Design of Experiments for Reverse Engineering Formulations

Predicting Chemical Composition from Fewer Test Blends by Modeling Spectra using Functional Data Analysis

Tom Donnelly, PhD, CAP
Principal Systems Engineer
JMP Defense & Aerospace Team
tom.donnelly@jmp.com

Outline

- Why use? & What is Design of Experiments (DOE)?
- What is Functional Data Analysis (FDA)?
- Reverse Engineering Case #1 – Mixture of alcohol blends
- Reverse Engineering Case #2 – Mineral formulations

Let’s Review the Poll

Review of DOE – Why use it?

There's no better way
to get the most information
from the least amount of testing

- Identify important factors when faced with many
- Do sensitivity and trade-space analysis
- Optimize across multiple responses
- Characterize the operating region

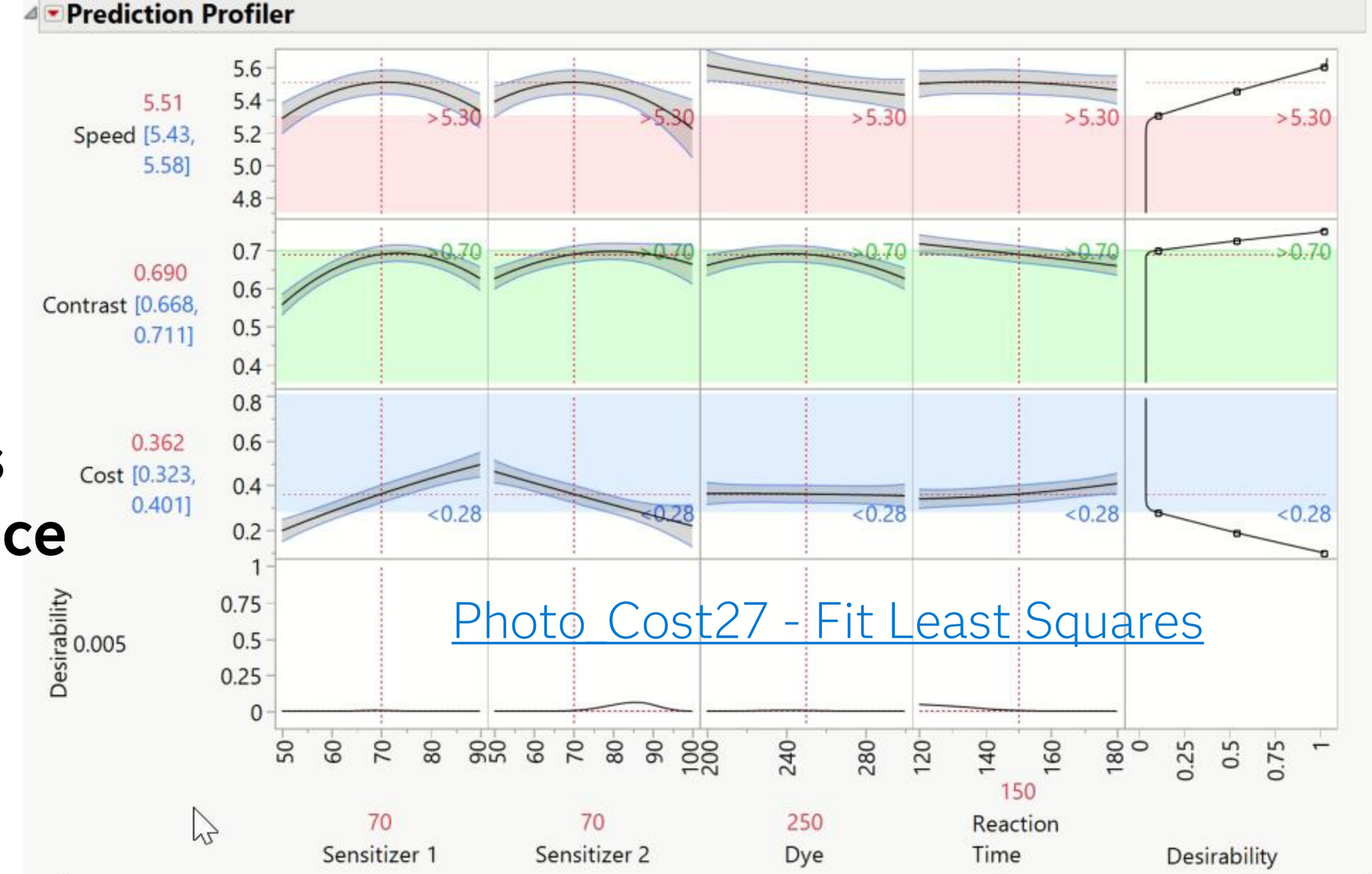
Let’s go to JMP…

Use Prediction Profiler to Answer Questions about the DOE-Characterized/Modeled Process

Back from JMP…

How do you discuss trading off performance and cost with your decision makers?

1. Use JMP
2. Record in PPT
3. Interactive HTML
   (CI & PI - new in JMP 18)

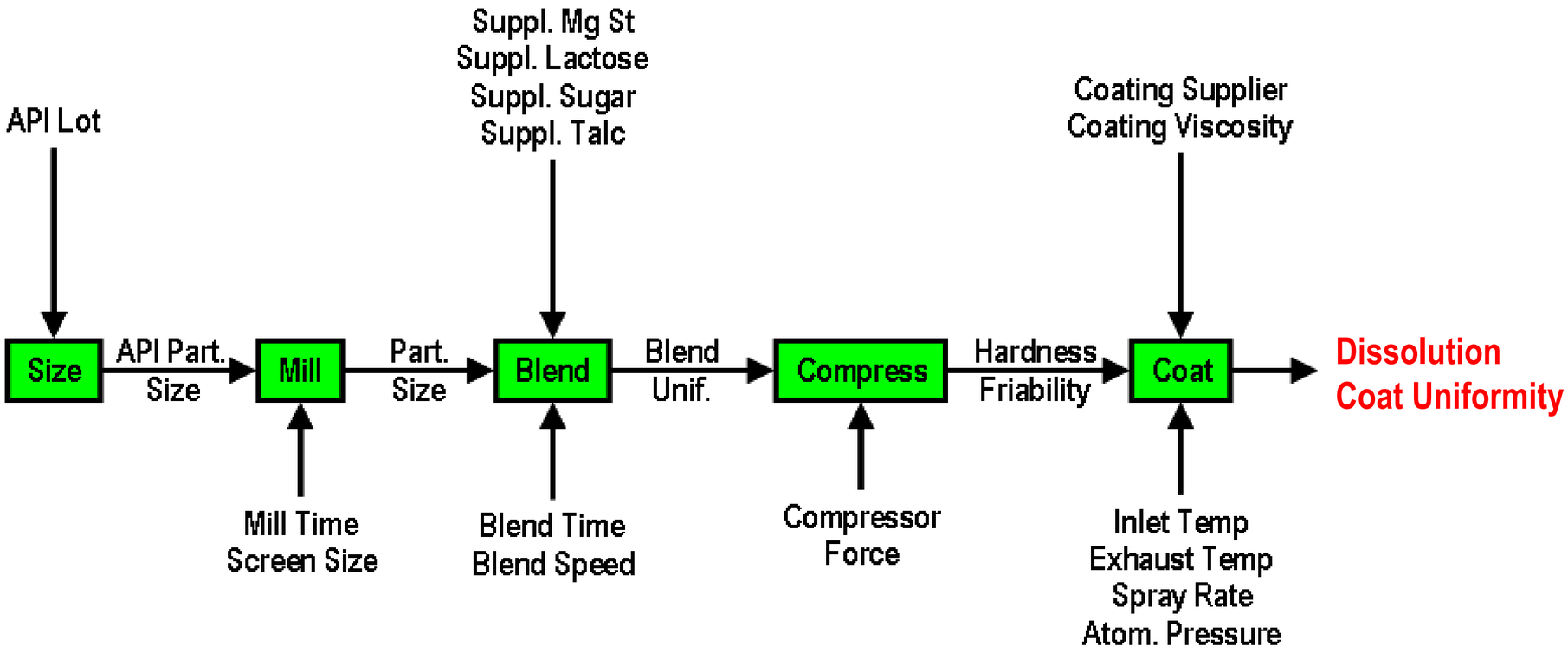


Remembered Settings

Setting	Sensitizer 1	Sensitizer 2	Dye	Reaction Time	Speed	Contrast	Cost	Desirability
Equal Importance Opt	80.8	91.3	251	120	5.35	0.747	0.250	0.348
Mid Point Settings	70.0	70.0	250	150	5.51	0.690	0.362	0.005
Cost 6X Speed & Contrast	84.0	93.7	283	120	5.29	0.725	0.199	0.214
Opt Spd3X-Cntr1X-Cost6X	82.0	90.7	287	120	5.33	0.718	0.221	0.264

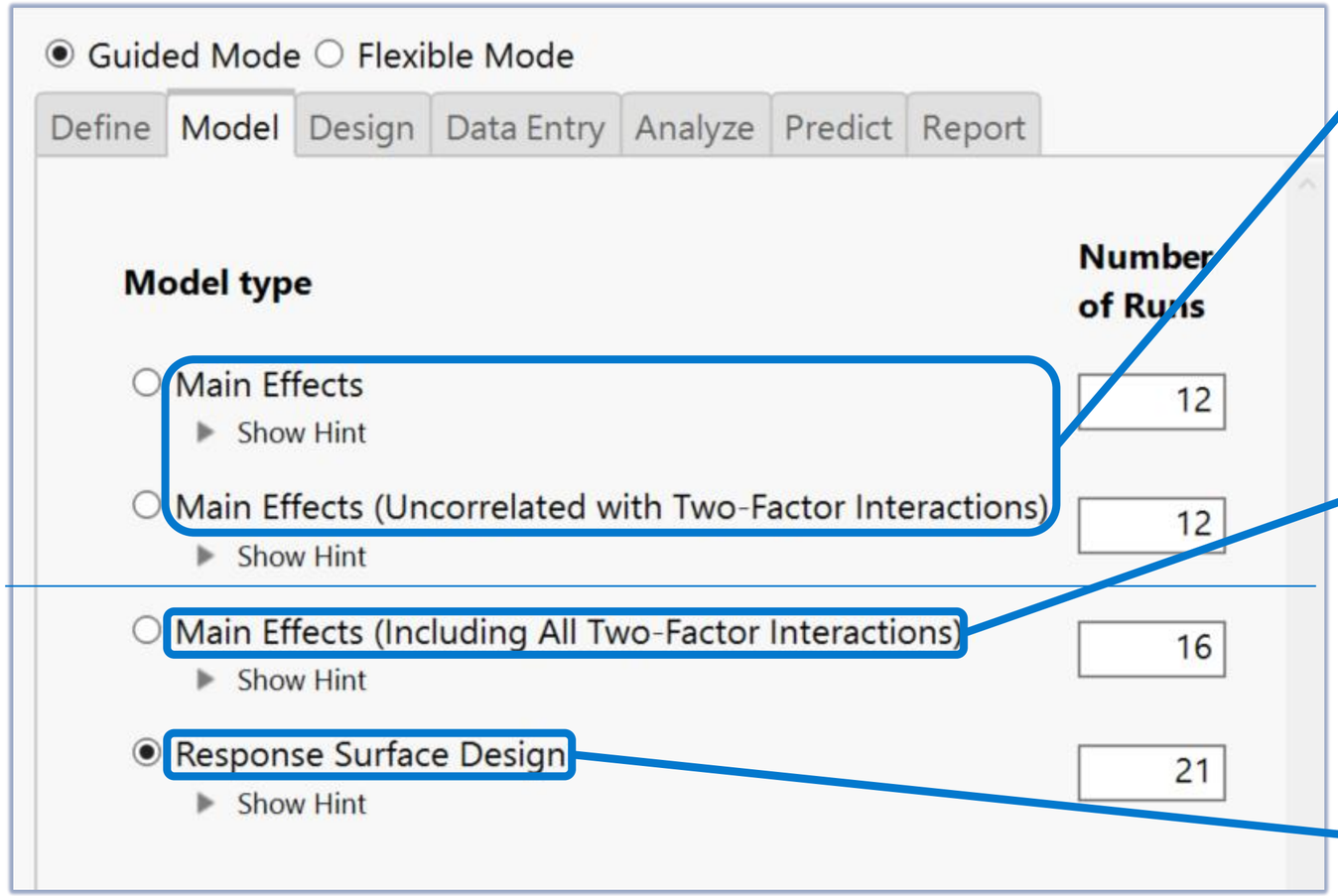
Review of DOE - Classic Definition

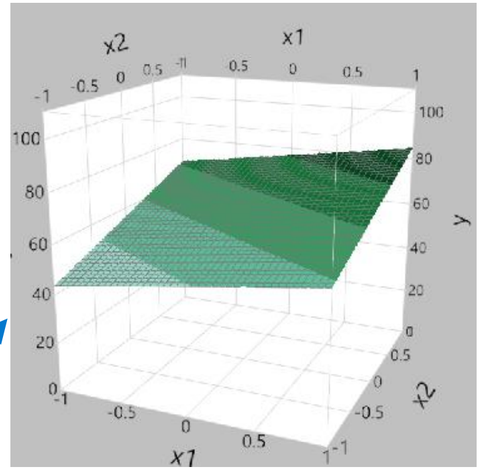
Purposeful control of the inputs (factors) in such a way as to deduce their relationships (if any) with the outputs (responses).

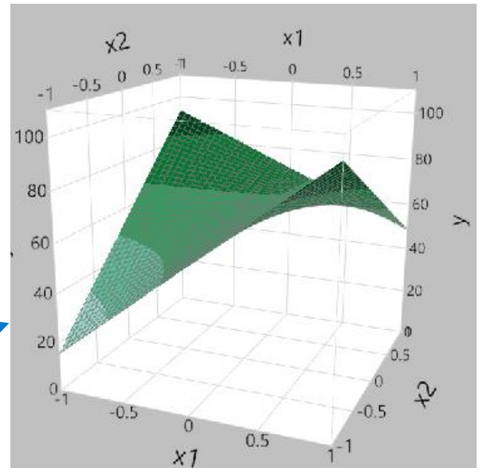


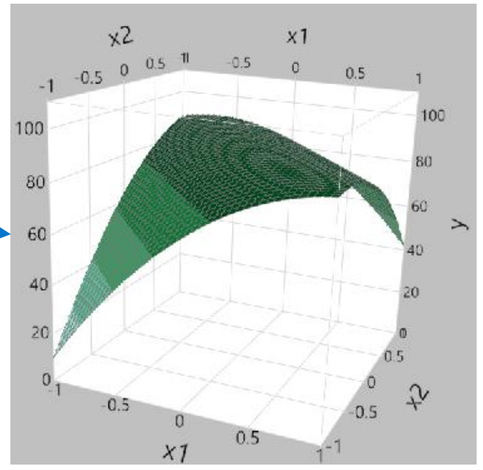
Alternative Definition of DOE

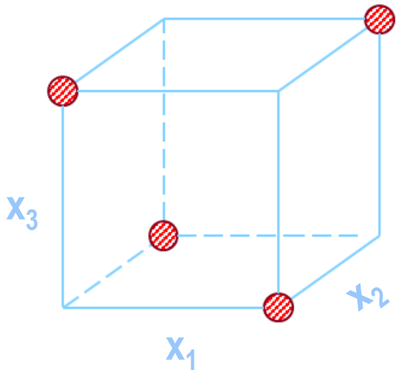
A DOE is a specific collection of trials run to support a *proposed* model.

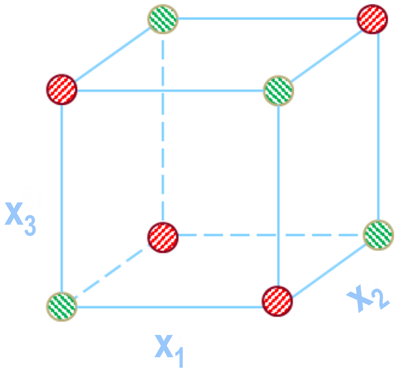


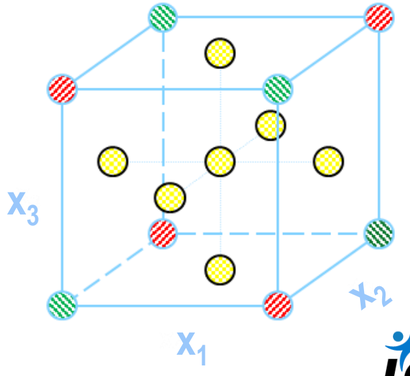










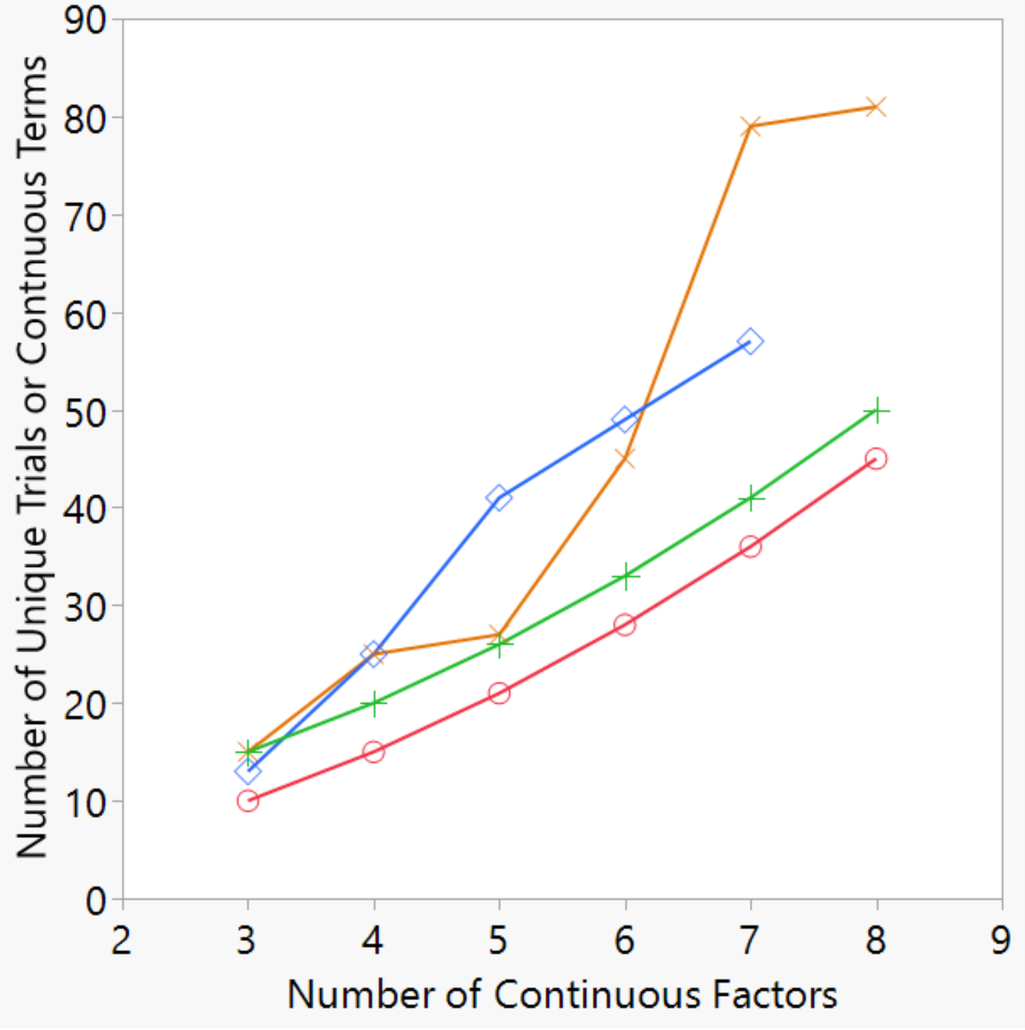


Real-World Design Issues

Model-Driven Custom DOE Makes Design Fit your Problem – NOT your Problem Fit the Design!

