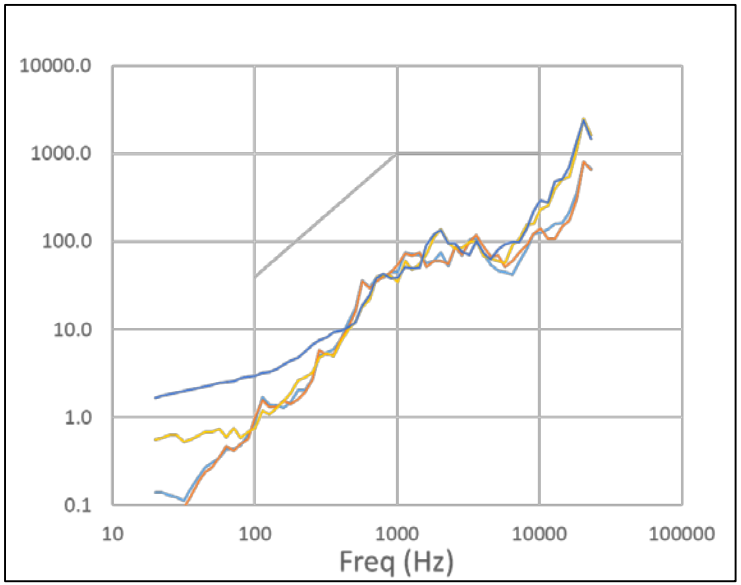
Functional Data Analysis for Aerospace

Dataworks 2019

Lockheed Martin Corporation

Functional Data

- Sensor data—based on time or frequency
  - Accelerometers
  - Load Cells
  - Temperature
  - Pressure
  - Shock
  - Vibe
  - Acoustic
  - Many more

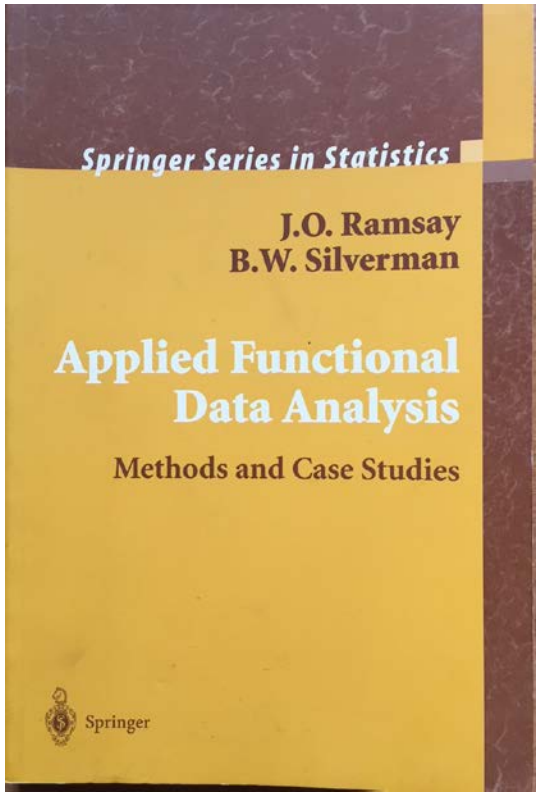


FUNCTIONAL DATA IS EVERYWHERE

Functional Data Analysis: 1997 through 2019

Ramsay and Silverman:
1997: Functional Data Analysis
2002: Applied Functional Data Analysis
2009: Functional Data analysis with R and MATLAB

Website: Springer-ny.com
Bernardsilverman.com



JMP released their "Functional Data Explorer" tool in April 2018

SUBJECT IS NOT OLD, BUT ALSO NOT NEW

Ramsey and Silverman Functional Data Analysis Examples:

Juggling Loop

Writing Analysis

Landmark vs FDA analysis of prehistoric bones
Growth of human children over time
Life course data for criminal activity
Cyclic variations in non-durable goods index
Lip acceleration during speech

Main Hypothesis: the data is "in-family"

