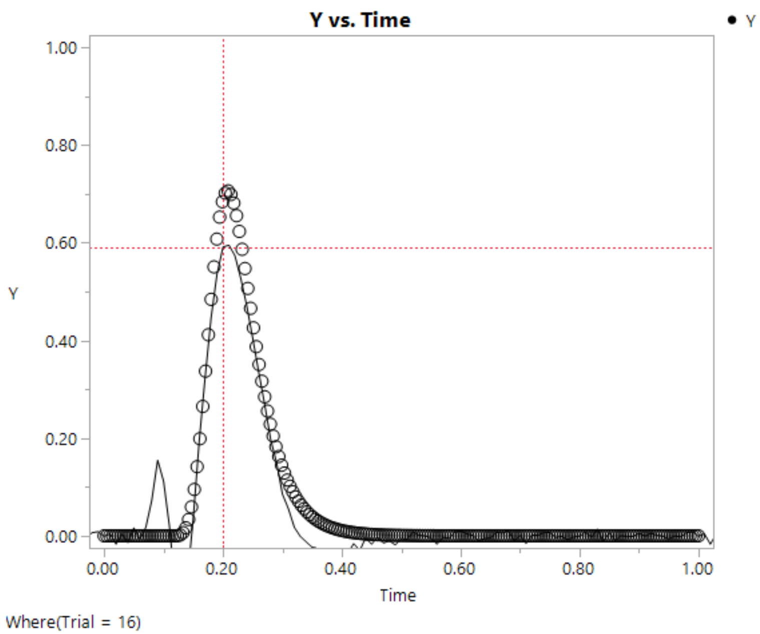


FPC Scores fit as function of DOE factors using Gaussian Process Model



Functional Data ***Curve***

DoE Factors



Test Trial #16

Overlay of simulation data on top of Functional Data Curve

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

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.

Graph Builder	***Text Exploration***	***Functional Data Analysis***
Tour of Graph Elements Video GB-1	***Intro to Visualization and Modeling***	***Functional DOE – with Target Function*** Video FDA-1
Modeling with GB Video GB-2	***Prepping the Data*** Video TXT-2	***Prepping the Data*** Video FDA-2
Adding Maps, Images, and Animation Video GB-3	***Analyses with JMP Pro*** Video TXT-3	***Functional Machine Learning*** Video FDA-3

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.

Data Wrangling	*Machine Learning*	*Design of Experiments*
Getting Data into JMP Video DW-1	***Intro & Honest Assessment*** Video ML-1	***Custom DOE – Making Design Fit Problem*** Video DOE-1
Prepping Data Video DW-2	***ML Analyses with JMP*** Video ML-2	***Screening Designs – Get More for Less*** Video DOE-2
Using Data Table Tools Addin Video DW-3	***ML Analyses with JMP Pro*** Video ML-3	***Comparing Designs & Fixing Broken Designs*** Video DOE-3

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

Thank You.

Questions?

Webcast recordings at 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
302-489-9291

Statistical Discovery.™ From SAS.

ABSTRACT:
Sensors that record sequences of measurements are now embedded in many systems. There is information in the shapes of the sensor stream that is highly predictive of the likelihood of a system failure or performance. These data are often being used inefficiently due to lack of knowledge and tools for how to properly leverage it. In this presentation we will show how to fit splines to data streams and extract features called functional principal component scores. This method is called Functional Data Analysis. Then, we use these features as inputs into machine learning models like neural networks. Answering a wide variety of questions becomes a two-step process of functional feature extraction followed by modeling using those features as inputs. Additionally, it will be shown how when combined with Design of Experiments, one can then model the principal component scores to predict the shapes of data streams as functions of the factors in the design.