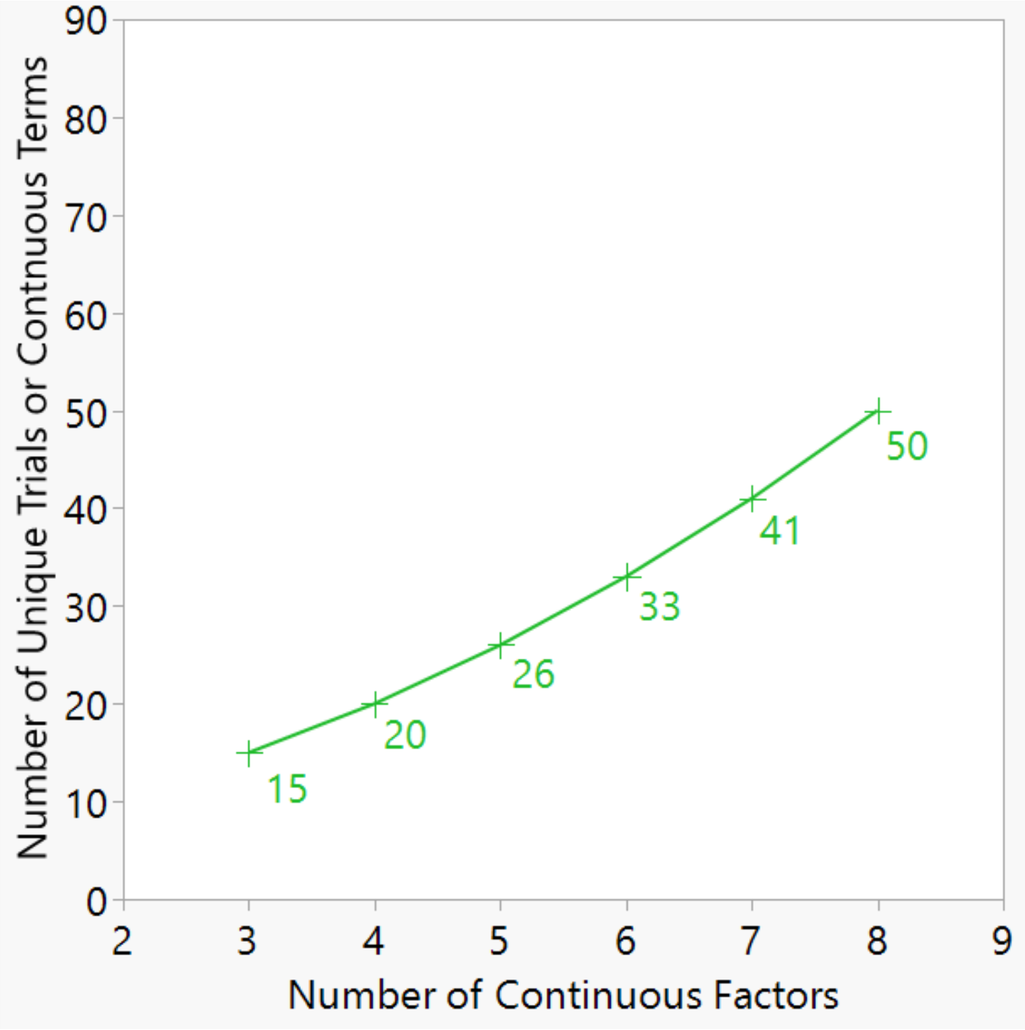
- Work with different kinds of control variables/factors:
  - Continuous/quantitative
  - Categorical/qualitative
  - Mixture/formulation
  - Blocking
- Work with combinations of these four kinds of variables
- Certain factors are hard-to-change
- Certain factor combinations cannot be run
- Want to add onto existing trials
- Need to repair broken design

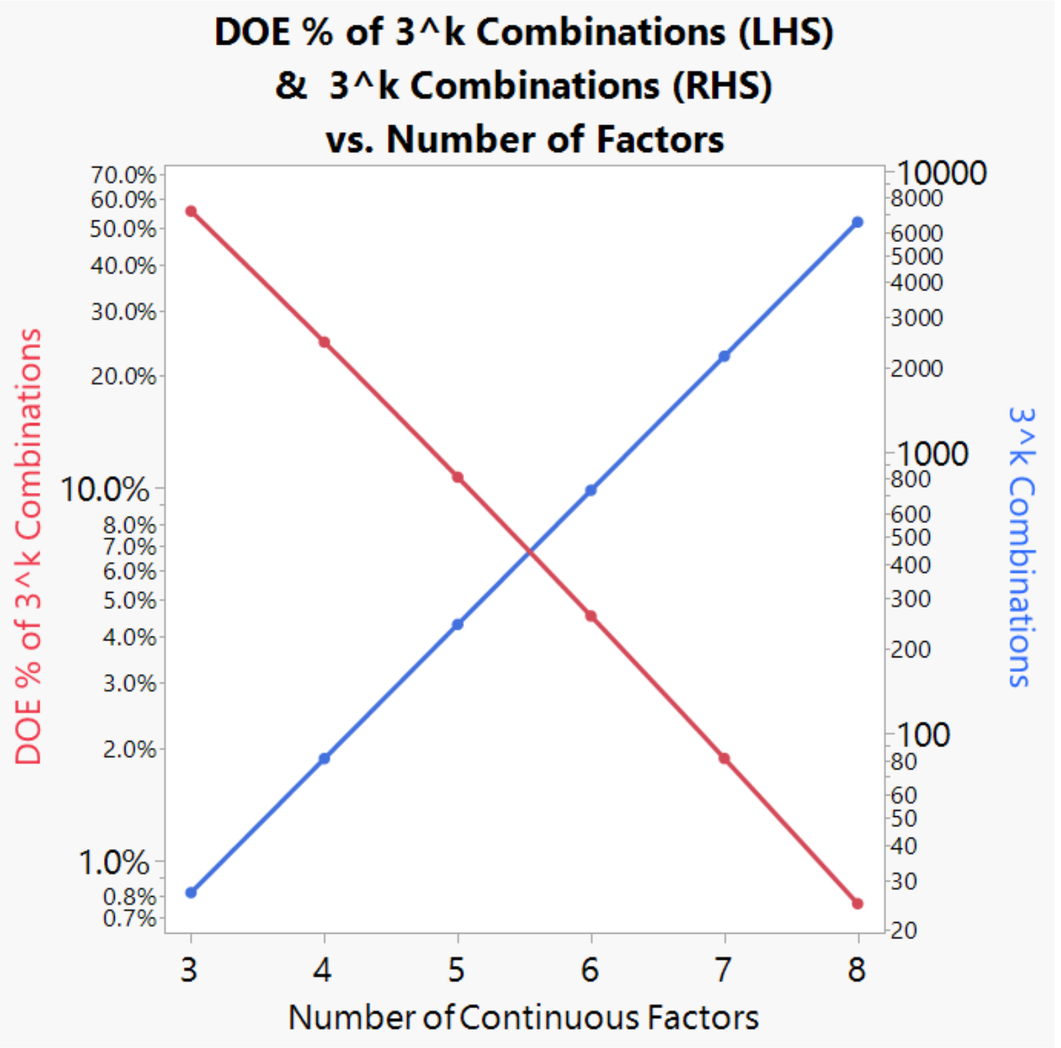
Number Unique Trials & Number Quadratic Model Terms vs. Number Continuous Factors

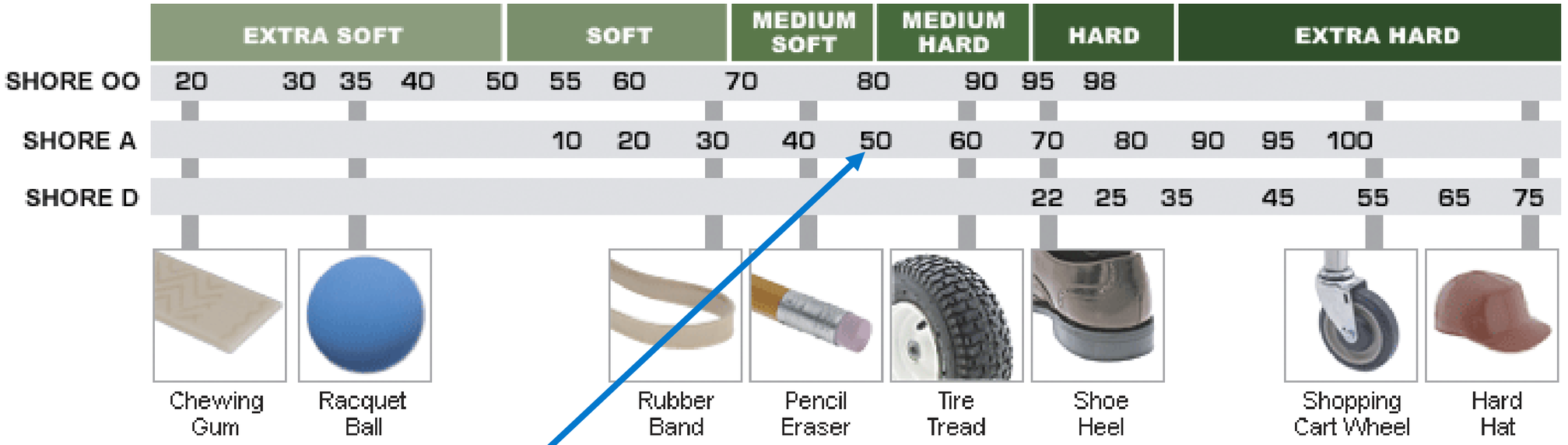


- Unique Trials in Central Composite Design
- Unique Trials in Box-Behnken Design
- Unique Trials in Custom Design with 6 df for Model Error
- Terms in Quadratic Model

Number of Custom DOE Trials Rises Slowly
Number of Possible Trial Combinations Rises Rapidly (3^k)
DOE Trials as a Percentage of All Combinations Falls Rapidly

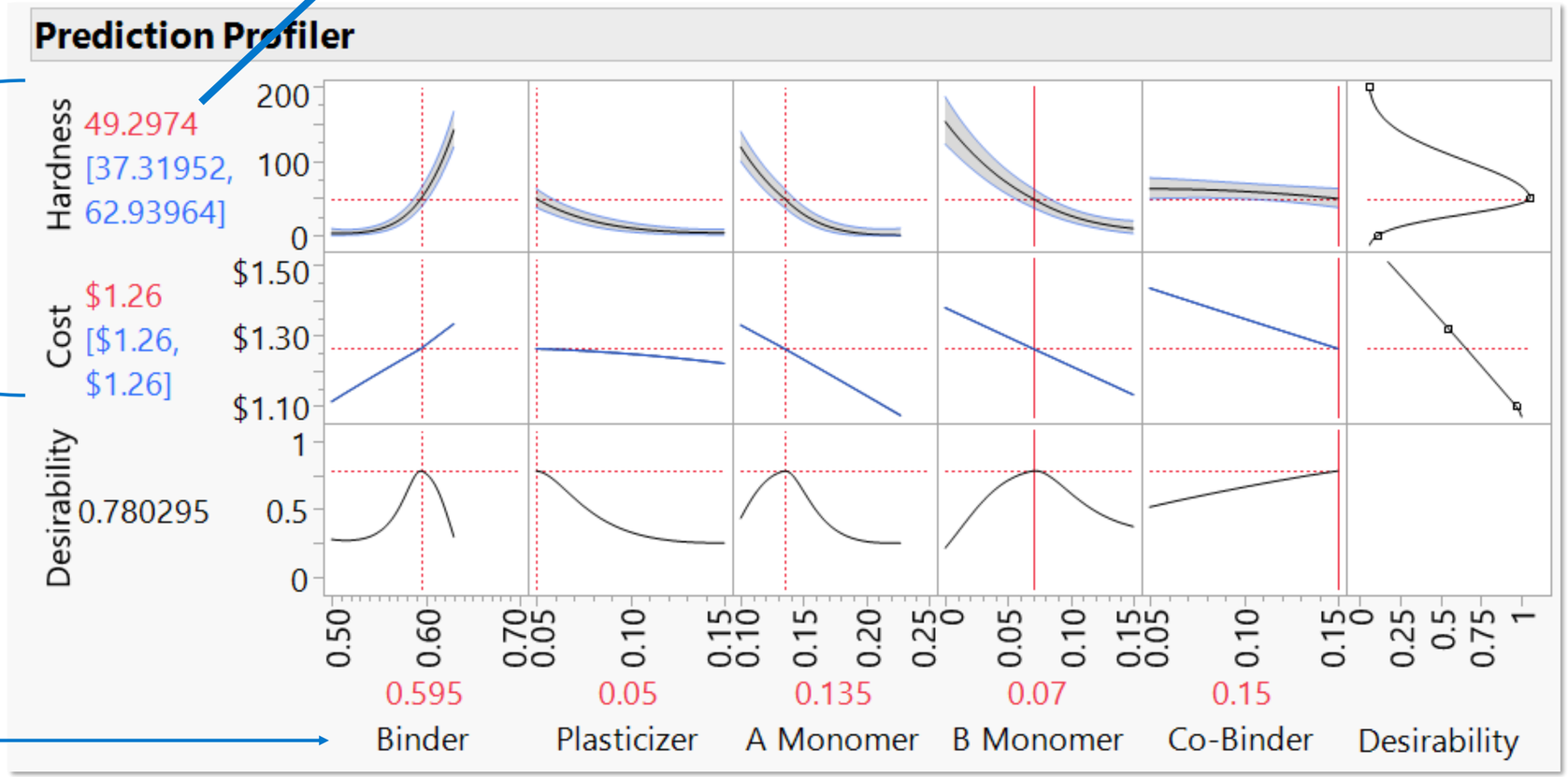






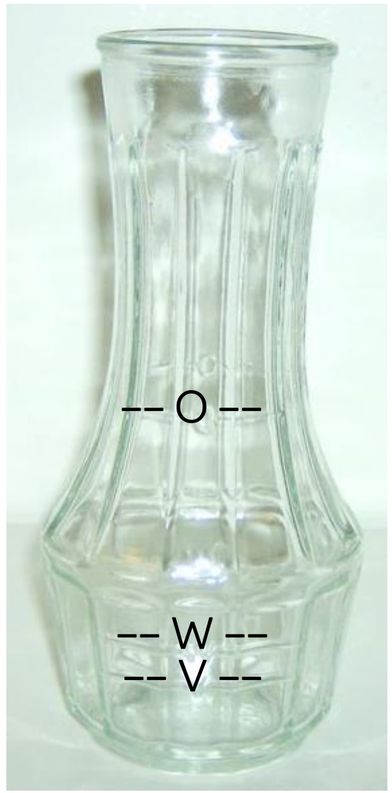
Mixture DOE

2 Responses & 5 Formulation Components



Mixture or Formulation DOE

Making salad dressing is a simple 3-component blend of *Oil*, *Water*, & *Vinegar*



100.0%

37.5%

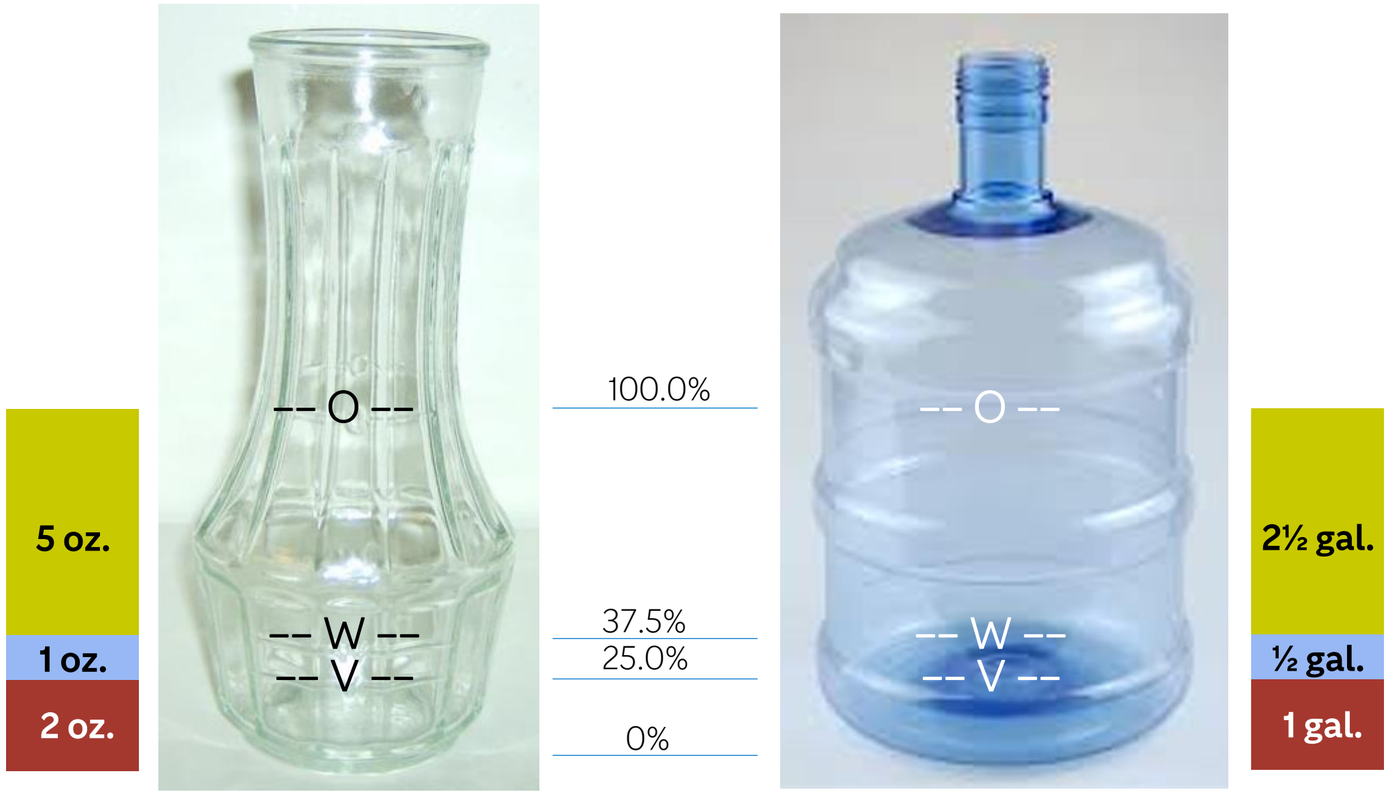
25.0%

0%

- Response depends on **proportions** not quantity.
- Sum of proportions equals 1. This ***constraint*** is what makes mixture DOE different.

Mixture or Formulation DOE

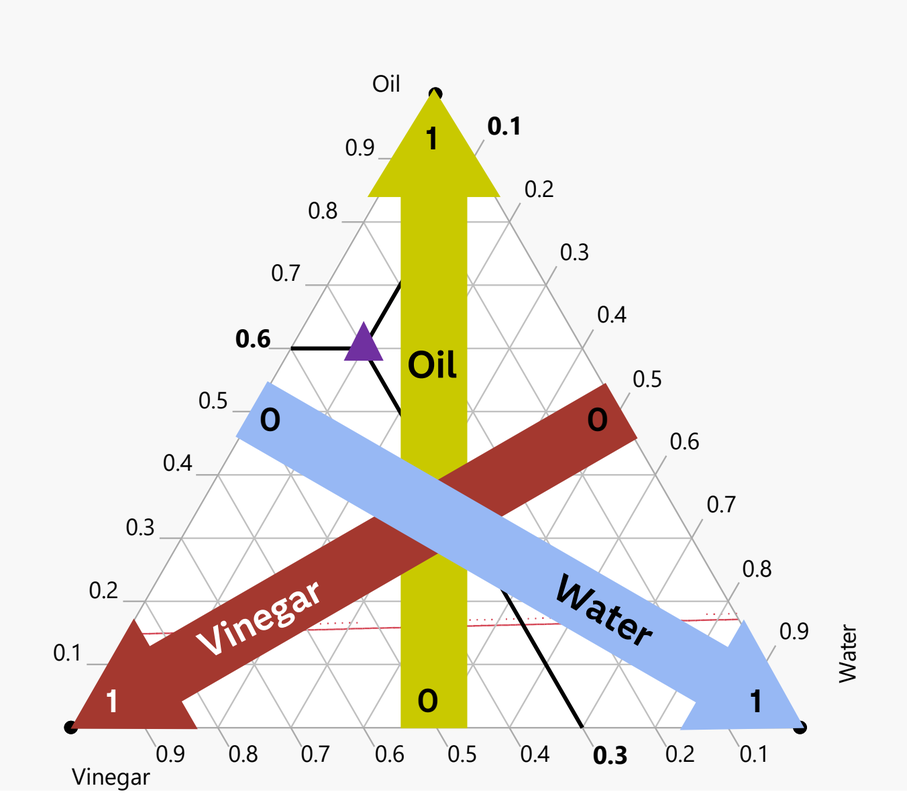
Making salad dressing is a simple 3-component blend of *Oil*, *Water*, & *Vinegar*

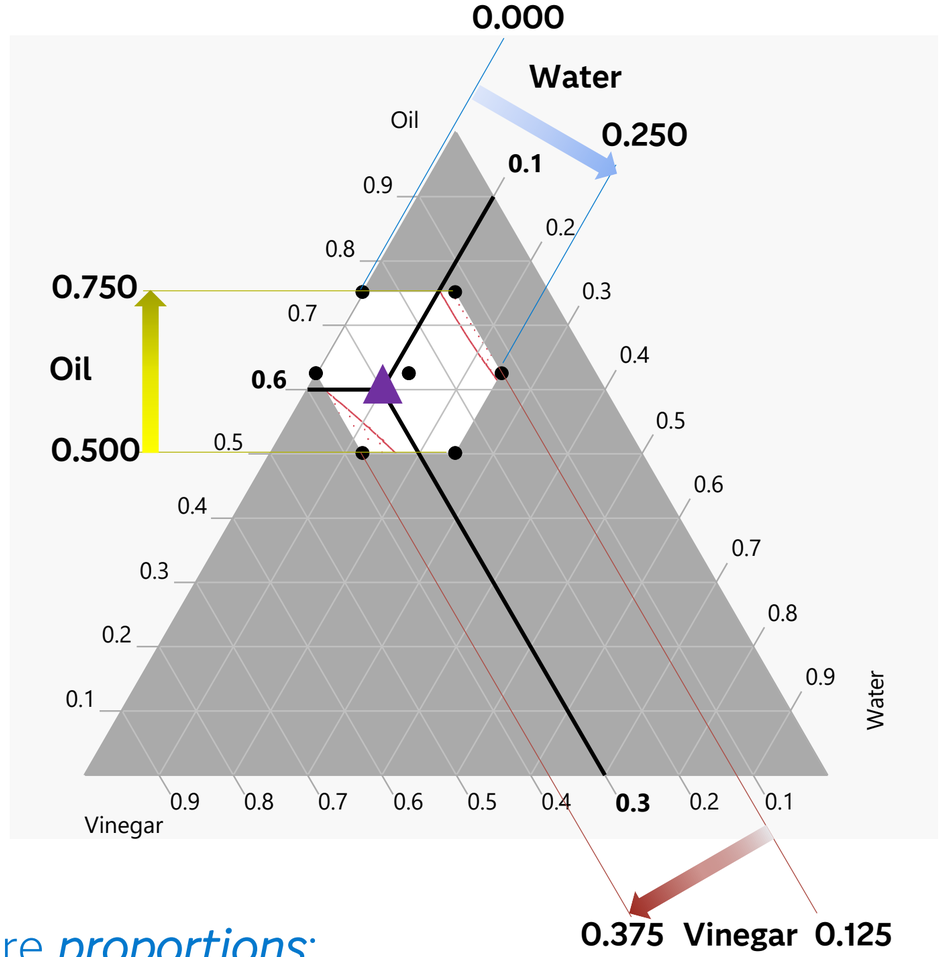


- Response depends on **proportions** not quantity.
- Sum of proportions equals 1. This *constraint* is what makes mixture DOE different.

Use Ternary Plots to Visualize the Mixture Constraint

$$O + W + V = 1$$





Mixture components in a DOE use ranges that are *proportions*:

O:	0.500 to 0.750
W:	0.000 to 0.250
V:	0.125 to 0.375

If Oil = 0.6 and Vinegar = 0.3, then

Water = 1 – (0.6 + 0.3) = 0.1 (See ▲)

Want to something more complex than salad dressing?

Design of Experiments for Complex Biochemical Systems

Presenters

Matthew Lux

John Davies

David Garcia

Cell-free expression (CFE) systems are a suite of methods that reconstitute complex cellular functions like transcription, translation, and metabolism outside the confines of a living cell. CFE systems have numerous biotechnological uses in sensing, biomanufacturing, medicine, basic research, and education. Most CFE systems are made from combining cellular lysates with a complex blend of excipients that improve activity. While the number of excipients makes exploring the combinatorial spaces challenging, high-throughput experimentation with acoustic liquid handling makes it feasible to optimize formulations if paired with an appropriate statistical framework.