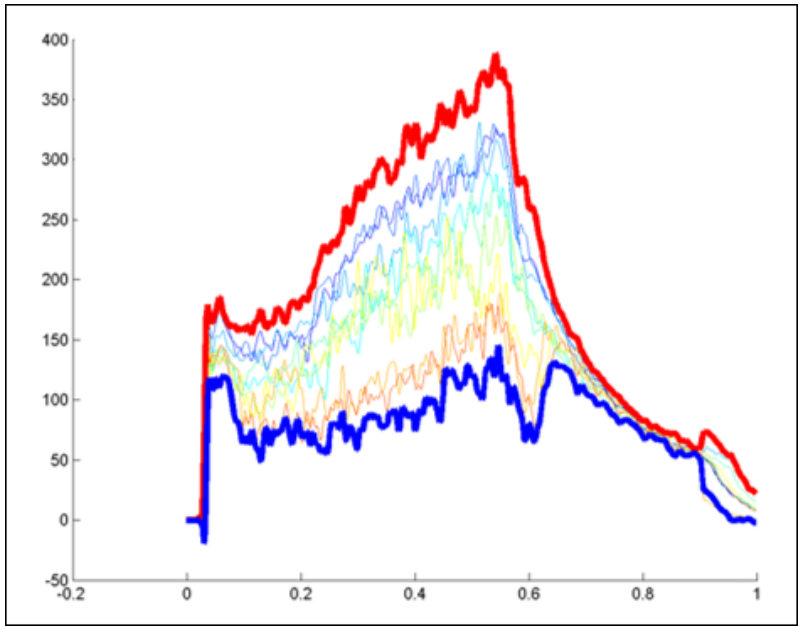
Teams are looking to trigger engineering discussion

Type I Error: Reject when actually true—the team goes looking for a root cause of the change when there is not a change

Type II Error: Team accepts the data as statistically similar when there has actually been a change in the process

Example: Multi-company project to compare designs

- A 1980's component was re-designed to be more adaptable to multiple vehicles
- If the new design produced "in-family data" over the course of multiple tests many of the design changes could be accepted without further review
- Both the 1980's and 2016 test design was OFAT (i.e. statistically weak)
- Analysis processes for the data were not decided in conjunction with the test design
- Multiple analysis teams struggled to analyze the results



1-SECOND WORTH OF DATA,
30 DIFFERENT TEST SENSORS,
MID-1980'S TEST REPEATED IN 2016

Three Analysis Options Covered and Compared

- Single-point comparison

  Develops go/no-go limits for each time period

- Landmark comparison

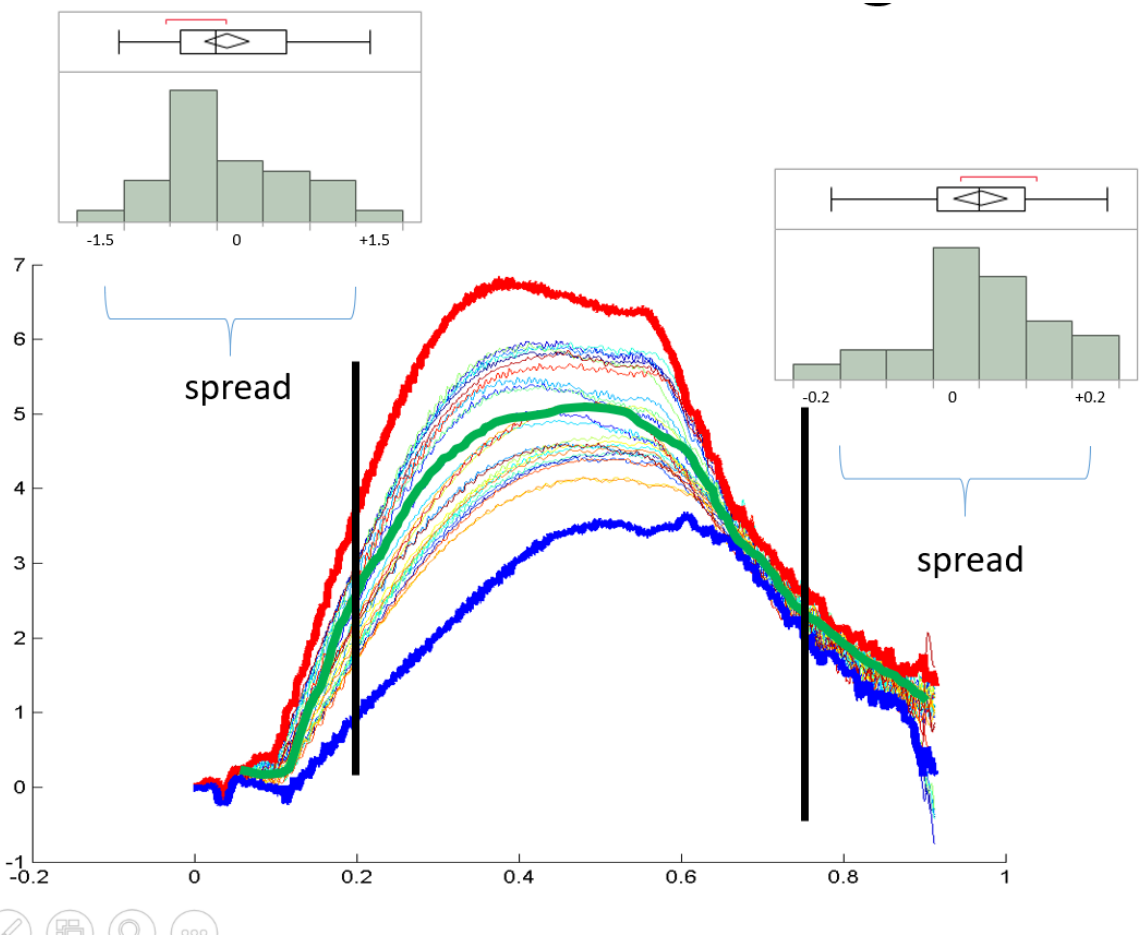
  Compares specific, strategic points along each curve