Here we describe our use of design of experiments (DOE) to optimize excipient combinations for specific use cases of CFE. We pair our DOE with functional data analysis (FDA) to collapse activity over time measurements to metrics readily used for analysis. Initial formulation DOE examples range from five to 14 components. We further describe our efforts to push to higher scales, attempting mixture-process DOE designs with as many as 42 components using an experimental set-up that allows for 1,536 formulations to be tested at once.

Review of Functional Data Analysis (FDA)

Functional Data Analysis (FDA) is a branch of statistics that analyzes data providing information about curves, surfaces, or anything else varying over a continuum.

Traditional Rectangular Data

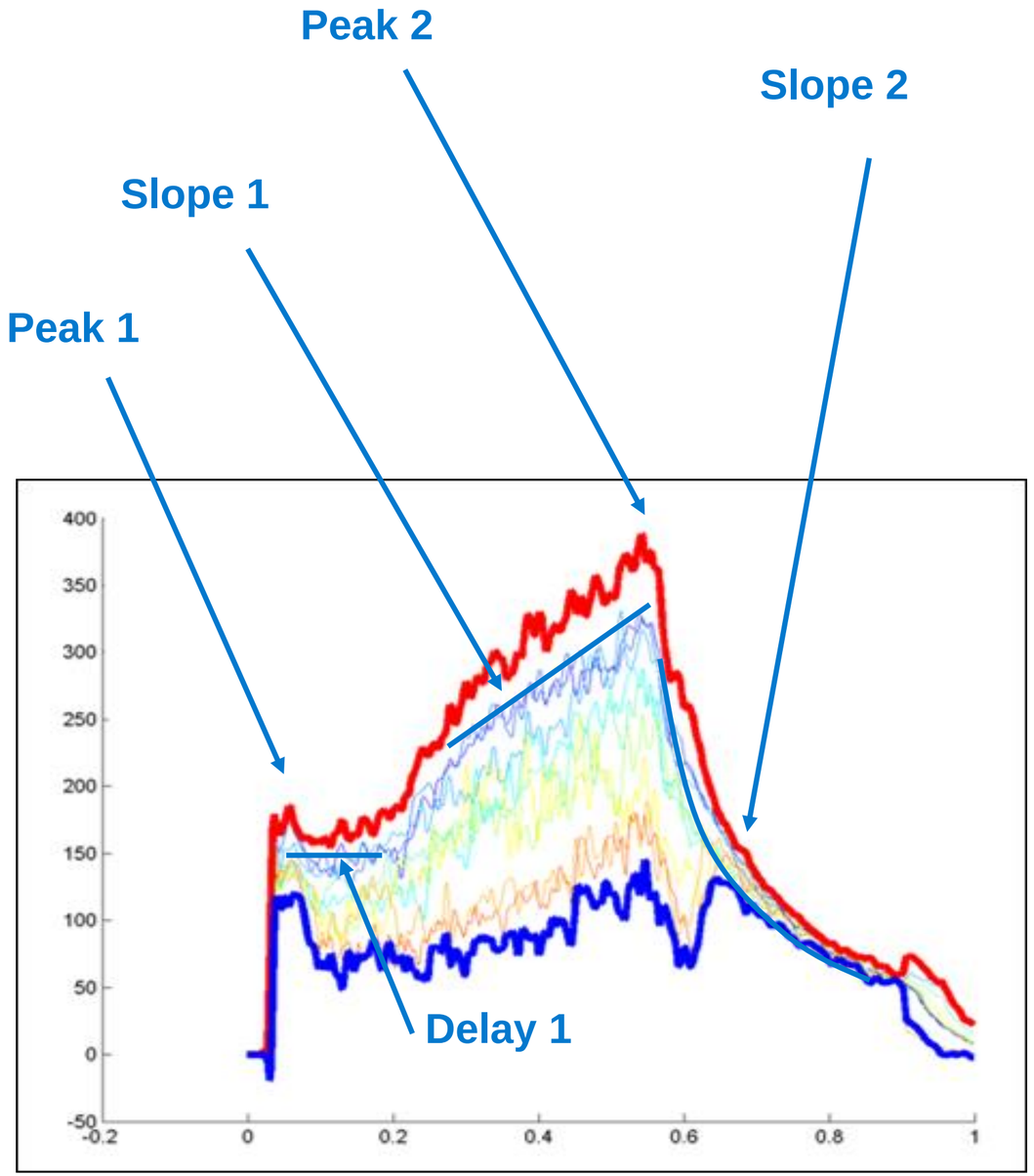
	Batch	X1	X2	Y
1	001	1.00	1.00	2.17
2	002	0.94	1.01	0.00
3	003	1.06	1.01	2.70
4	004	0.94	0.99	0.26
5	005	1.06	0.99	2.87
6	006	1.00	1.00	1.97

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

Landmark Analysis



Moderately effective non-FDA option, but NOT as good as using FDA

Does not use information from the whole curve

Based on slide from David Harrison of Lockheed Martin Corporation

FDA Examples

- Sensor streams
- Measurements taken over a range
- Vibration signals
- **Spectral data**
- Radar/sonar signatures
- Trajectories of flights between cities
- Electrocardiograms (EKGs)

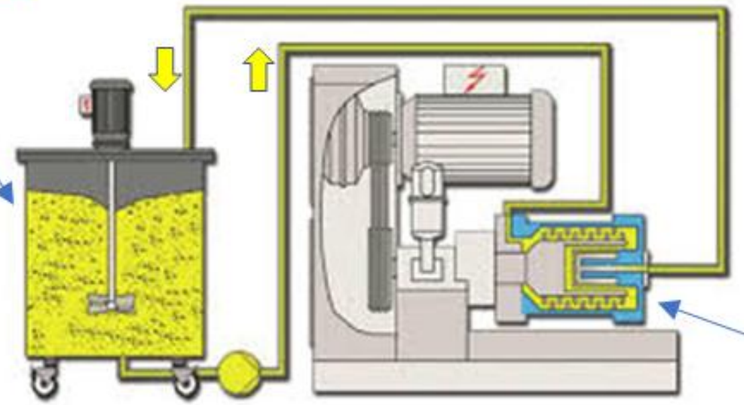
    - *…and more!*

- *Almost any response in a longitudinal order*

FDA-DOE Example

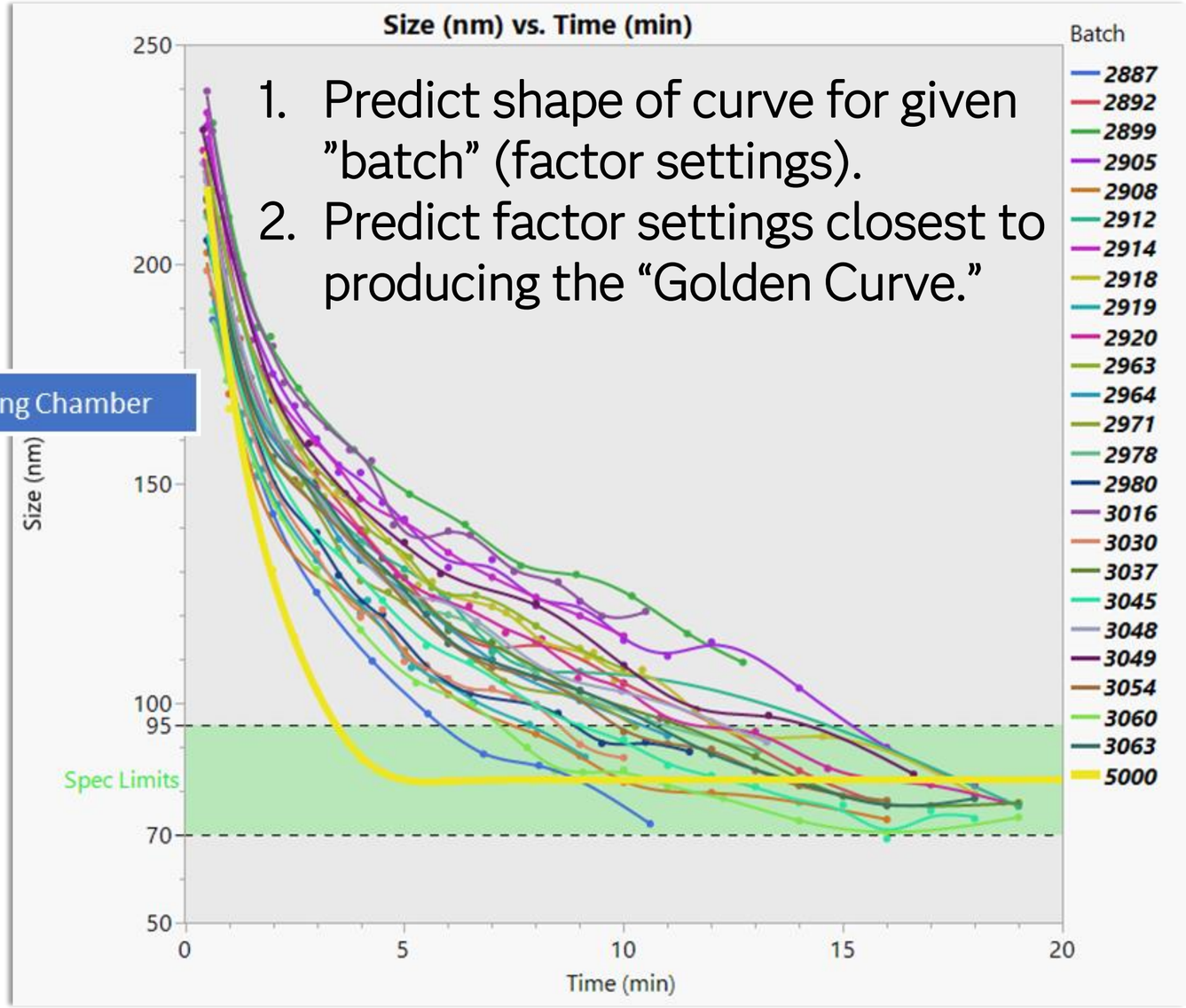
Holding Vessel

https://www.netzsch.com



Milling Chamber

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



Functional Data Analysis – DOE (5-min)

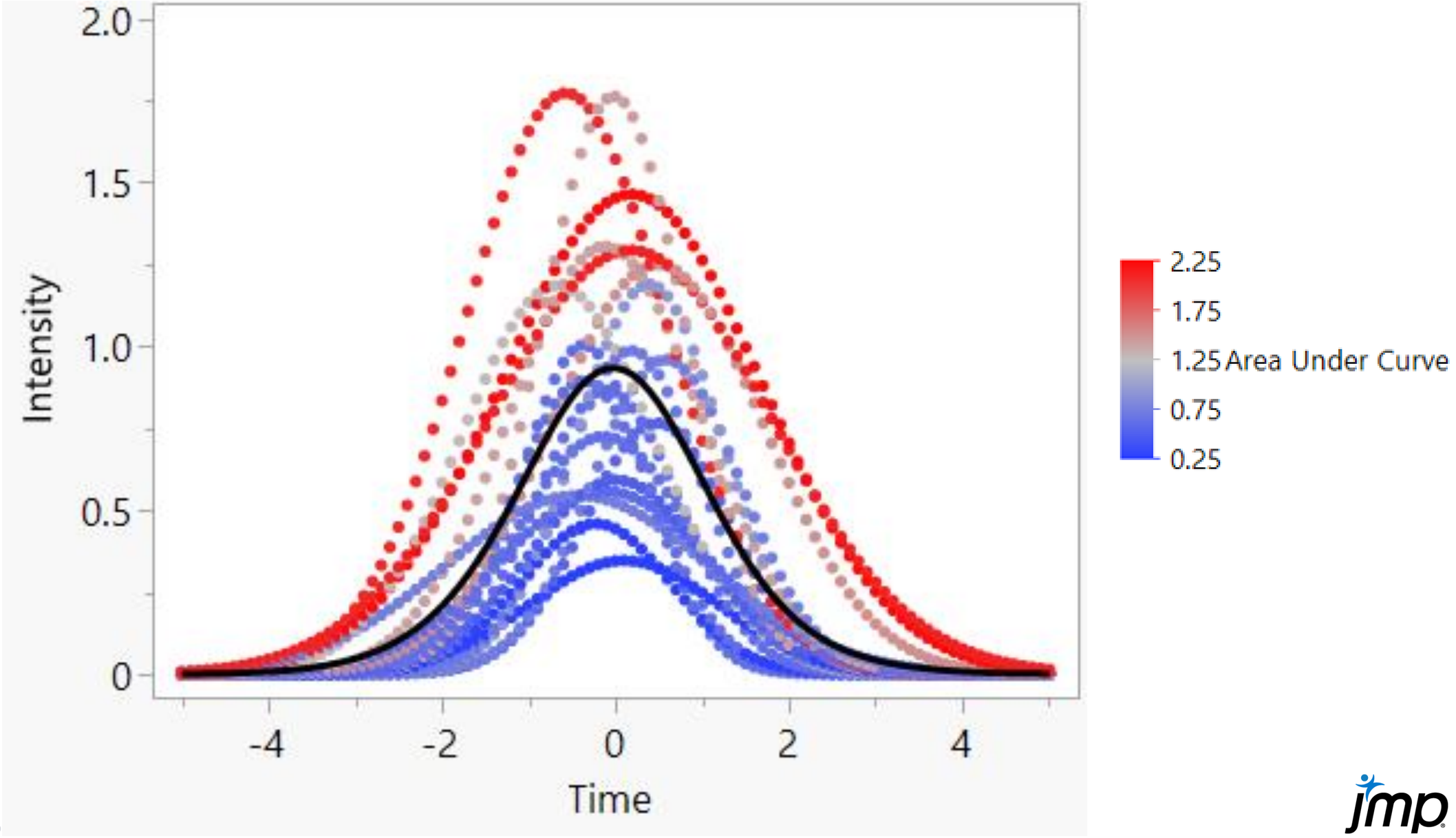
FUNCTIONAL DATA ANALYSIS

- Use Functional Principal Components Analysis (FPCA) to do dimension reduction and break "curve" data into *FPC Scores* & *Shape Components*
  - FPC Scores explain *function-to-function variation*
  - Shape Components explain the *longitudinal variation* (e.g., *time*, *distance*, *frequency*, or *wavelength*)

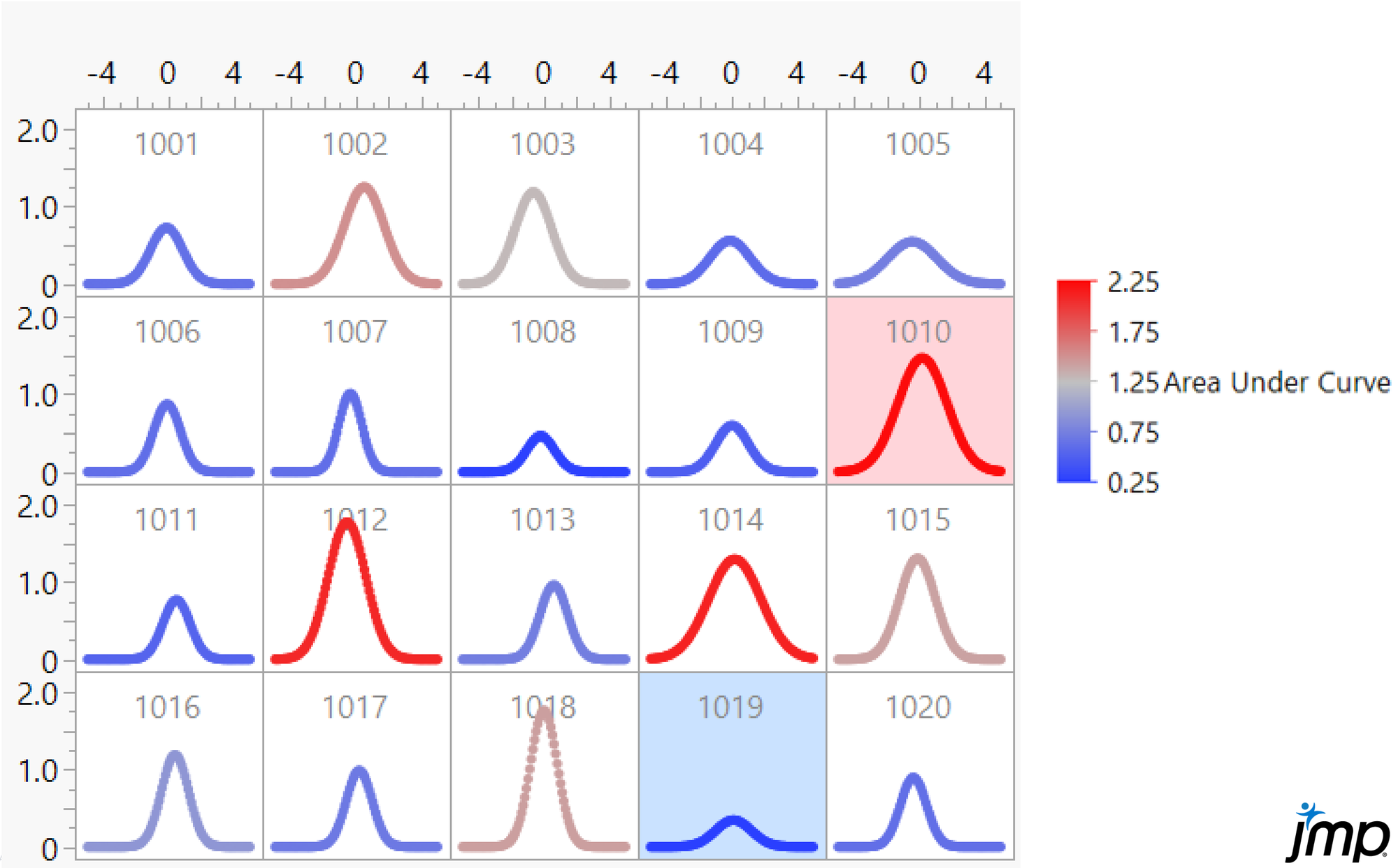
- Fit models with FPC scores, cluster and graph them - *just like any other continuous data*

- *Model the FPC scores as functions of the DOE factors*

FUNCTIONAL DATA ANALYSIS

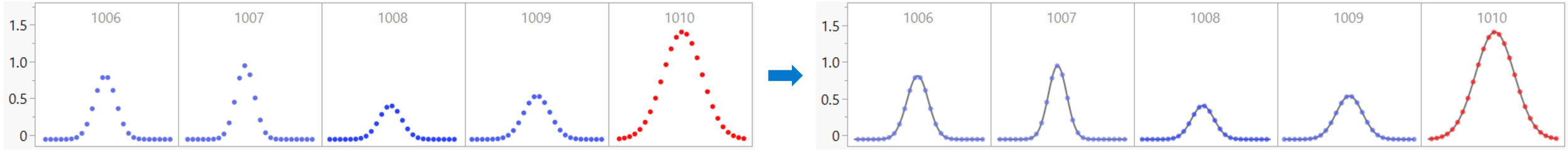


FUNCTIONAL DATA ANALYSIS

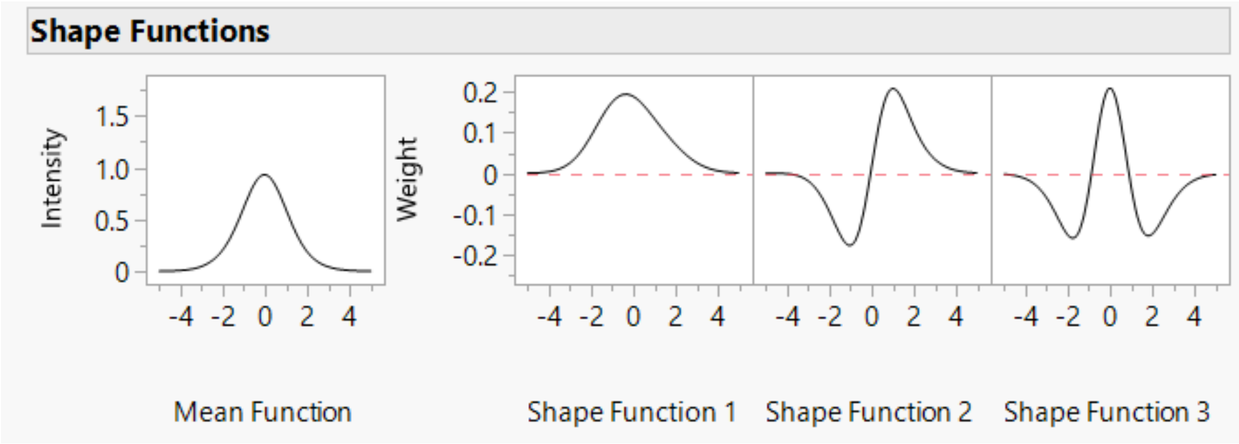


FUNCTIONAL DATA ANALYSIS

1. Convert streams of data into functions - “smoothly connect the dots.”



2. Do FPCA to create
   a) Shape Functions to explain the longitudinal variation.
   b) FPC scores to explain function-to-function variation.



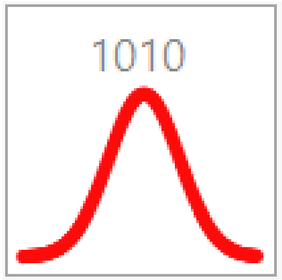
Function Summaries

Batch	FPC 1	FPC 2	FPC 3
1006	-0.97	-0.30	0.48
1007	-0.86	-1.13	0.54
1008	-2.33	-0.48	-0.22
1009	-1.61	-0.07	-0.15
1010	3.42	1.05	-0.76

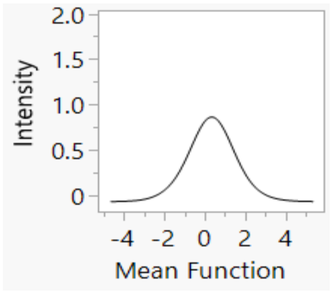
FUNCTIONAL DATA ANALYSIS

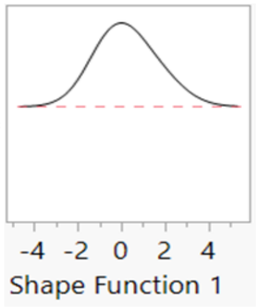
Batch	FPC 1	FPC 2	FPC 3
1001	-1.08	-0.32	0.01
1002	1.39	1.77	-0.29
1003	1.02	-1.86	-0.44
1004	-1.42	-0.24	-0.48
1005	-1.20	-0.56	-0.91
1006	-0.97	-0.30	0.48
1007	-0.86	-1.13	0.54
1008	-2.33	-0.48	-0.22
1009	-1.61	-0.07	-0.15
1010	3.42	1.05	-0.76
1011	-1.53	0.99	-0.00
1012	3.78	-2.15	-0.10
1013	-0.91	1.57	-0.01
1014	2.80	0.88	-1.07
1015	1.63	-0.21	0.40
1016	-0.07	1.45	0.64
1017	-0.75	0.55	0.67
1018	2.14	0.15	1.90
1019	-2.53	-0.08	-0.51
1020	-0.91	-1.03	0.29

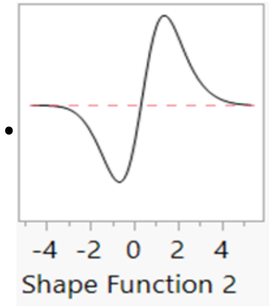
3. Products of FPC scores multiplying their corresponding shape functions, when added to the Mean shape closely reproduce the original (batch) raw function curves.