- Functional Data Analysis

  Develops spline curves and then compares principal components

A KEY GOAL OF ANY OF THESE ANALYSES IS TO TRIGGER ENGINEERING DISCUSSIONS; THE LAST TWO CAN DEVELOP FULL MODELS

Single-Point Review: Overview



- Each histogram of normalized values has a “spread” (variance/standard deviation)
- In this case, the spread on the rise is much larger than the spread on the fall.

- This creates a “spread” for each point in time…+/-2*STDEV = 95% confidence limits for in-family limits

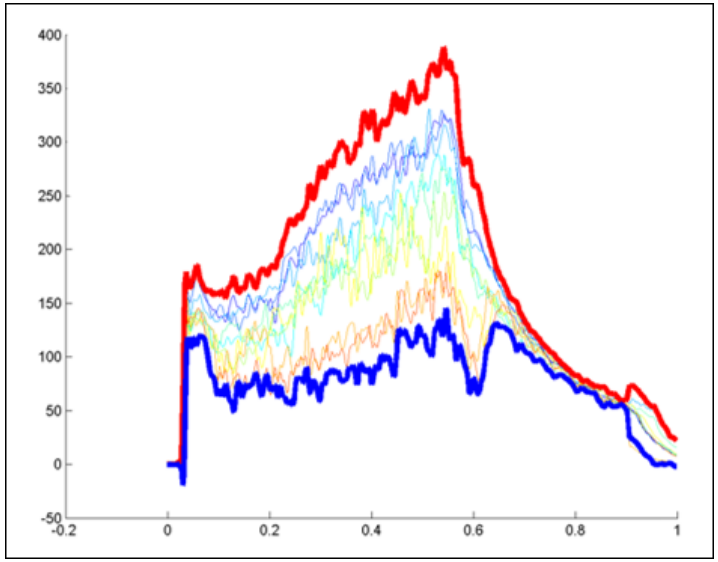
- A similar process for the median (using the interquartile range) produces results less affected by outliers or low percentage results

(the average is affected by data point leverage, the median is not)

Two Options Based on Normality

- The process is based on setting 2-sigma (95%) and 3-sigma (99%) limits based on previous data.

- A normality assessment on each set of data points hints at the data structure.

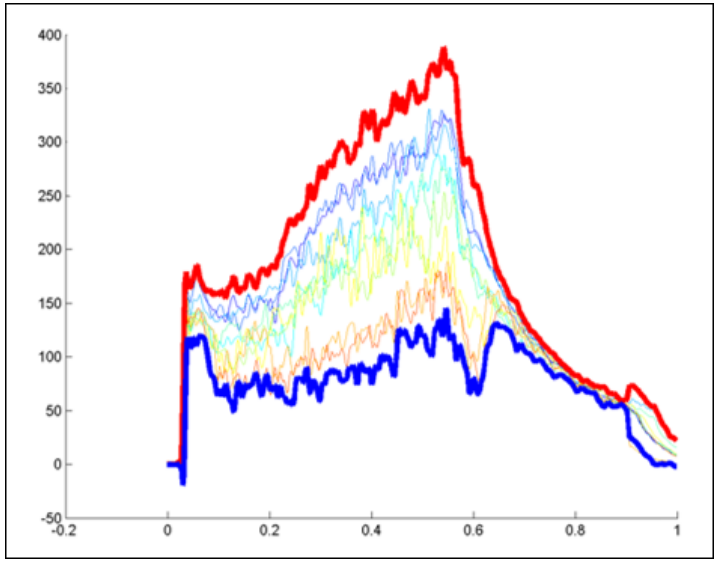


IF THE DATA IS NORMALLY DISTRIBUTED A MEAN/STANDARD DEVIATION METHOD IS APPLICABLE

Two Options Based on Normality

- If data is a small data set, or the data is not normally distributed, then the assessment can rely on a median/pseudo-standard deviation.