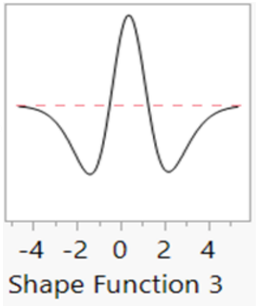


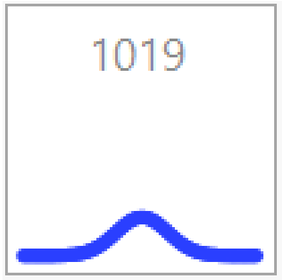
$+3.42$ $+1.05 \cdot$ $-0.76 \cdot$

$=$ 

$+FPC1 \cdot$ 

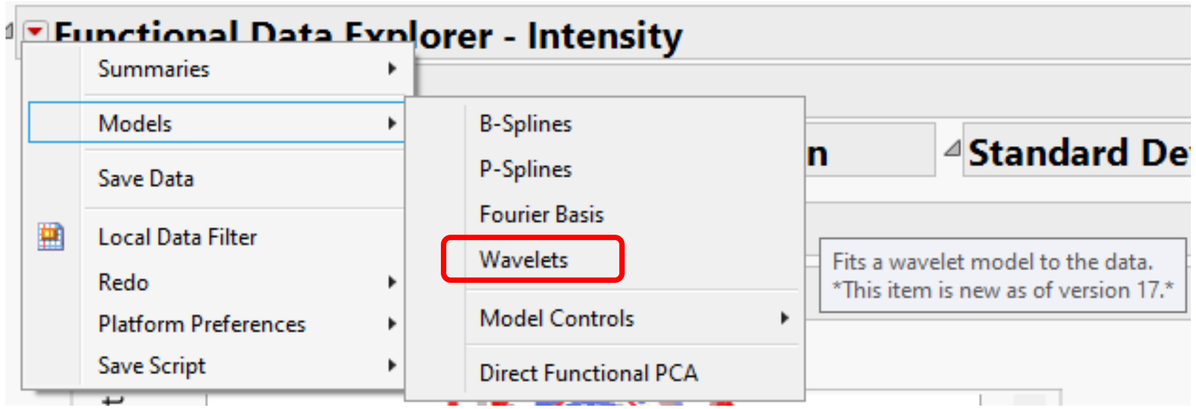
$+FPC2 \cdot$ 

$+FPC3 \cdot$ 



-2.53 $-0.08 \cdot$ $-0.51 \cdot$

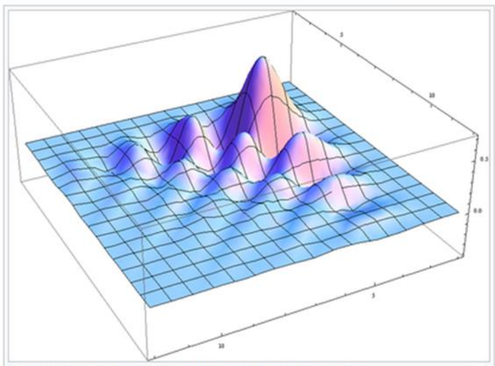
Wavelets often Outperform Splines & Fourier Basis Functions



The **Daubechies wavelets**, based on the work of Ingrid Daubechies, are a family of orthogonal wavelets defining a discrete wavelet transform and characterized by a maximal number of vanishing moments for some given support. With each wavelet type of this class, there is a scaling function (called the *father wavelet*) which generates an orthogonal multiresolution analysis.

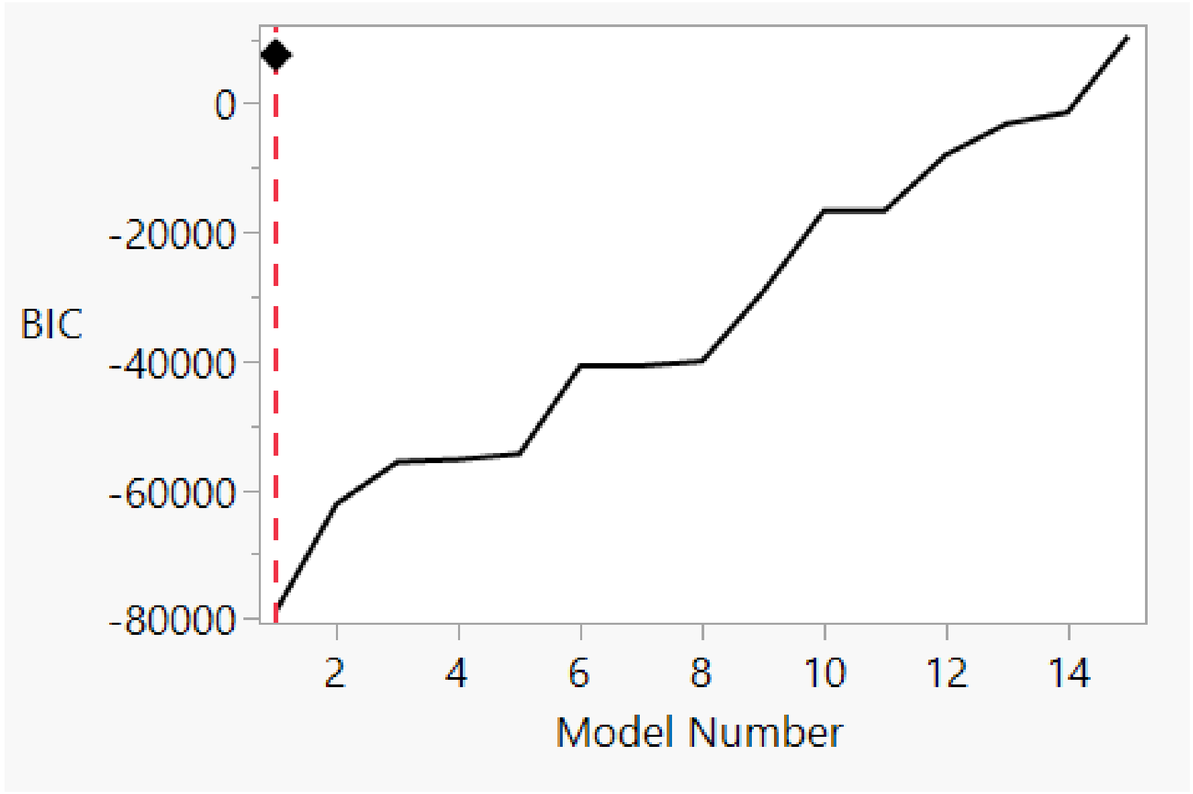
Properties [edit]

In general the Daubechies wavelets are chosen to have the highest number *A* of vanishing moments, (this does not imply the best smoothness) for given support width (number of coefficients) 2*A*.[1] There are two naming schemes in use, D*N* using the length or number of taps, and db*A* referring to the number of vanishing moments. So D4 and db2 are the same wavelet transform.



Daubechies 20 2-d wavelet (Wavelet Fn X Scaling Fn)

A **wavelet** is a wave like oscillation with an amplitude that begins at zero, increases or decreases, and then returns to zero one or more times. They are used to approximate more complex functions.



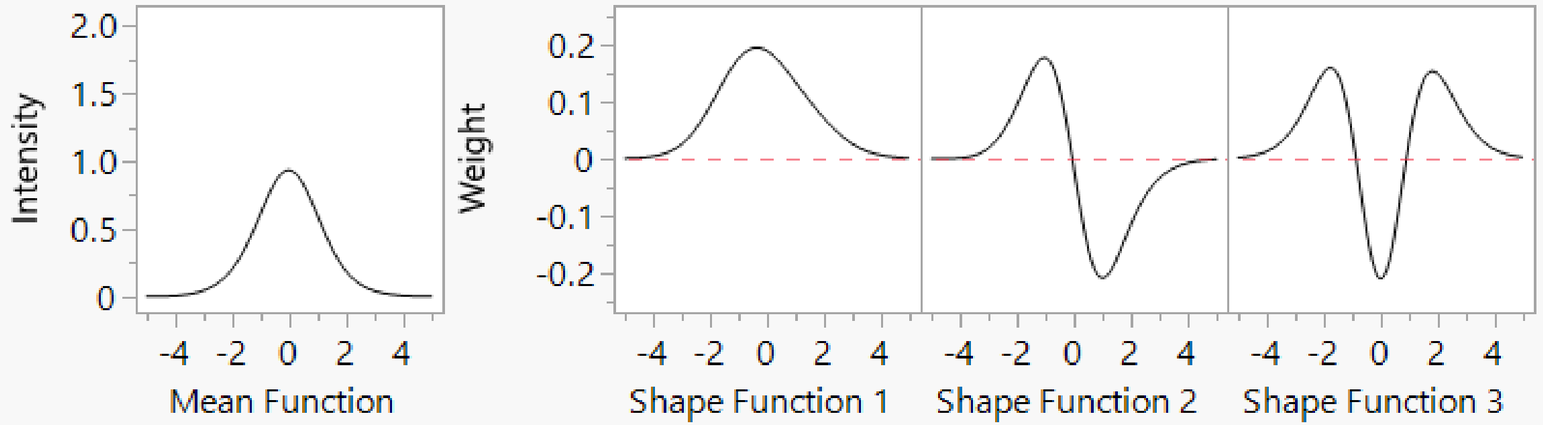
Model		Wavelet	AICc	BIC	GCV
	1	Daubechies 20	-59532.1	-78876.4	3.44e-15
	2	Symlet 20	-44379.4	-62187.8	1.35e-11
	3	Symlet 10	-52644.6	-55665.3	5.25e-10
	4	Daubechies 10	-52369.7	-55315.7	6.27e-10
	5	Coiflet 5	-48479.4	-54470.6	8.06e-10
	6	Daubechies 6	-39144.6	-40758.9	9.397e-7
	7	Symlet 6	-39174.2	-40758.8	9.423e-7
	8	Coiflet 3	-39355	-40067.7	1.448e-6
	9	Symlet 4	-30273.7	-29262.9	0.000388
	10	Coiflet 1	-19733.8	-16681.6	0.363082
	11	Daubechies 2	-19678.2	-16665.7	0.358012
	12	Haar	-11619.8	-8022.99	45.13531
	13	Biorthogonal 1.3	-5971.84	-3170.2	257.5009
	14	Biorthogonal 4.4	-2245.76	-1401.35	368.4419
	15	Biorthogonal 2.6	11632.9	10392.76	96774.19

FUNCTIONAL DATA ANALYSIS

Component Strength

FPC	Eigenvalue	20 40 60 80	Percent	Cumulative
1	3.6484		68.3%	68.3%
2	1.1660		21.8%	90.1%
3	0.4546		8.51%	98.6%

Shape Functions



longitudinal variation

	Batch	FPC 1	FPC 2	FPC 3
1	1001	-1.08	-0.32	0.01
2	1002	1.39	1.77	-0.29
3	1003	1.02	-1.86	-0.44
4	1004	-1.42	-0.24	-0.48
5	1005	-1.20	-0.56	-0.91
6	1006	-0.97	-0.30	0.48
7	1007	-0.86	-1.13	0.54
8	1008	-2.33	-0.48	-0.22
9	1009	-1.61	-0.07	-0.15
10	1010	3.42	1.05	-0.76
11	1011	-1.53	0.99	-0.00
12	1012	3.78	-2.15	-0.10
13	1013	-0.91	1.57	-0.01
14	1014	2.80	0.88	-1.07
15	1015	1.63	-0.21	0.40
16	1016	-0.07	1.45	0.64
17	1017	-0.75	0.55	0.67
18	1018	2.14	0.15	1.90
19	1019	-2.53	-0.08	-0.51
20	1020	-0.91	-1.03	0.29

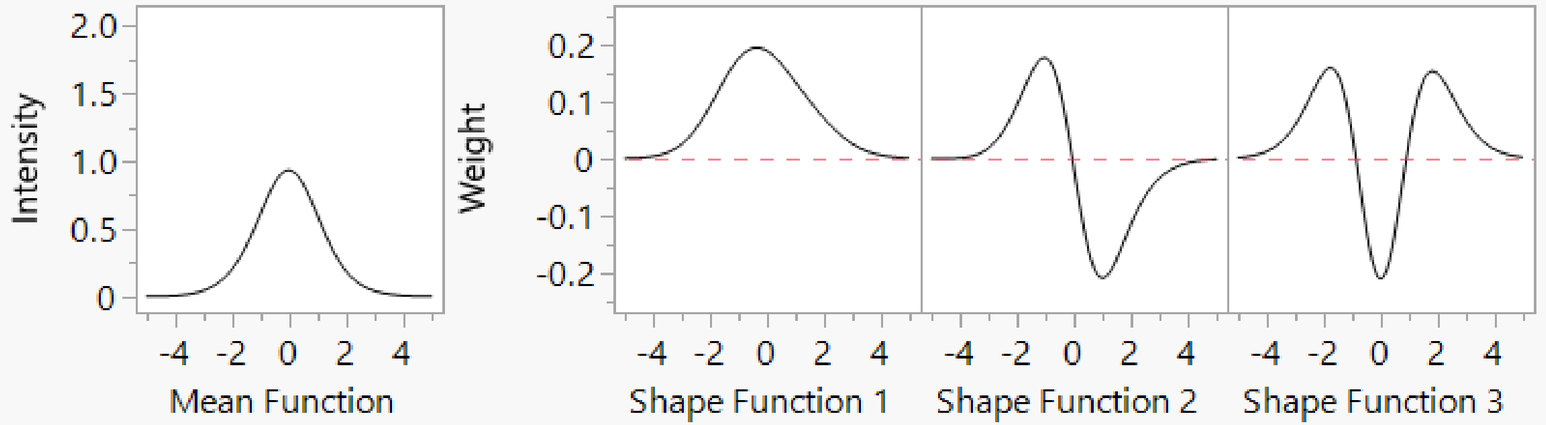
function-to-function variation

FUNCTIONAL DATA ANALYSIS

Component Strength

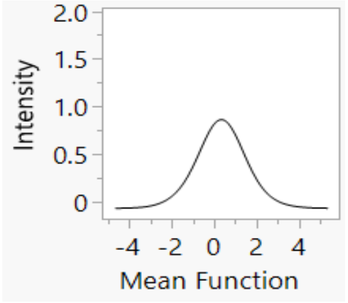
FPC	Eigenvalue	20 40 60 80	Percent	Cumulative
1	3.6484		68.3%	68.3%
2	1.1660		21.8%	90.1%
3	0.4546		8.51%	98.6%

Shape Functions



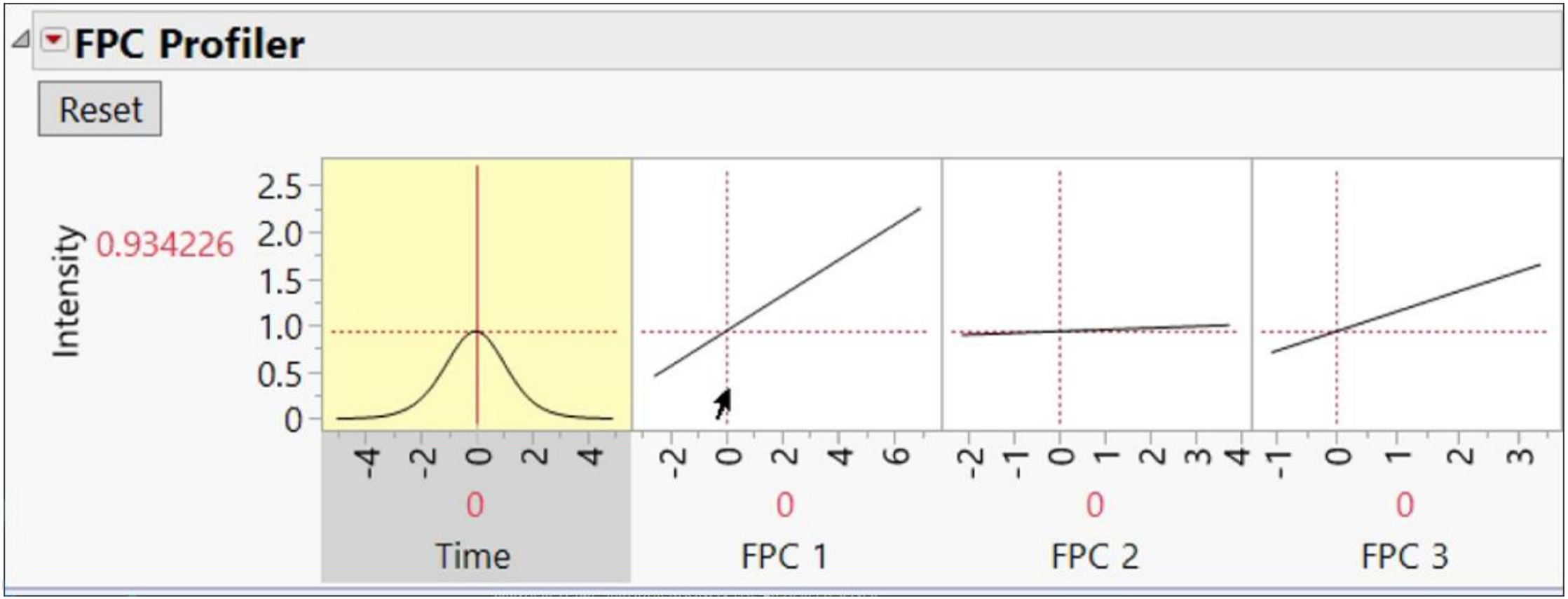
	Batch	FPC 1	FPC 2	FPC 3
1	1001	-1.08	-0.32	0.01
2	1002	1.39	1.77	-0.29
3	1003	1.02	-1.86	-0.44
4	1004	-1.42	-0.24	-0.48
5	1005	-1.20	-0.56	-0.91
6	1006	-0.97	-0.30	0.48
7	1007	-0.86	-1.13	0.54
8	1008	-2.33	-0.48	-0.22
9	1009	-1.61	-0.07	-0.15
10	1010	3.42	1.05	-0.76
11	1011	-1.53	0.99	-0.00
12	1012	3.78	-2.15	-0.10
13	1013	-0.91	1.57	-0.01
14	1014	2.80	0.88	-1.07
15	1015	1.63	-0.21	0.40
16	1016	-0.07	1.45	0.64
17	1017	-0.75	0.55	0.67
18	1018	2.14	0.15	1.90
19	1019	-2.53	-0.08	-0.51
20	1020	-0.91	-1.03	0.29

FUNCTIONAL DATA ANALYSIS

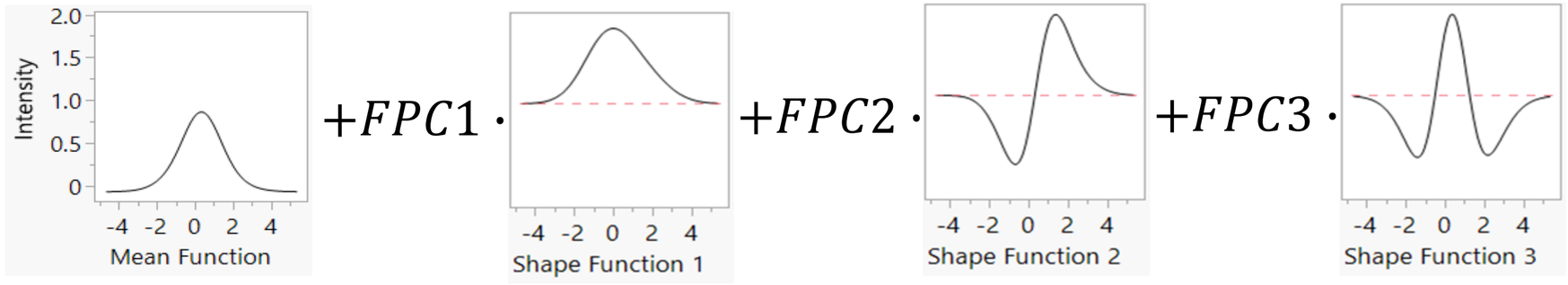


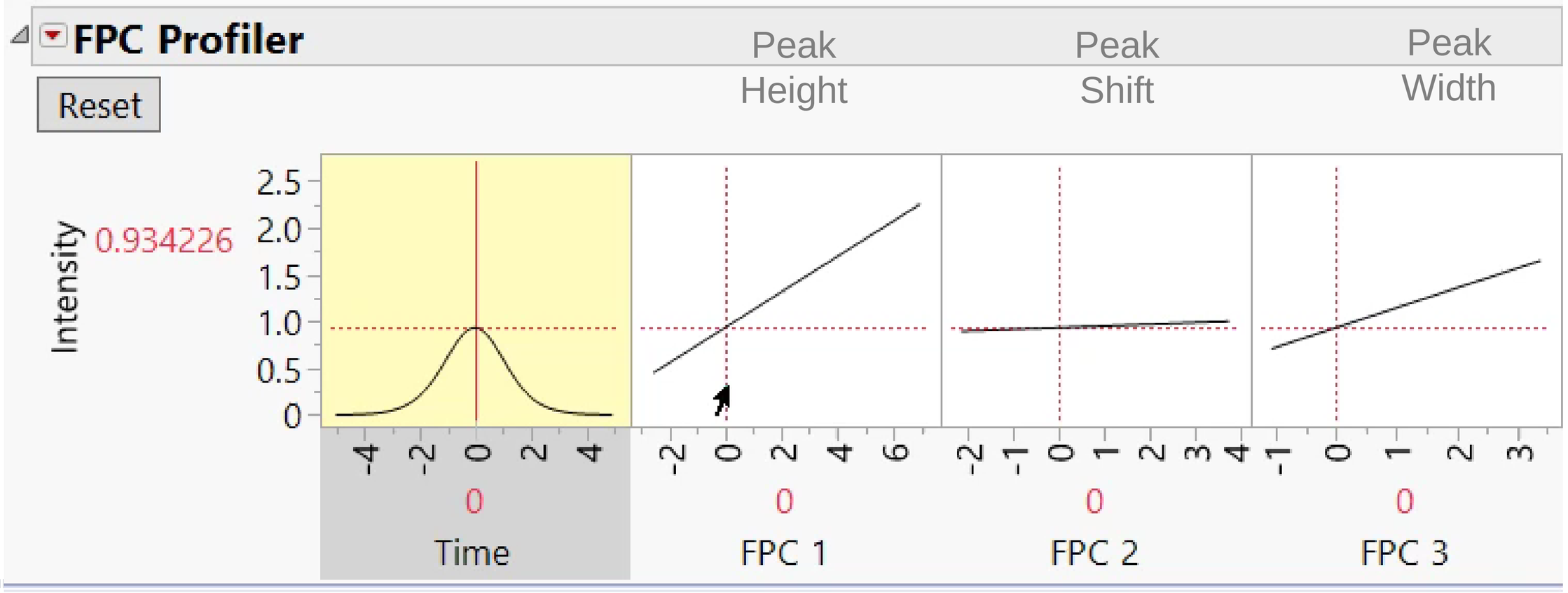
When
FPC Scores
all zero

→ Mean Curve



FUNCTIONAL DATA ANALYSIS





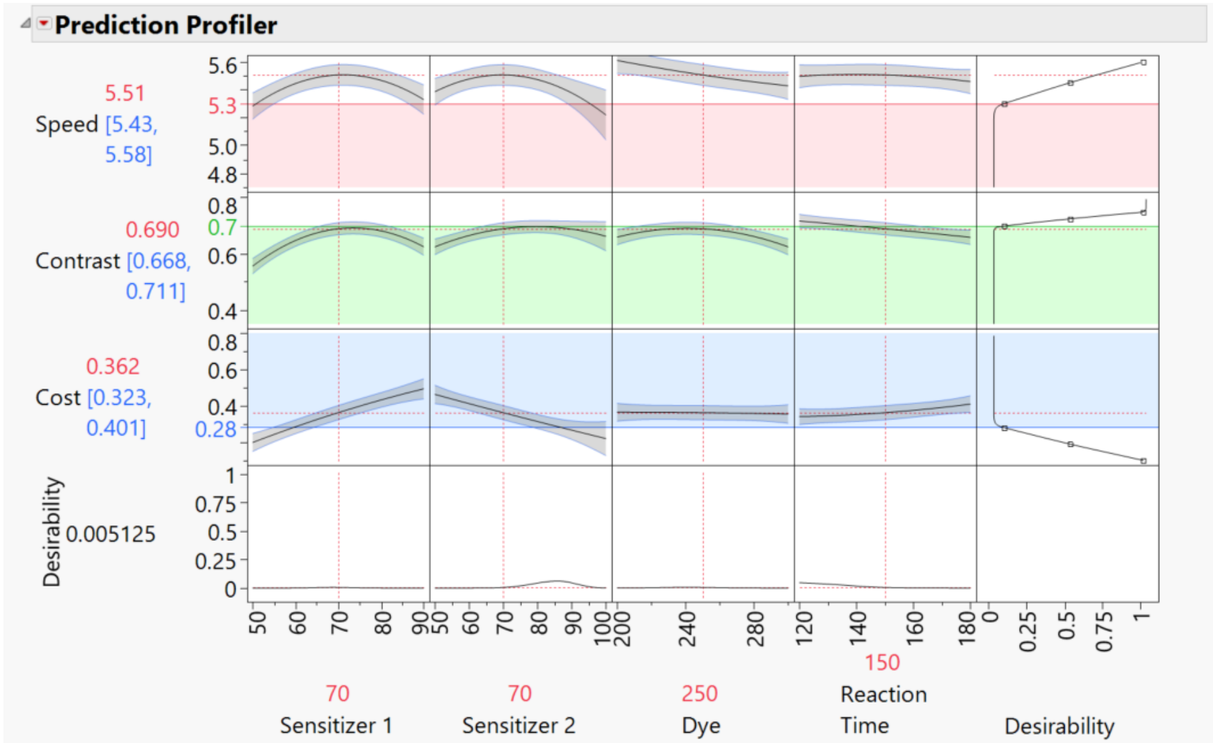
Outline

- Why / What DOE?
- What is Functional Data Analysis (FDA)?
- Reverse Engineering Case #1 – Modeling alcohol blends
- Reverse Engineer Case #2 – Mineral formulations

Takeaway #1

There's no better way to get the most information from the least amount of data than to use Design of Experiments methods

	Sensitizer 1	Sensitizer 2	Dye	Reaction Time	Speed	Contrast	Cost
7/4	90	90	300	180	5.59	0.73	0.73
27/0	50	50	200	120	4.79	0.36	0.17
1	50	50	250	120	5.36	0.616	0.198
2	50	50	200	180	5.39	0.537	0.175
3	90	70	200	120	5.31	0.623	0.447
4	50	90	200	150	5.13	0.431	0.177
5	70	70	250	180	5.37	0.643	0.445
6	50	90	300	120	4.79	0.375	0.231
7	90	90	200	180	5.45	0.626	0.471
8	90	50	250	150	5.00	0.470	0.670
9	50	50	300	150	5.22	0.478	0.283
10	70	90	200	120	5.41	0.668	0.226
11	90	90	250	120	5.33	0.734	0.310
12	50	50	250	120	5.32	0.574	0.257
13	70	50	200	150	5.49	0.596	0.456
14	50	70	250	180	5.22	0.558	0.166
15	70	70	250	150	5.57	0.689	0.390
16	90	90	300	150	5.26	0.653	0.226
17	70	70	250	150	5.47	0.688	0.356
18	70	70	300	120	5.42	0.657	0.337
19	50	70	200	120	5.43	0.518	0.222
20	50	50	300	150	5.15	0.505	0.287
21	90	70	200	120	5.33	0.661	0.457
22	50	90	300	120	4.97	0.411	0.191
23	90	50	300	120	5.09	0.492	0.588
24	90	50	300	180	5.03	0.358	0.733
25	70	70	250	150	5.59	0.707	0.318
26	70	90	300	180	5.25	0.605	0.290
27	50	90	200	150	5.24	0.476	0.177

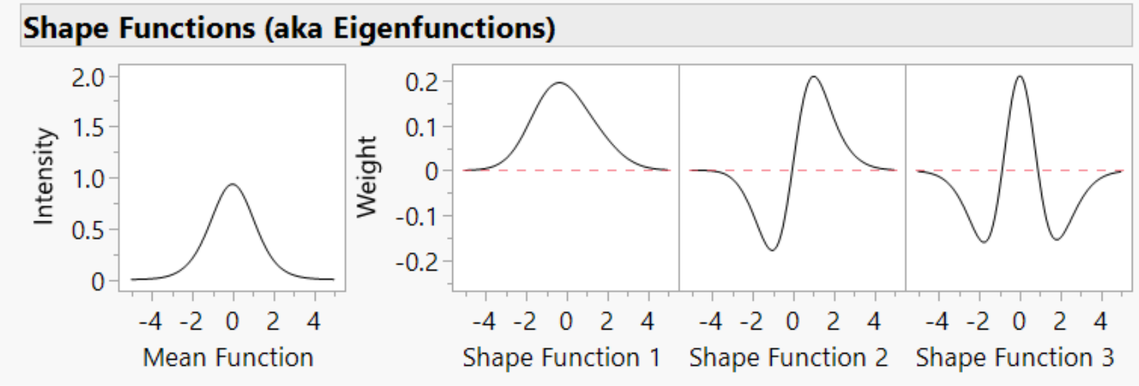


Remembered Settings

Setting	Sensitizer 1	Sensitizer 2	Dye	Reaction Time	Speed	Contrast	Cost
Equal Importance Opt	80.753574	91.269729	250.57625	120	5.3542877	0.7466933	0.2504014
Mid Point Settings	70	70	250	150	5.5054353	0.6895831	0.3623274
Cost 6X Speed & Contrast	84.016038	93.725925	283.02514	120	5.2902084	0.72549	0.1991539
Opt Spd3X-Cntr1X-Cost6X	81.958309	90.706277	286.82246	120	5.3269582	0.7177857	0.2211116

Takeaway #2

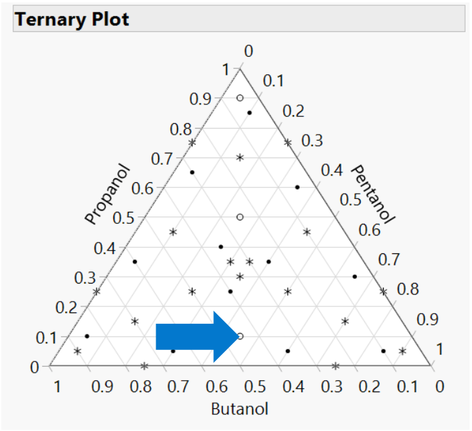
- FDA breaks apart highly correlated longitudinal data like spectra into two parts:
  - 1. Shape functions – explaining the longitudinal variation
  - 2. FPC Scores – explaining the function-to-function variation



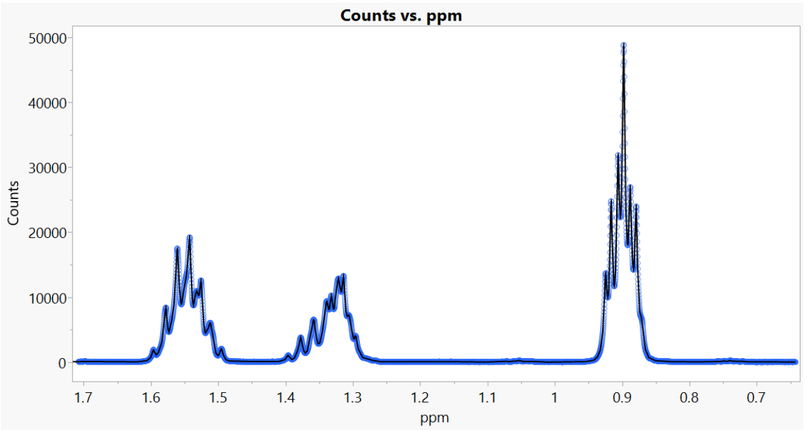
	Batch	FPC 1	FPC 2	FPC 3
1	1001	-1.08	-0.32	0.01
2	1002	1.39	1.77	-0.29
3	1003	1.02	-1.86	-0.44
4	1004	-1.42	-0.24	-0.48
5	1005	-1.20	-0.56	-0.91
6	1006	-0.97	-0.30	0.48

Takeaway #3

- FDA using wavelets can be combined with mixture DOE analysis to build models that predict spectra as a function of formulation component proportions.

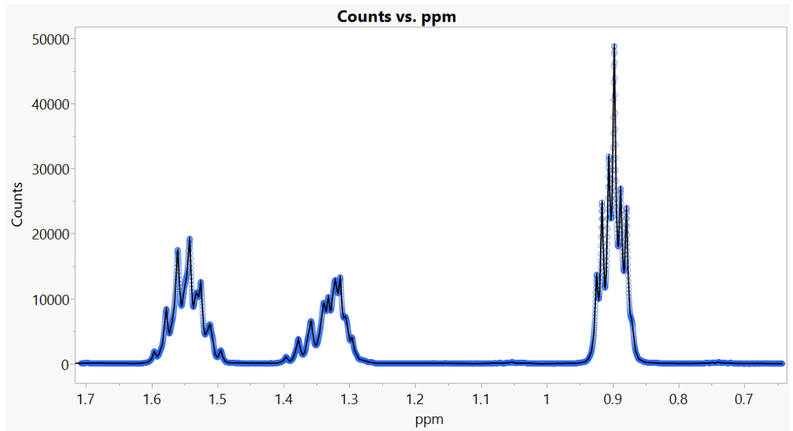


FDA-DOE Model

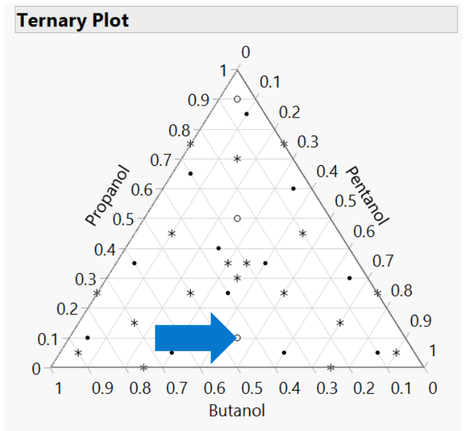


Takeaway #4

- Using spectra for an unknown blend as a “target” and an FDA-DOE mixture model, one can predict formulation component proportions. ***Reverse Engineer it!***

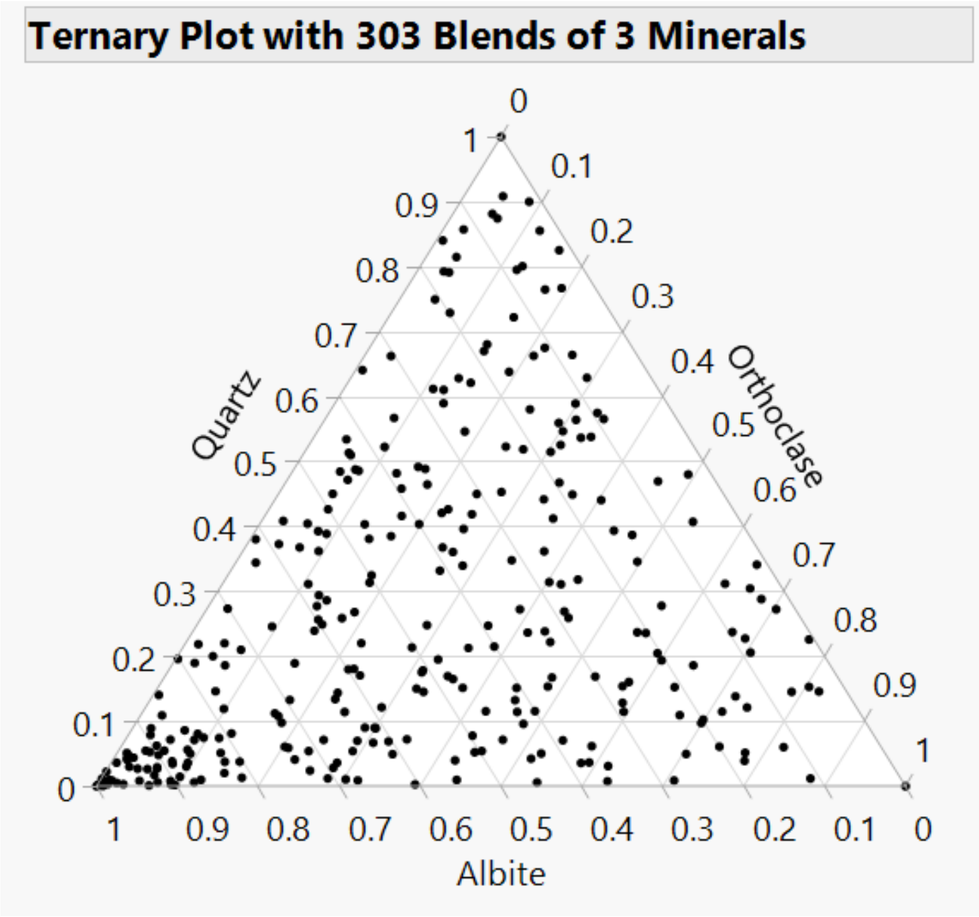


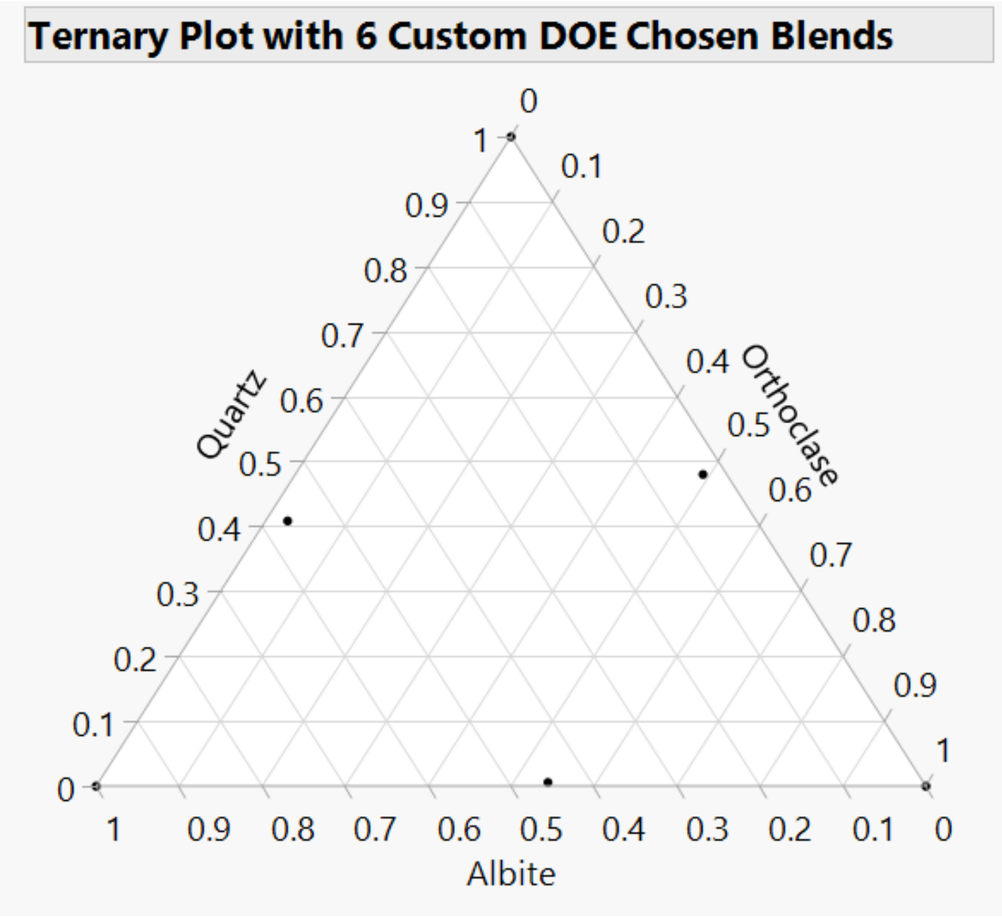
FDA-DOE Model



Takeaway #5

- Use Custom DOE to choose the most informative subset of trials (blends) from existing trials used as a candidate set.