- Pseudo-standard deviation?



IF THE DATA IS NOT NORMALLY DISTRIBUTED, A MEDIAN-BASED ANALYSIS IS MORE APPLICABLE

Pseudo-Standard Deviation

- Calculate the Inter-Quartile Range; 25% to 75% data points

- Z-score for normal distribution (25%/50%) = 1.349

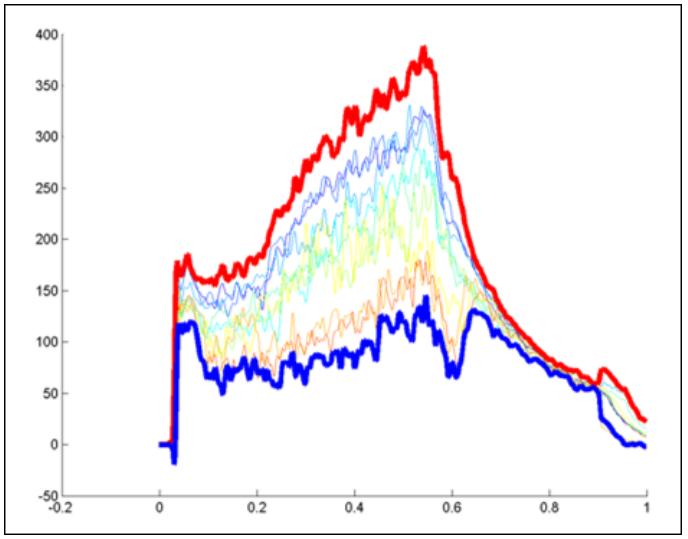
- $1.349*\sigma = IQR$[1]
- $\sigma_{pseudo} = IQR/1.349$

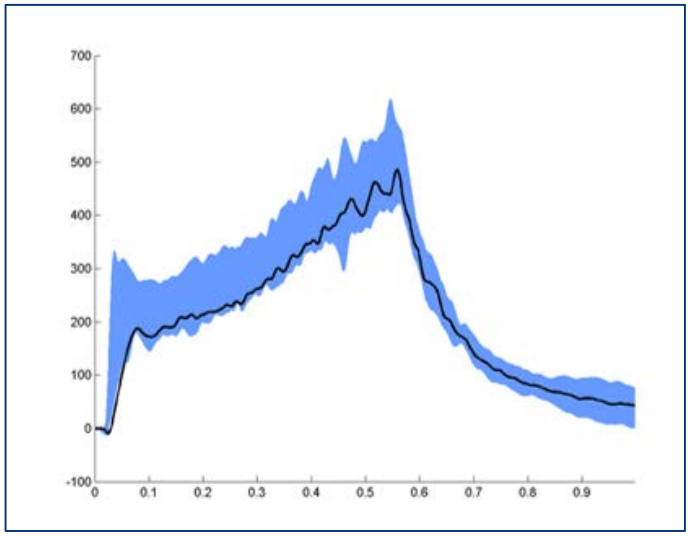
USING THE IQR GENERATES AN ACCEPTABLE 68%-95% COVERAGE VALUE COMPARABLE TO A STANDARD DEVIATION

Whaley, Dewey Lonzo, "The Interquartile Range: Theory and Estimation." (2005). *Electronic Theses and Dissertations.* Paper 1030. http://dc.etsu.edu/etd/1030

Analysis Method Overview: Compare Each Point

- 1980's data was used to generate "in-family" limits
- 20 years of performance confirmed "accuracy" of system
- In-Family Limits were used for 2016 test data to show 95% confidence for in-family