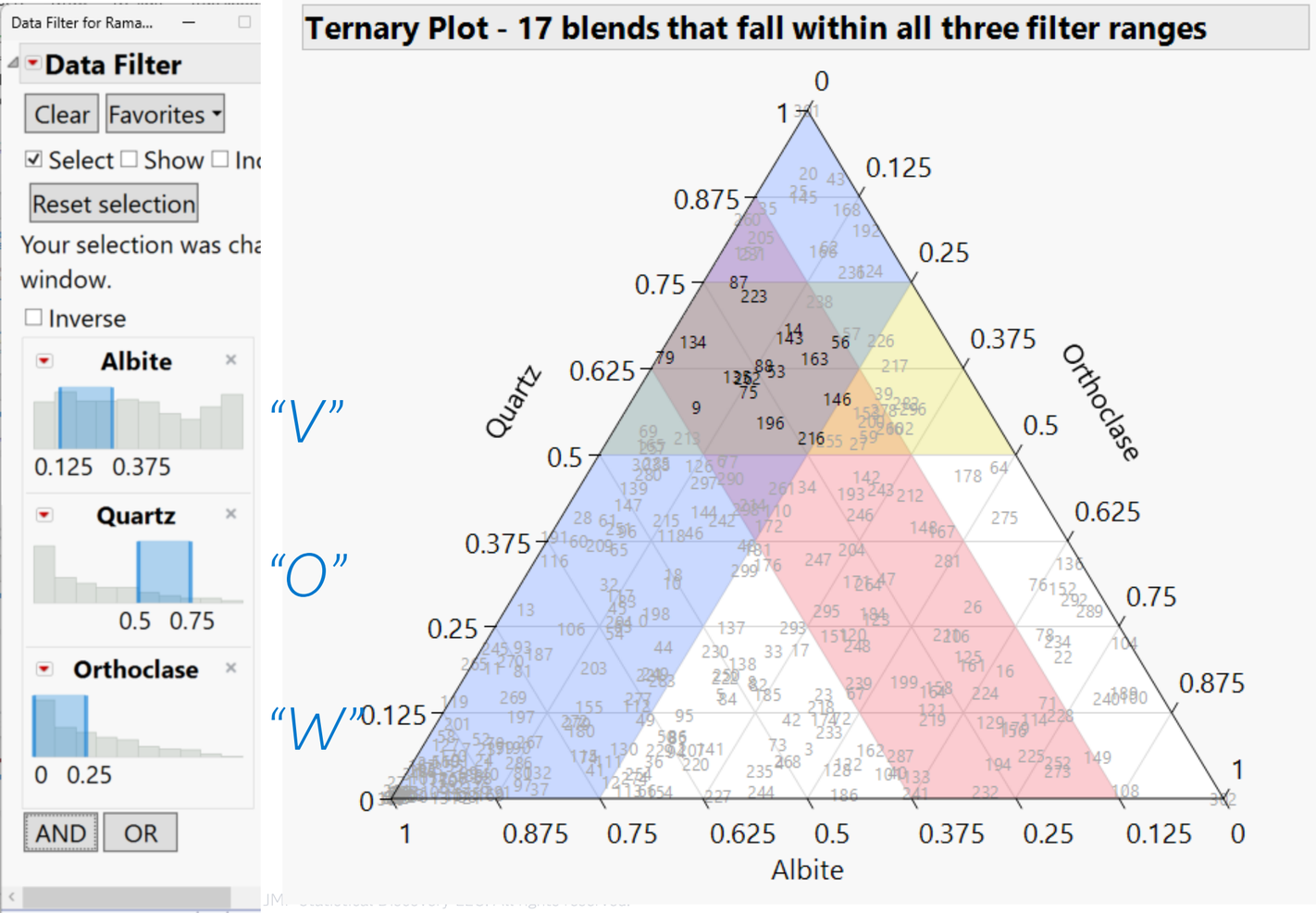


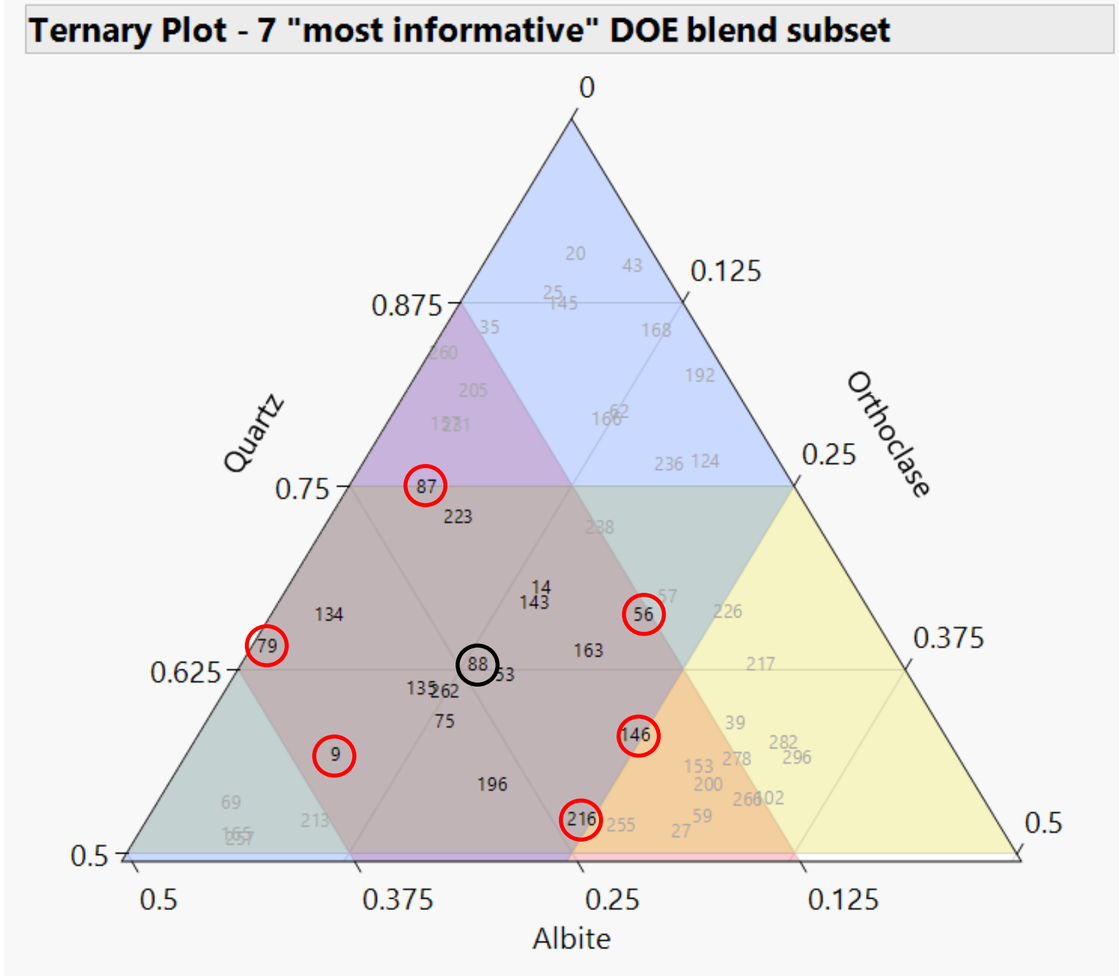
Takeaway #5 (continued)

- Use Custom DOE on a constrained region of design space

17 *filtered* blends used as candidate trials

7 circled trials are DOE subset





Case 1: Reanalysis of NMR Spectral Data for 3-Alcohol Mixture DOE using Functional Data Analysis

Available online at www.sciencedirect.com

ScienceDirect

Journal of Magnetic Resonance 190 (2008) 26–32

JMR
Journal of
Magnetic Resonance

www.elsevier.com/locate/jmr

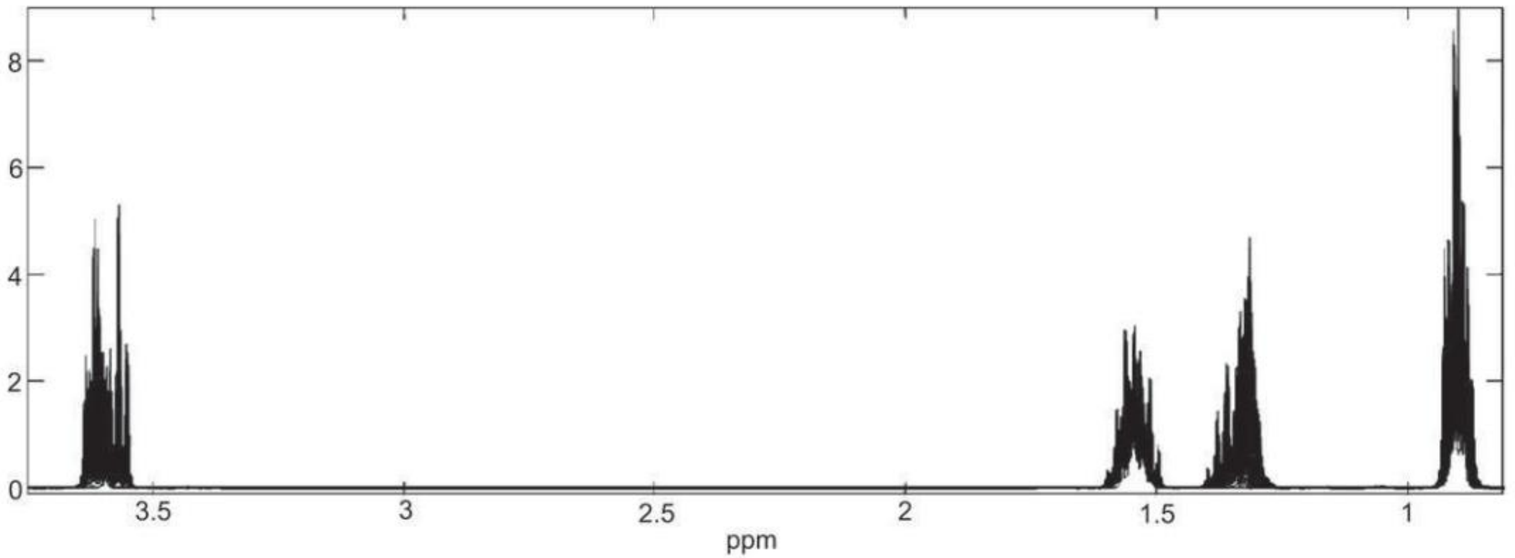
Quantitative analysis of NMR spectra with chemometrics

H. Winning *, F.H. Larsen, R. Bro, S.B. Engelsen

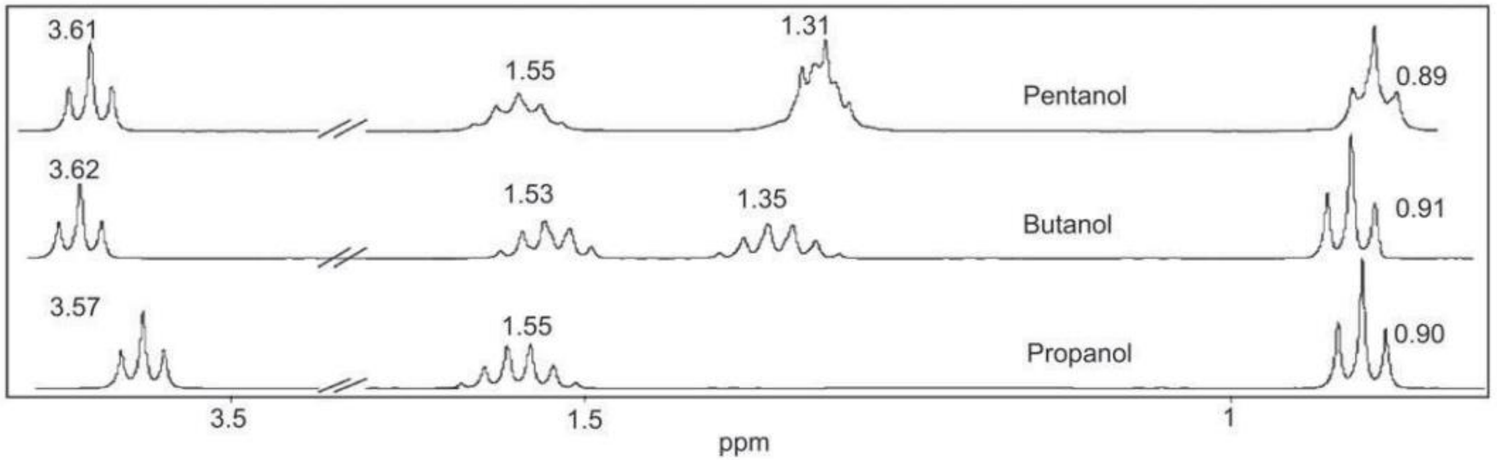
Quality and Technology, Department of Food Science, Faculty of Life Sciences, University of Copenhagen, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark

Received 11 April 2007; revised 8 October 2007
Available online 14 October 2007

Quantitative analysis of NMR spectra with chemometrics



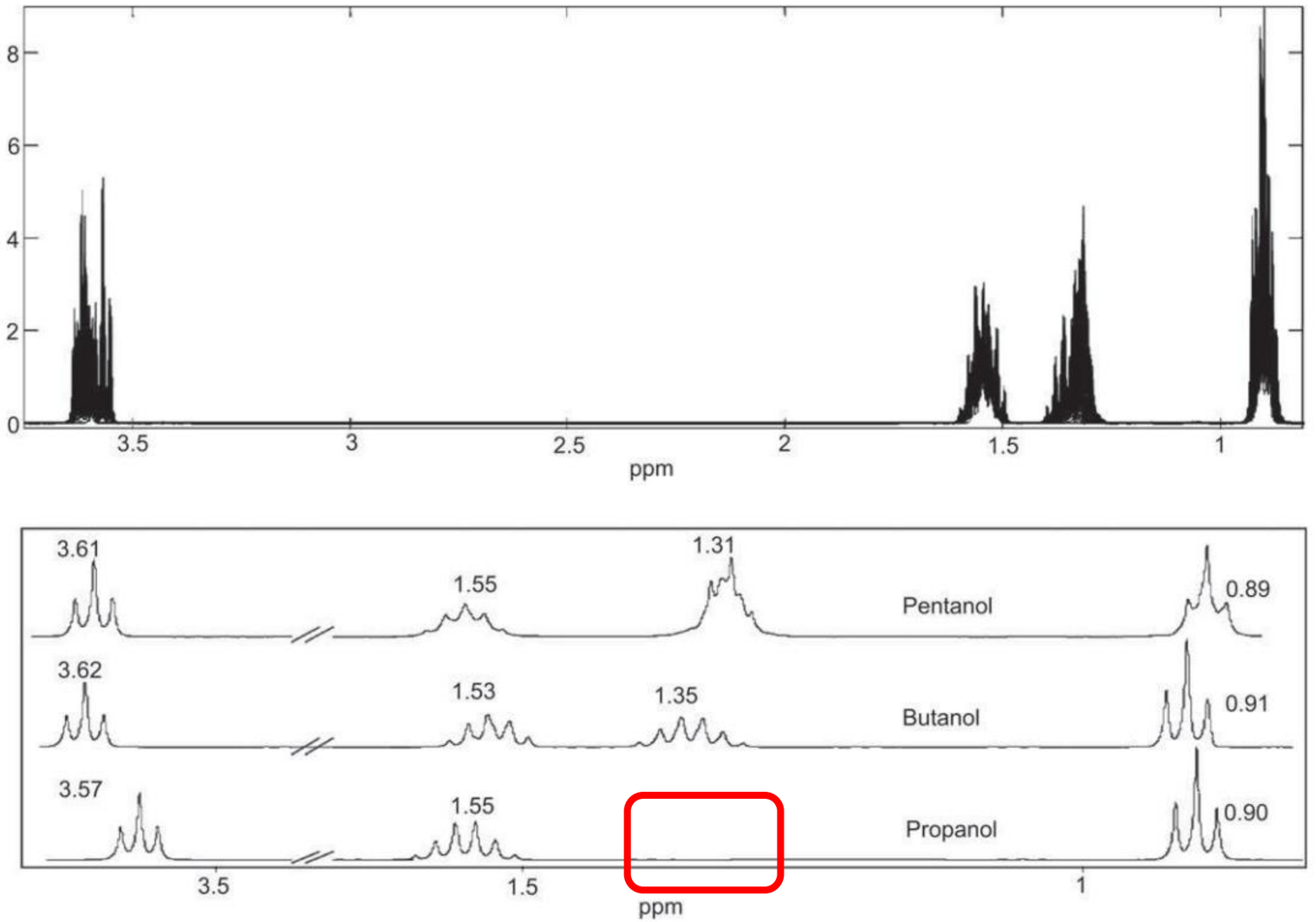
Overlay of 231 alcohol-blend spectra



Spectra for 3 pure alcohols

Fig. 2. (Top) NMR spectra of the 231 alcohol mixtures from 3.85 to 0.65 ppm. The NMR spectra of mixtures show highly overlapping signals. (Bottom) The ^{1}H NMR spectra of the pure alcohol samples of propanol, butanol and pentanol.

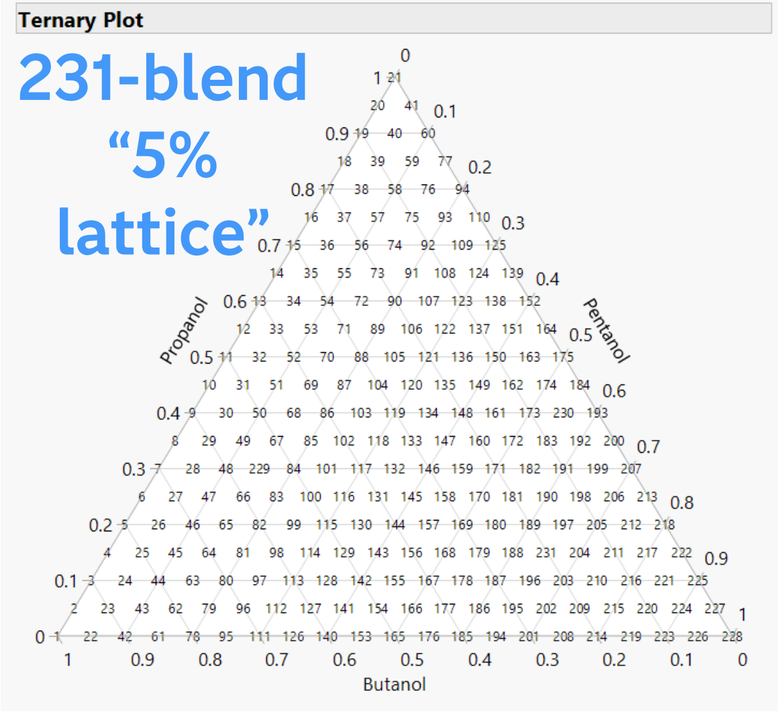
Quantitative analysis of NMR spectra with chemometrics

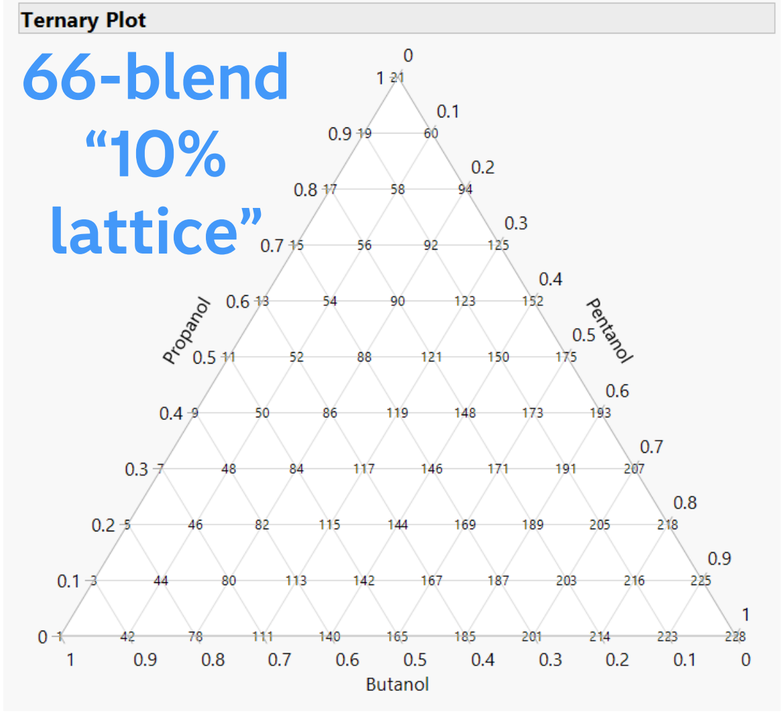


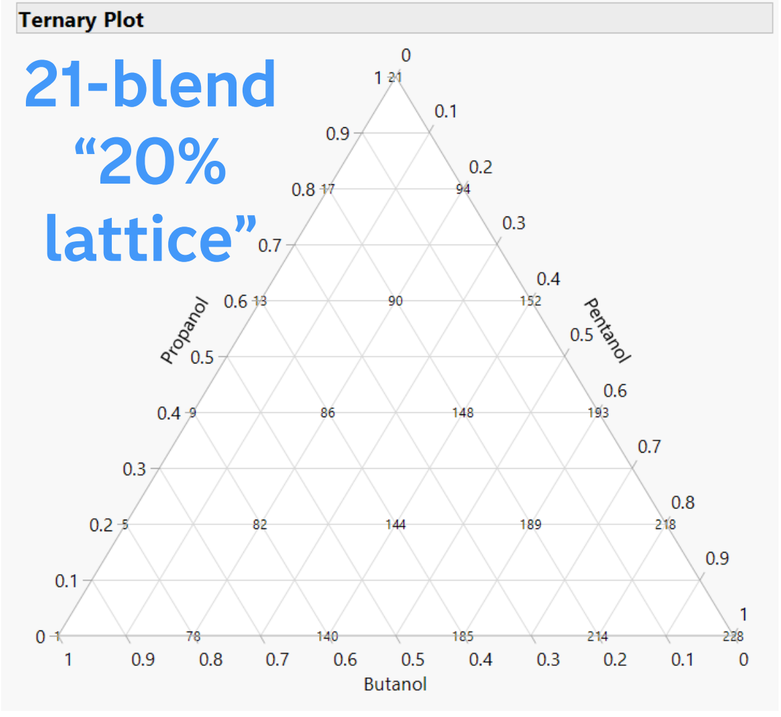
Note the absence of spectral group in Propanol

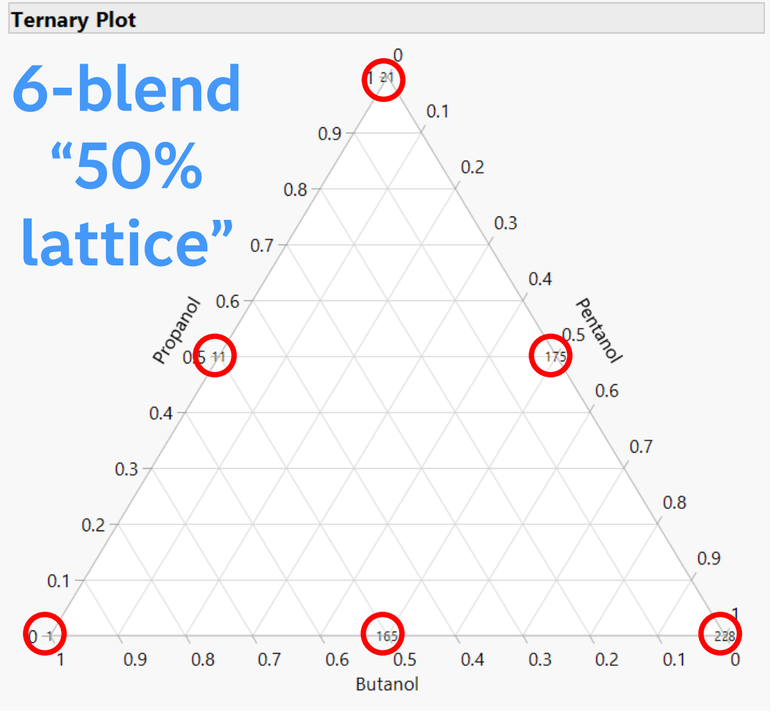
Fig. 2. (Top) NMR spectra of the 231 alcohol mixtures from 3.85 to 0.65 ppm. The NMR spectra of mixtures show highly overlapping signals. (Bottom) The ^{1}H NMR spectra of the pure alcohol samples of propanol, butanol and pentanol.

Various size “lattice” subset mixture designs



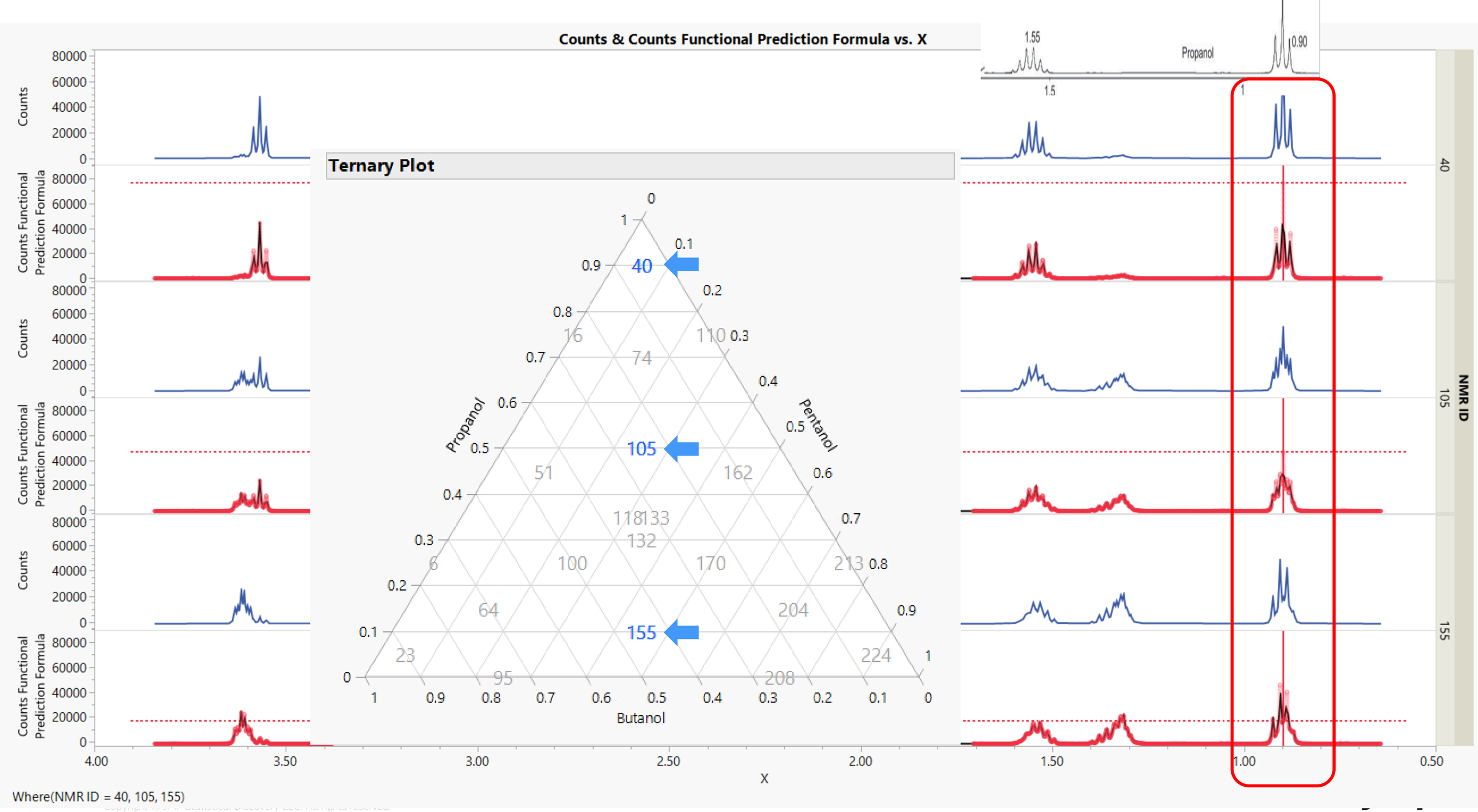


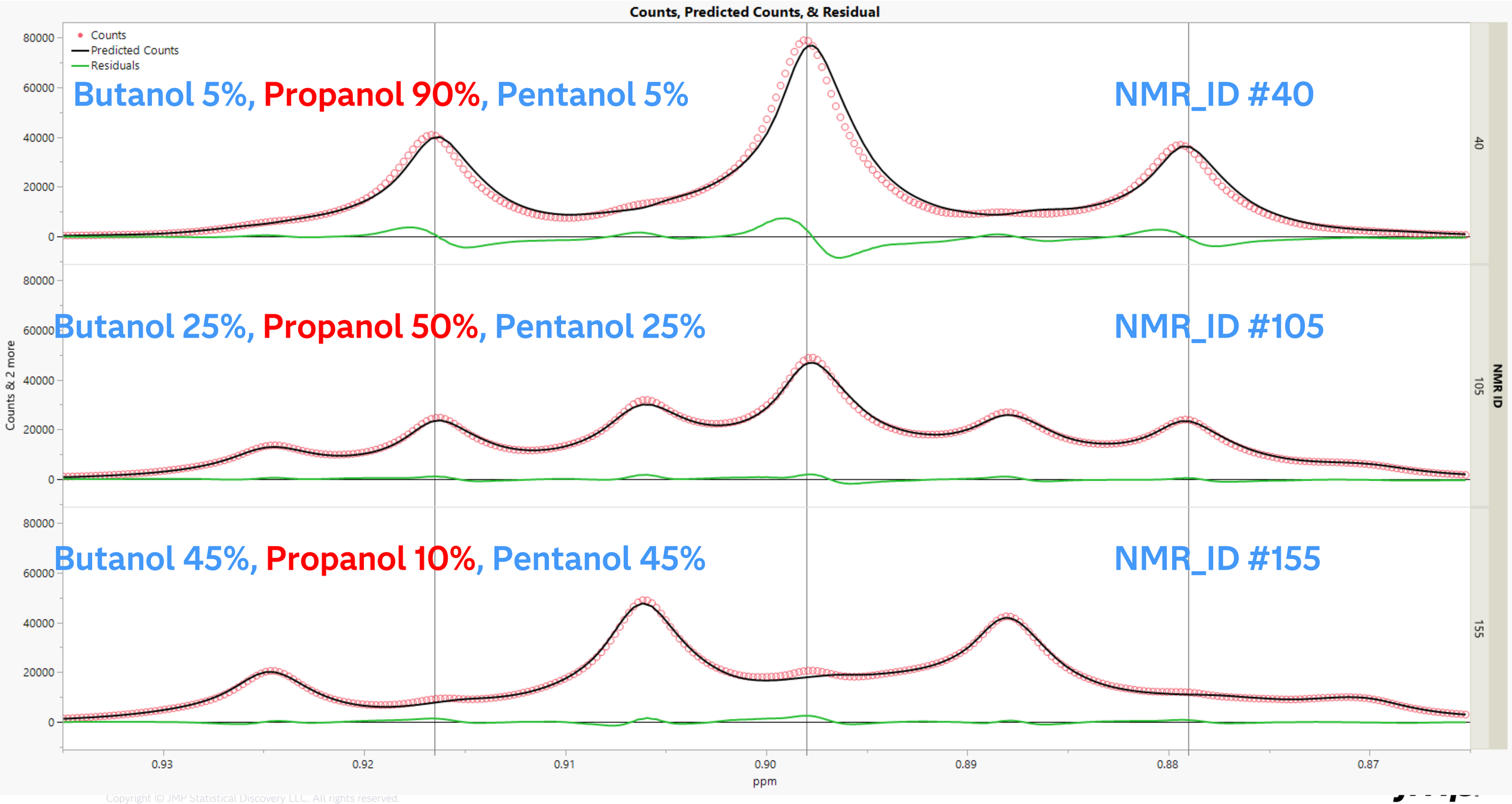




Given proportions of each component in a formulation - *that was not used in the functional data analysis* –

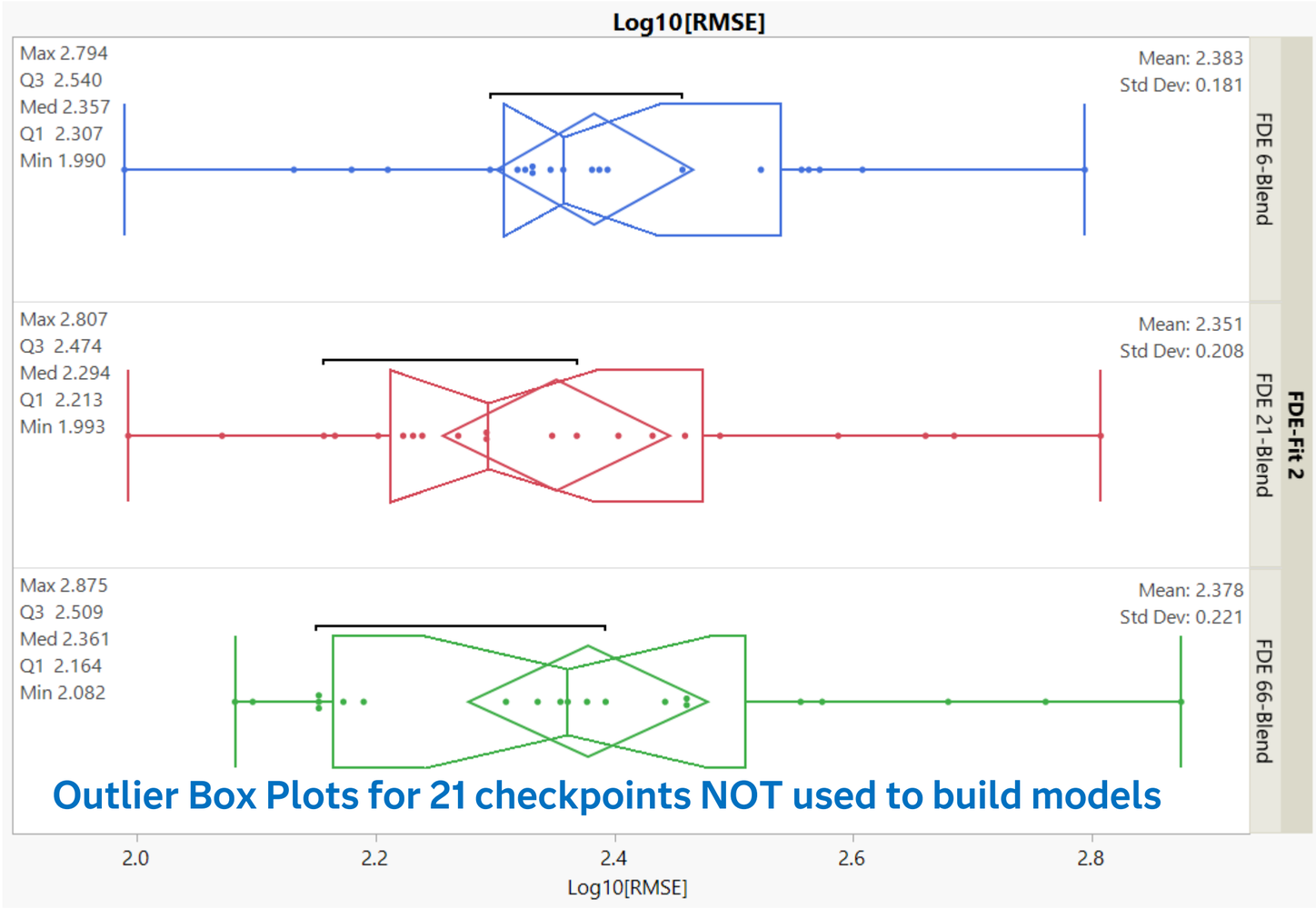
How well can the spectral shape be predicted?


Counts & Counts Functional Prediction Formula vs. X
Ternary Plot
Propanol
Pentanol
Butanol
40
105
155
Counts
Counts Functional Prediction Formula
NMR ID
X
Where(NMR ID = 40, 105, 155)

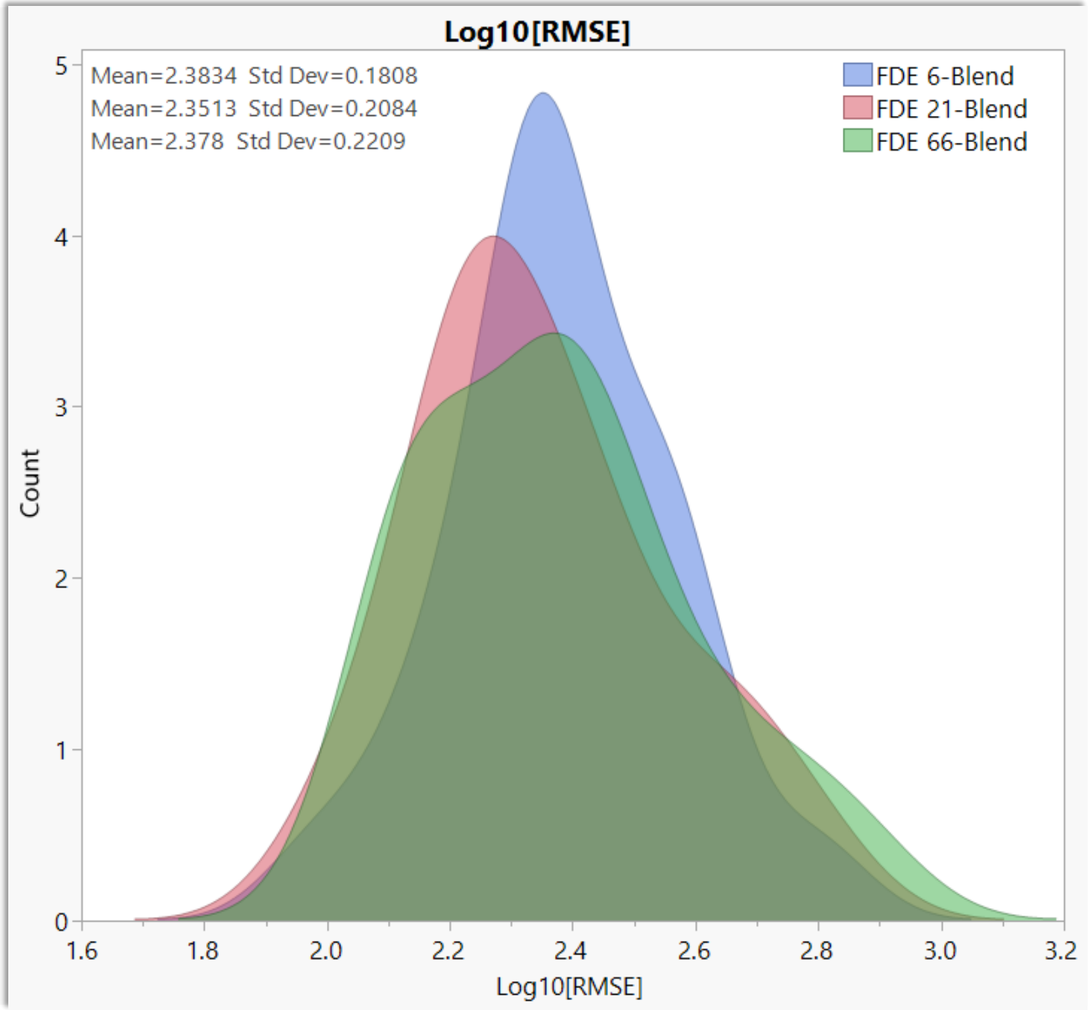


Actual vs. Predicted Plots for 21 Test Blends

No Statistical or Practical Improvement Observed when Increasing from 6 to 21 or 66 Blends to Train FDA Model



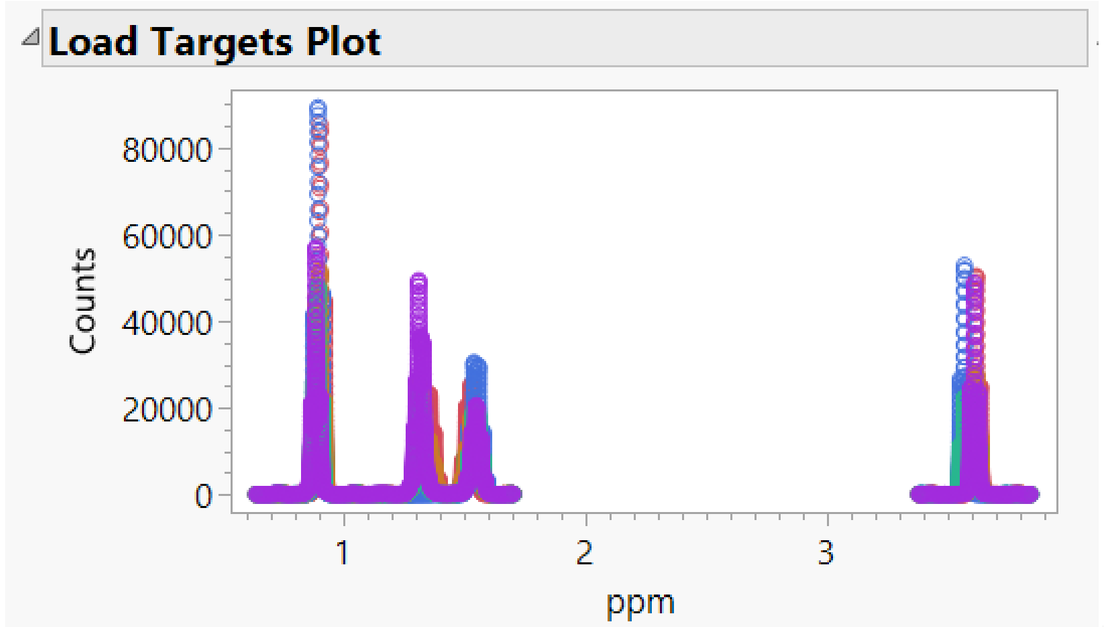
Outlier Box Plots for 21 checkpoints NOT used to build models

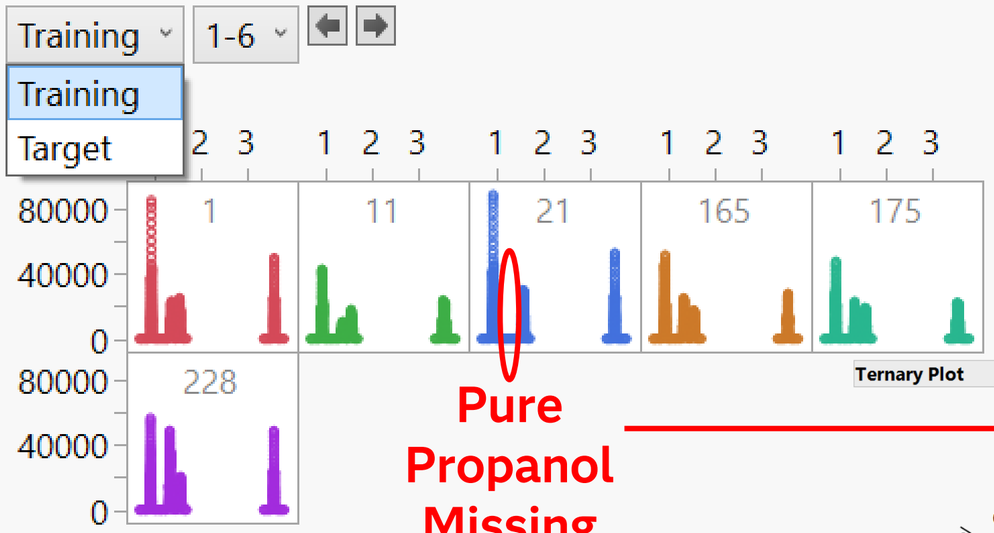


Given the spectral shape of a formulation
– *that was not used in the functional data analysis* –

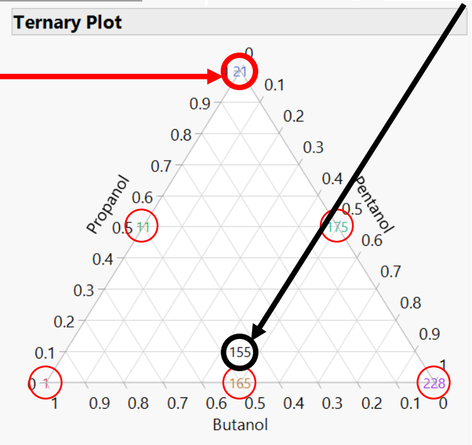
How well can the proportion of each component be predicted? Can we *Reverse Engineer it?*

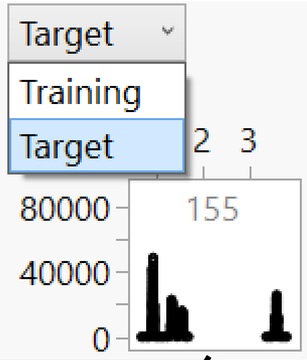
Functional Data Explorer:
After loading 6 *Training* spectra and 1 *Target* spectra to predict composition of NMR ID #155

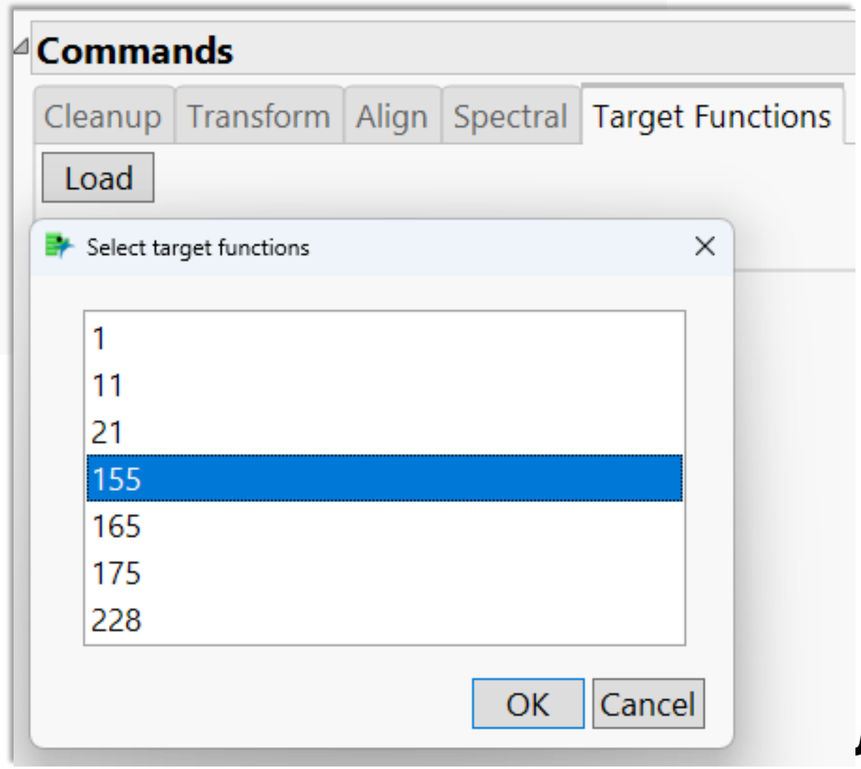




Pure Propanol Missing Spectral Group





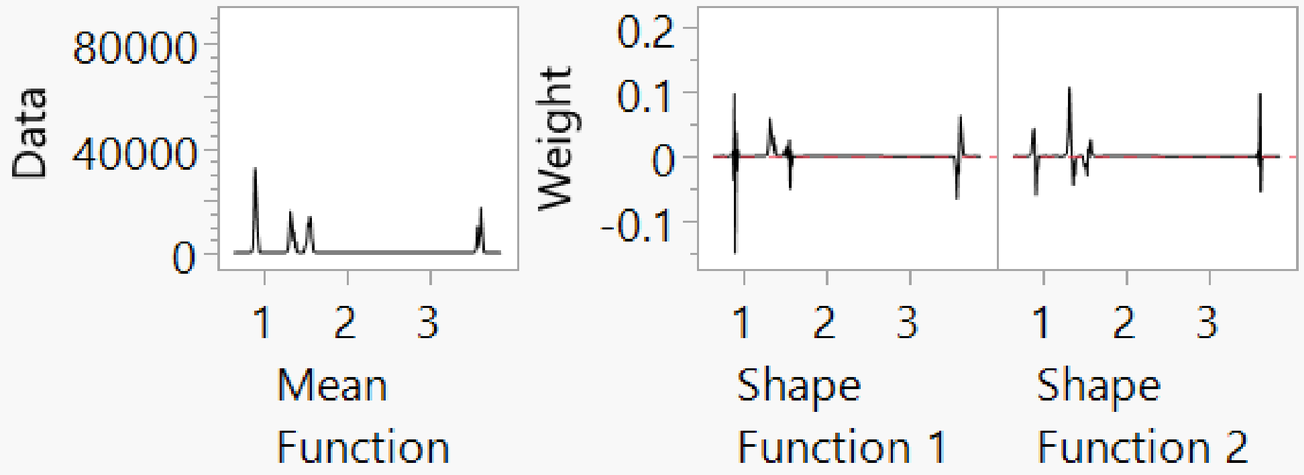


FUNCTIONAL DATA ANALYSIS

Component Strength

FPC	Eigenvalue	20406080	Percent	Cumulative
1	2.978e+10		56.8%	56.8%
2	2.213e+10		42.2%	99.1%