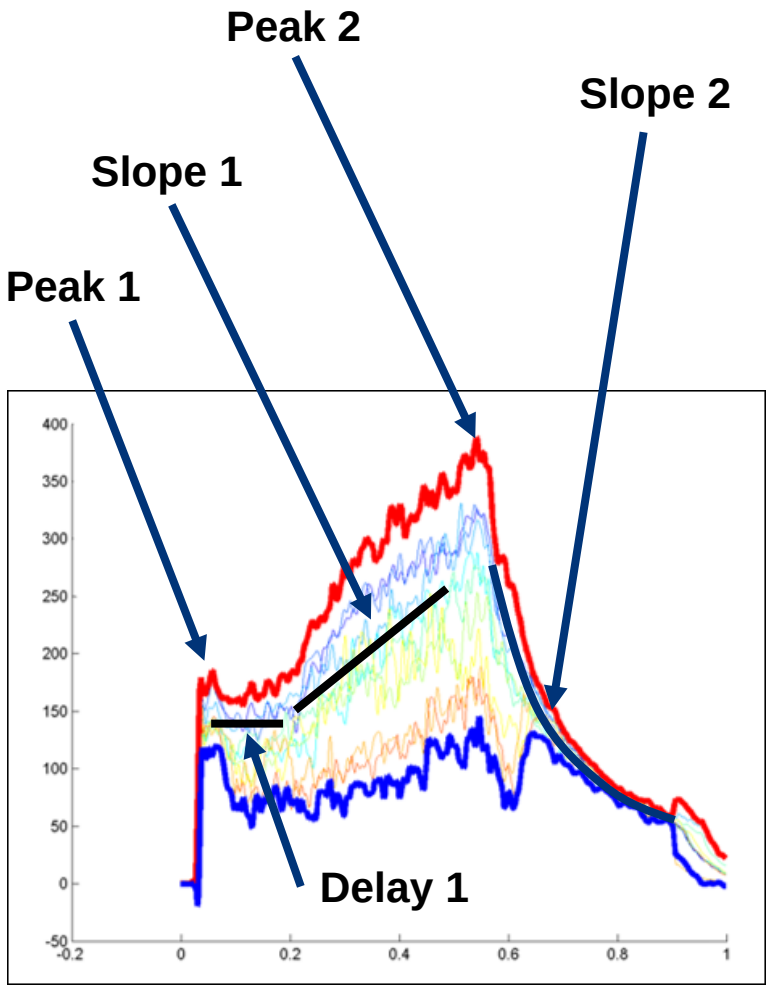




THIS METHOD WAS SIMPLE AND EFFECTIVE

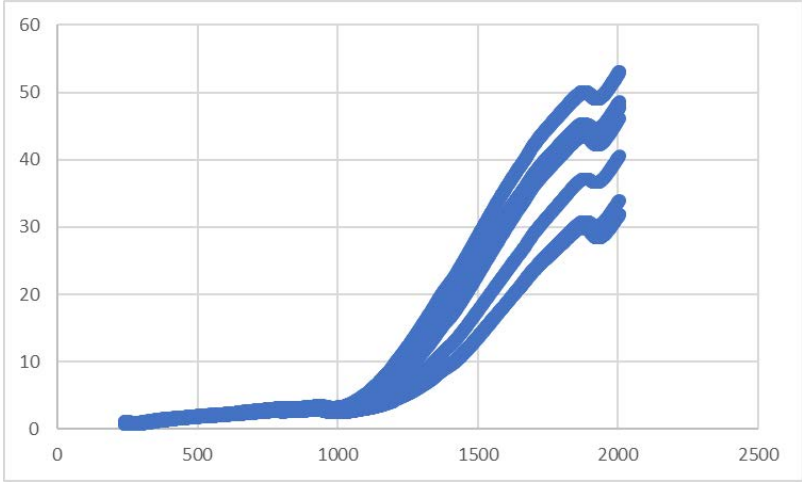
Analysis Method Overview: Data Landmarks

- 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



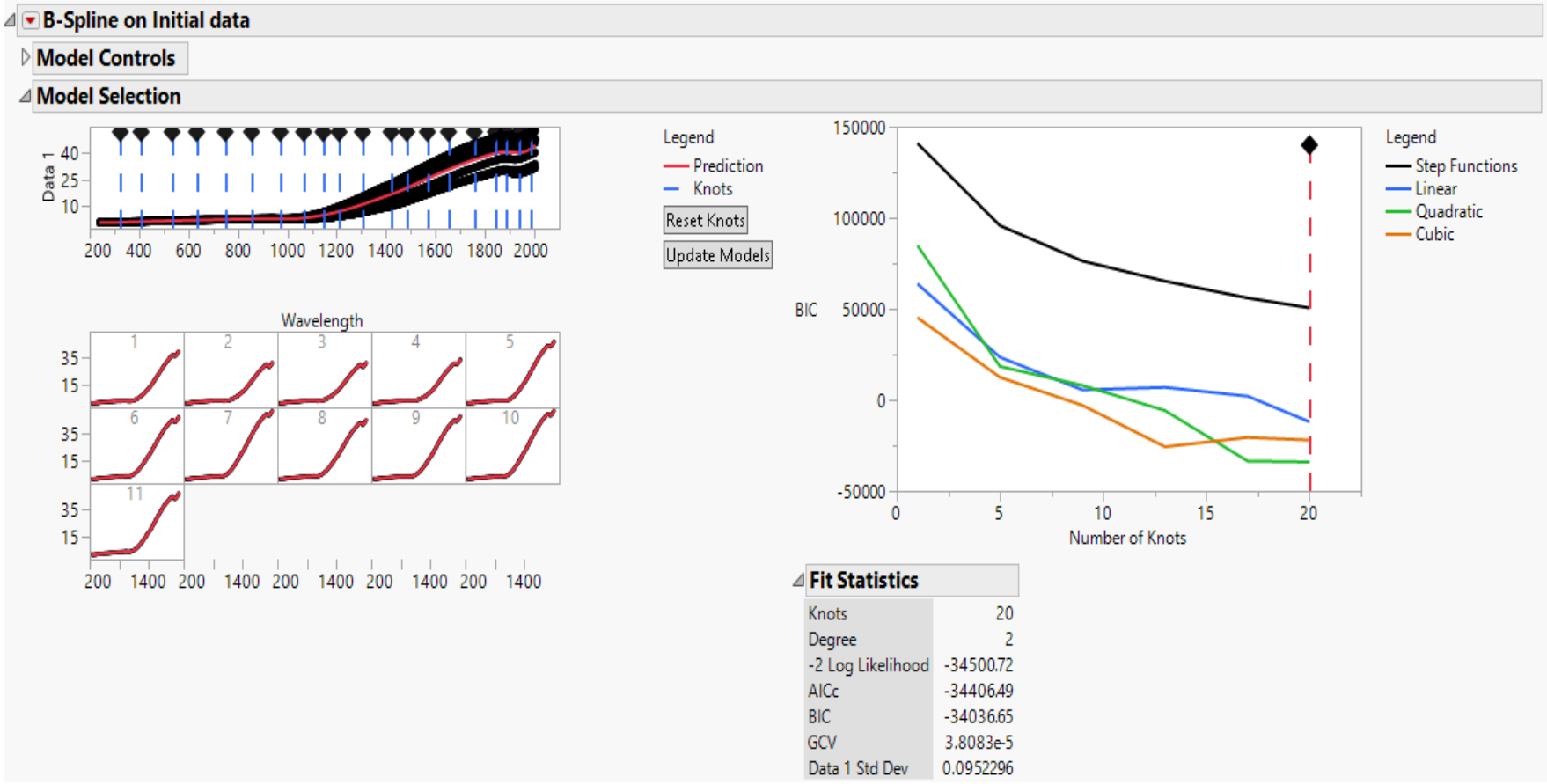
Coating Process FDA Example

- Coating process was generating occasional defects
- Long-time technicians had retired
- Long list of possible root causes
- No one had ever studied process
- Never had a problem before…



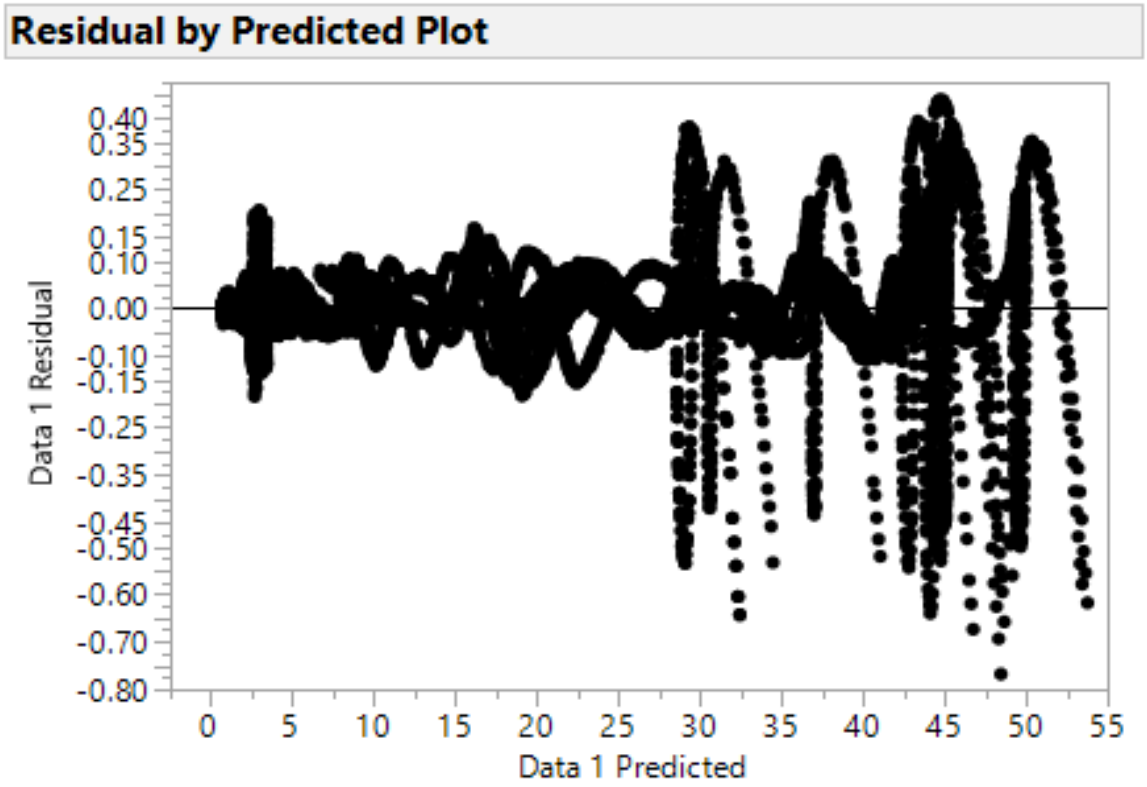
OPTICAL PROPERTIES OF COATING PROCESS

JMP FDA Explorer—Fit Spline Curves

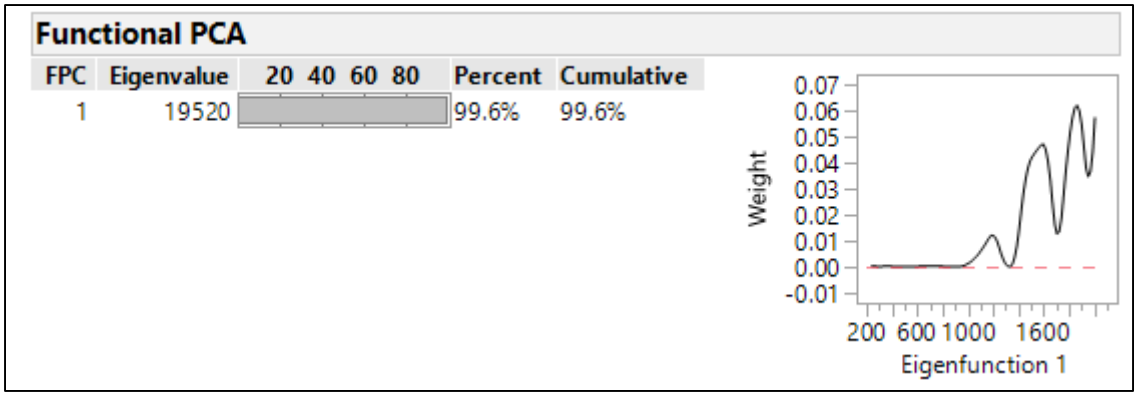


Data was already aligned at zero and adjusted as needed.

JMP FDA Explorer—Check Fit



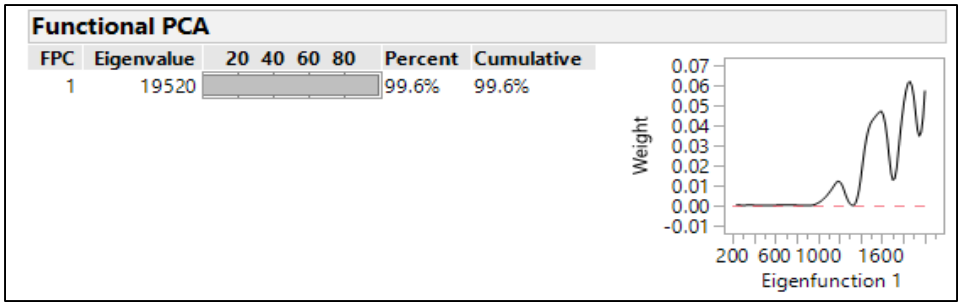
JMP FDA Explorer- Principal Component Analysis



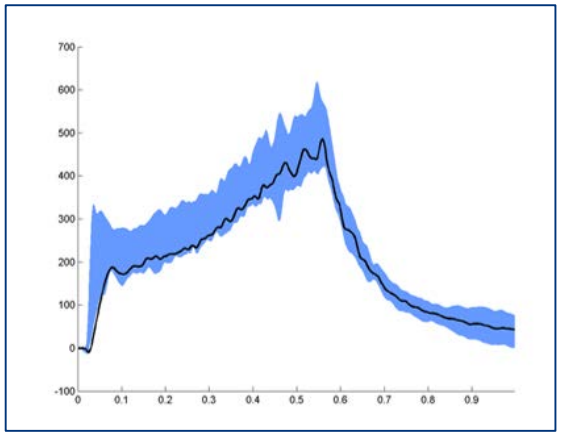
Data Set Number	Data 1 FPC 1
1	-86.2
2	-214.0
3	-214.0
4	-201.2
5	56.8
6	89.6
7	179.1
8	49.8
9	78.9
10	181.3
11	79.8

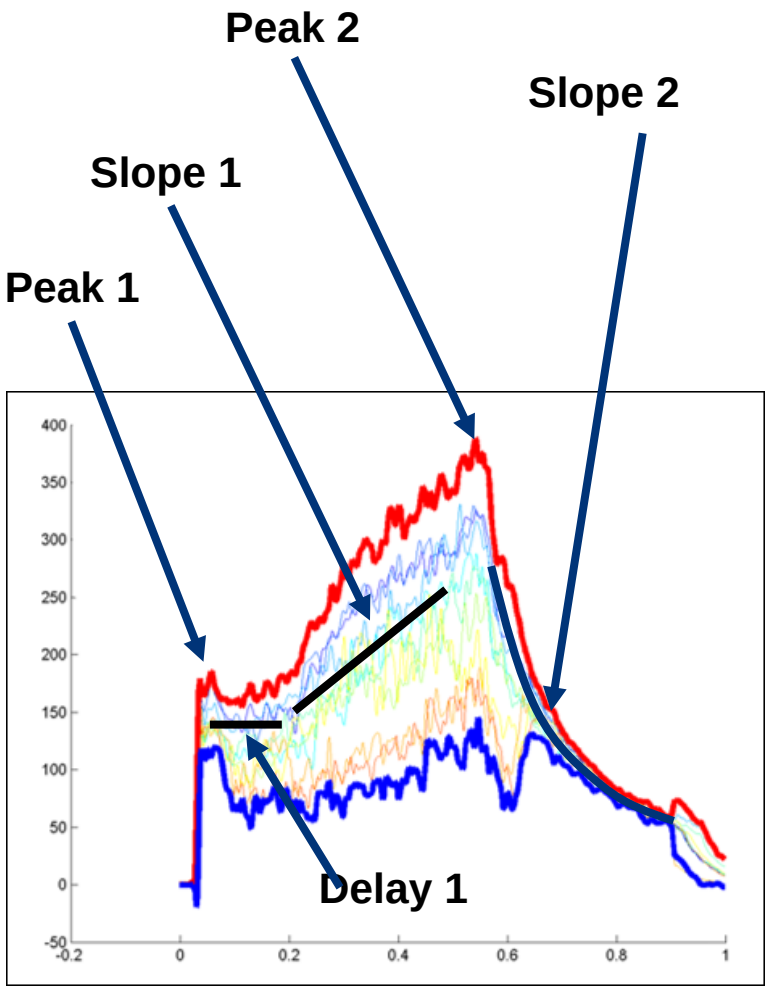
- This data is 99.6% covered by one function
- Each data set is represented by the function and a single multiplier
- Move forward with standard statistical analysis techniques

Three Methods for Functional Data Analysis



Data Set Number	Data 1 FPC 1
1	-86.2
2	-214.0
3	-214.0
4	-201.2
5	56.8
6	89.6
7	179.1
8	49.8
9	78.9
10	181.3
11	79.8





Three options, each with their own advantages

References

Whaley, Dewey Lonzo, "The Interquartile Range: Theory and Estimation." (2005). *Electronic Theses and Dissertations.* Paper 1030. http://dc.etsu.edu/etd/1030

Ramsay, James, Silverman, B.W. "Applied Functional Data Analysis." Springer-Verlag New York. 2002.

https://www.jmp.com/en_us/events/ondemand/mastering-jmp/functional-data-explorer-part-1.html

https://community.jmp.com/t5/US-Federal-Government-JMP-Users/JMP-Federal-Government-User-Resources-Page/ta-p/24133

David R. Harrison
Test Design/Reliability/Data Analysis
David.R.Harrison@lmco.com
785-832-8470 (desk)
303-859-9235 (cell)


LOCKHEED MARTIN