Shape Functions



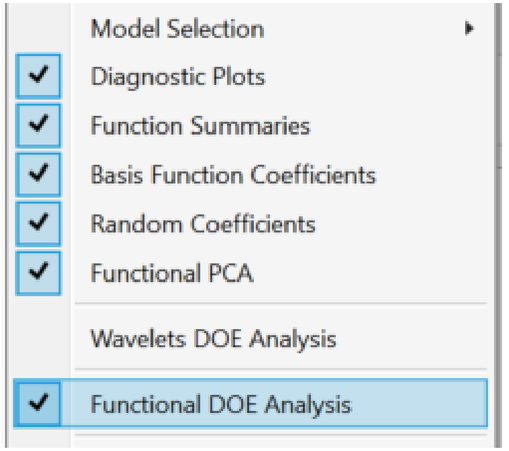
longitudinal variation

Function Summaries

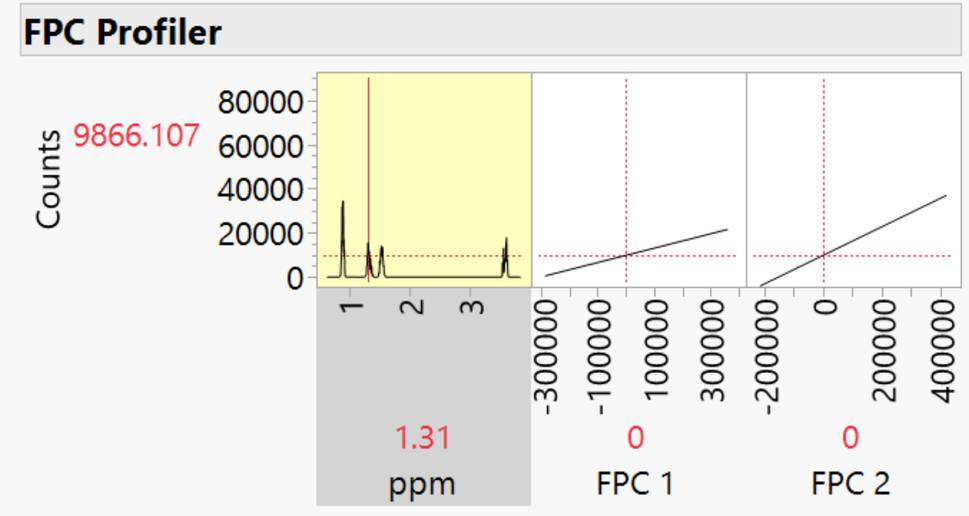
NMR ID	FPC 1	FPC 2
1	145028.69	-213039.7
11	-73207.5	-101250.7
21	-283858.6	-5560.768
165	140129.43	3883.9851
175	-63377.54	107639.32
228	135285.5	208327.83

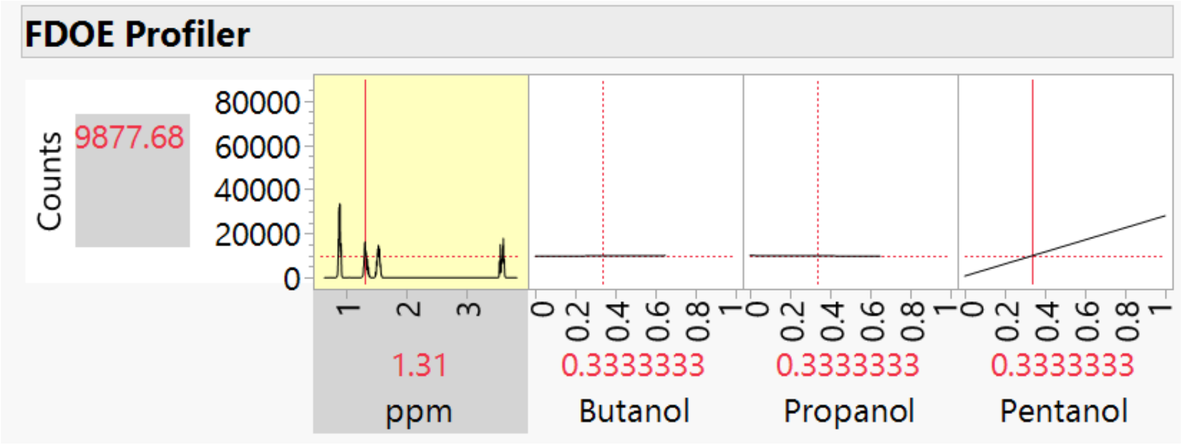
function-to-function variation

FUNCTIONAL DOE



- **One Problem!!** We don't blend FPCs. We blend alcohols!
- That is why **Functional DOE Analysis** option is so powerful
- Automatically models the FPC scores as functions of the DOE factors

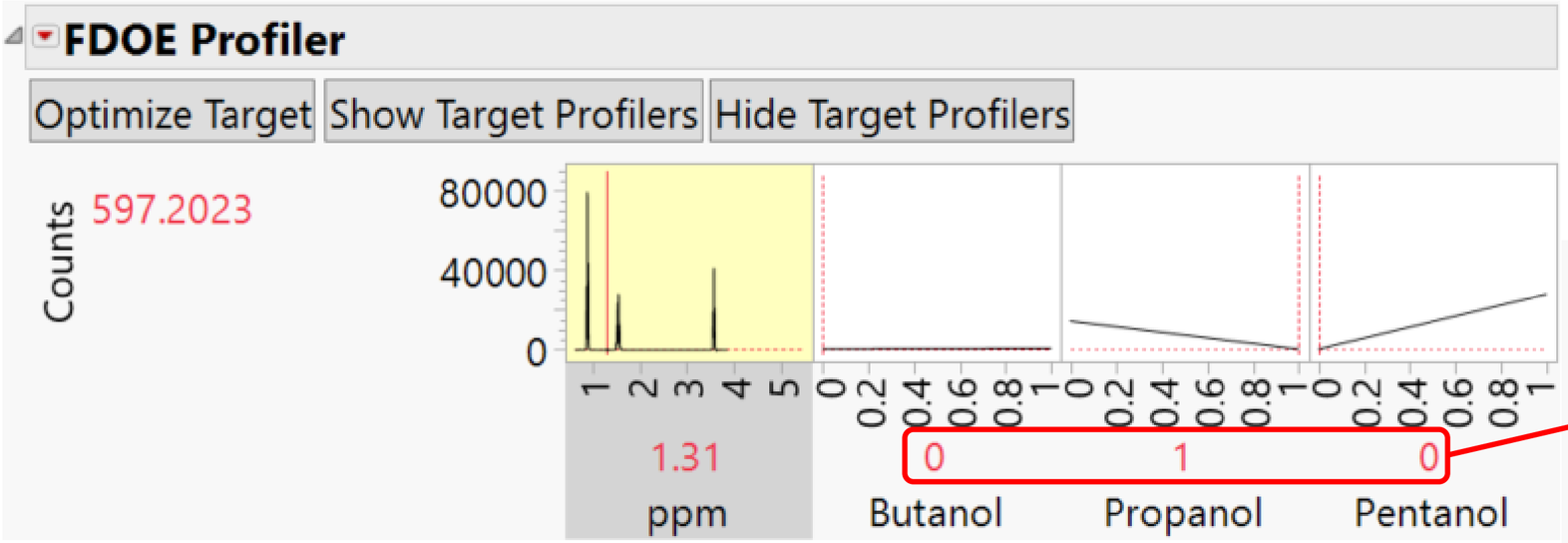


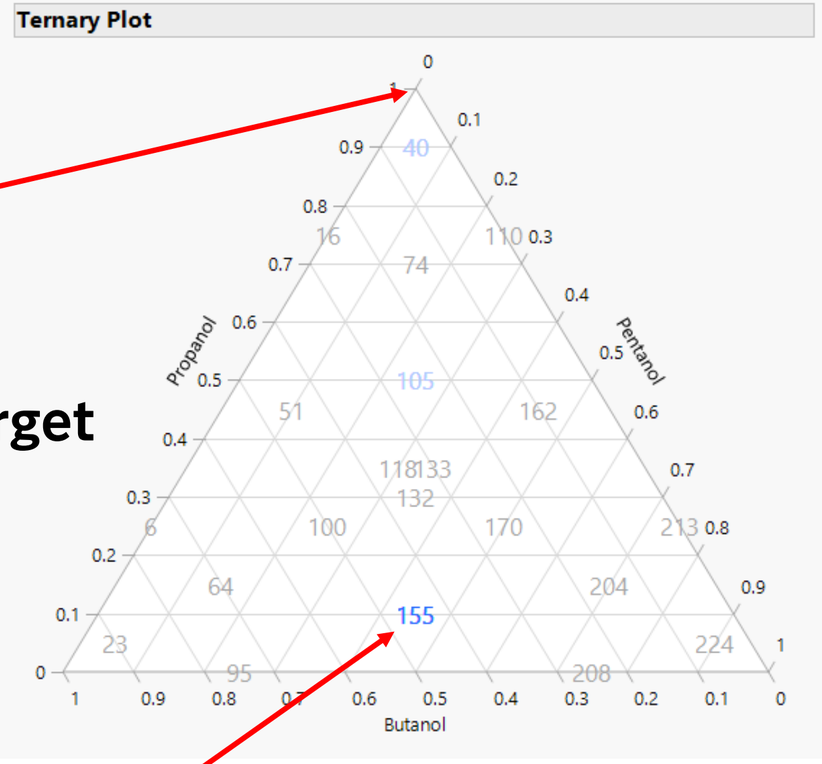


FPC 1 = f(Butanol, Propanol, Pentanol)
FPC 2 = f(Butanol, Propanol, Pentanol)

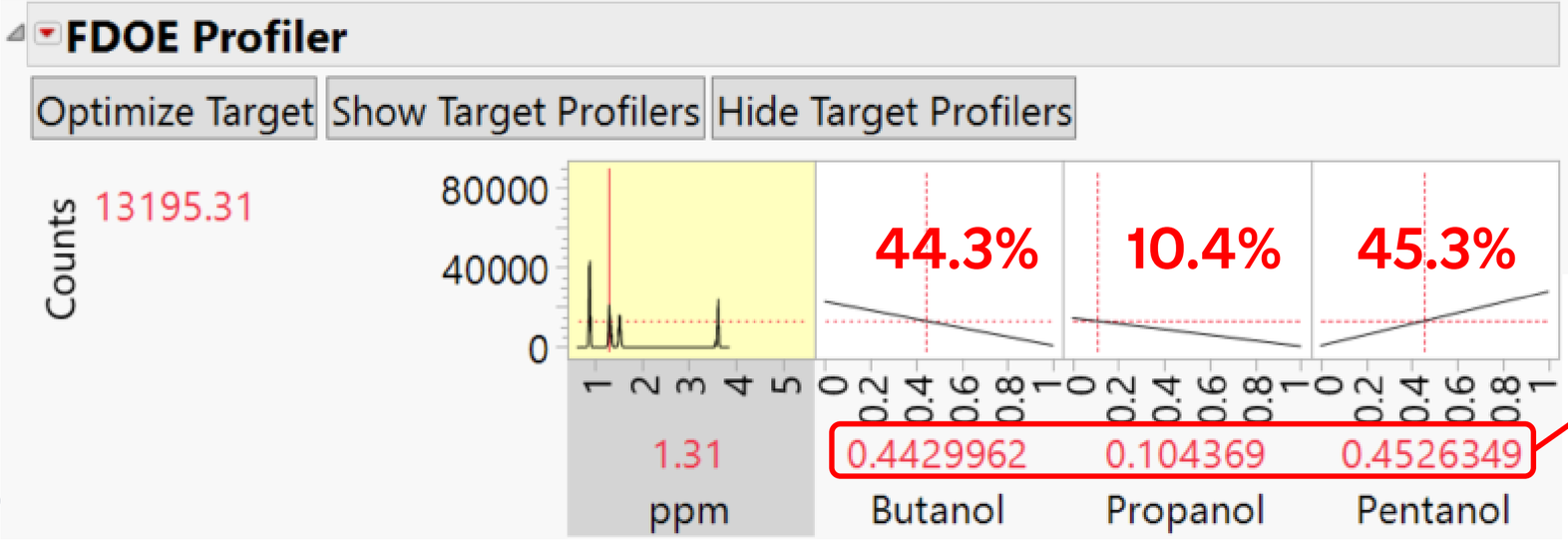
- **FDOE Profiler** model can readily be used - ***in a practical manner*** - to ***predict*** component/factor impact on spectra

Settings for pure Propanol (0% 100% 0%) – before Optimize Target

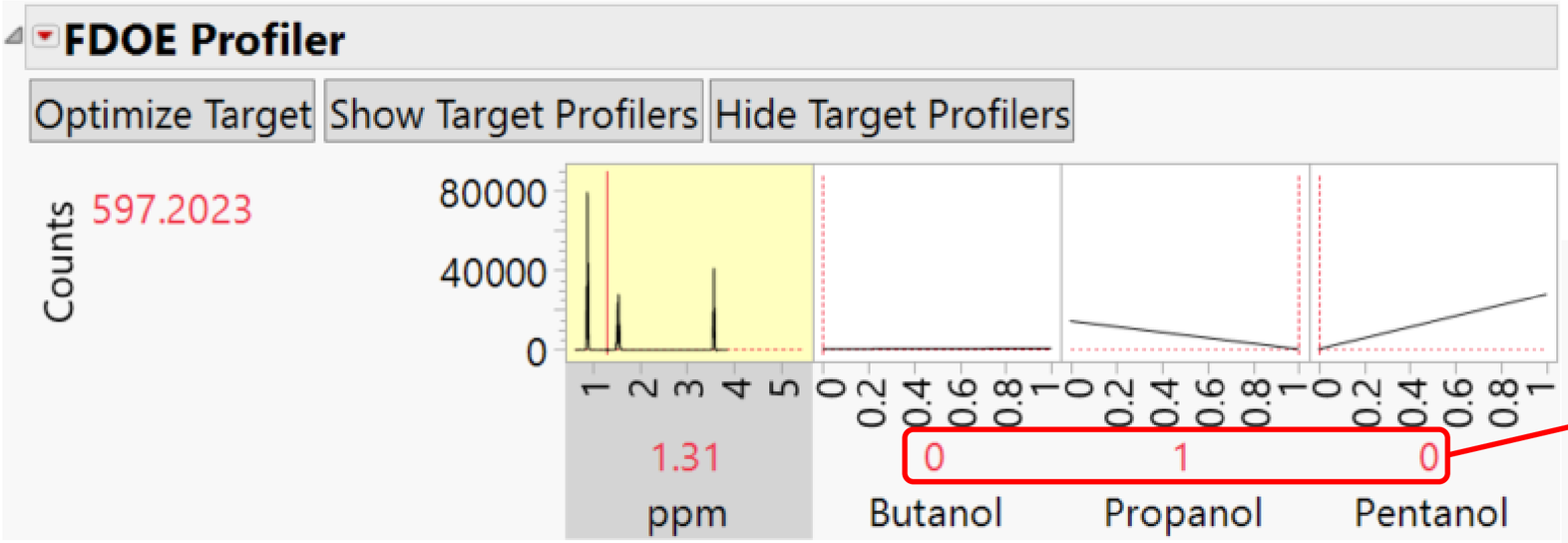


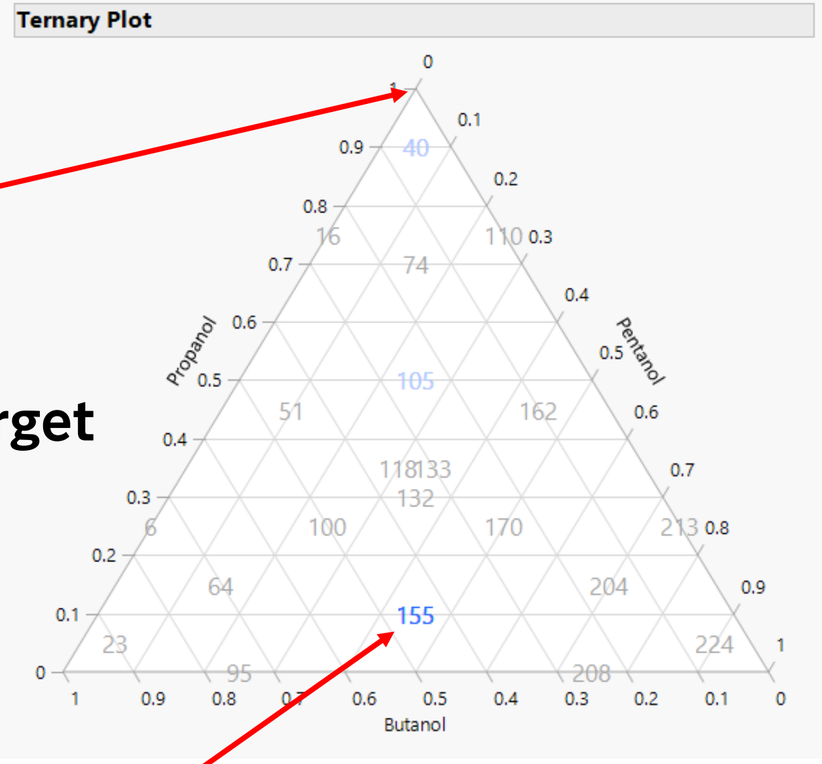


Predicted settings for target NMR #155 – after Optimize Target

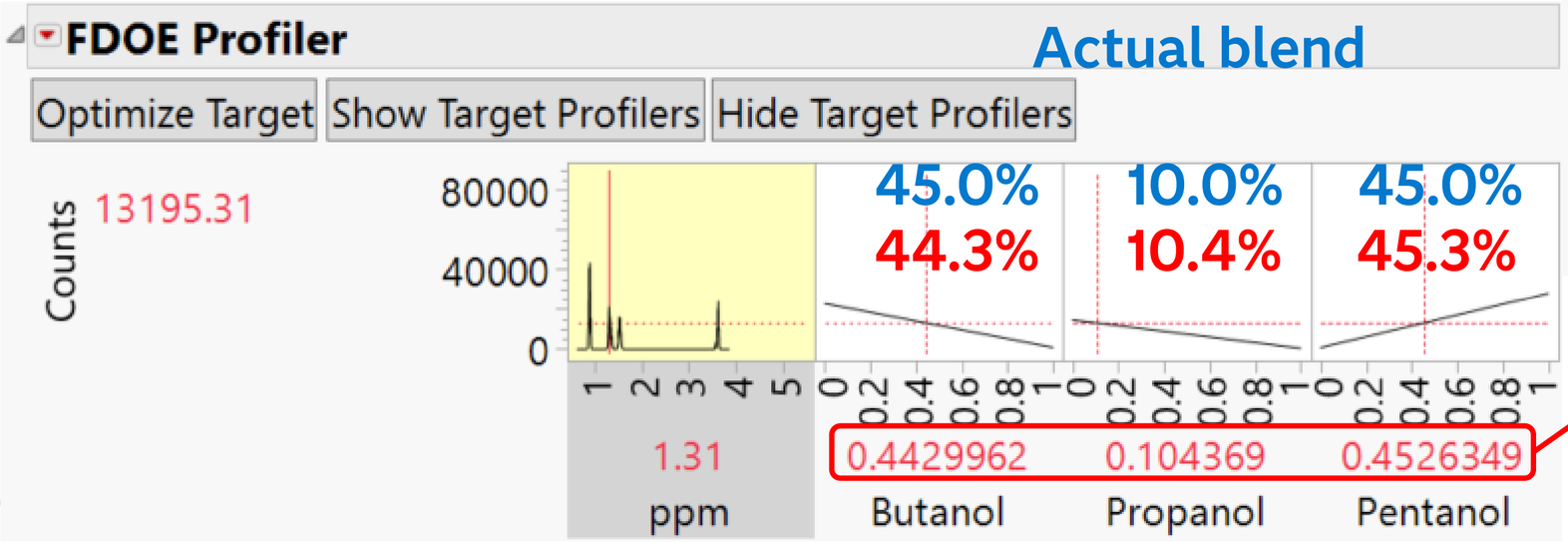


Settings for pure Propanol (0% 100% 0%) – before Optimize Target

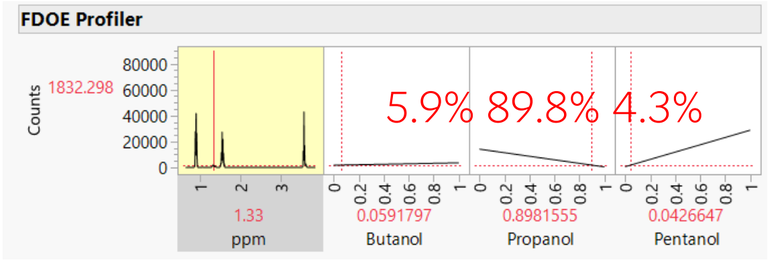




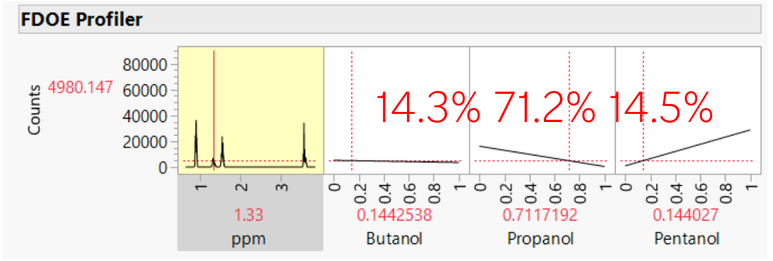
Predicted settings for target NMR #155 – after Optimize Target



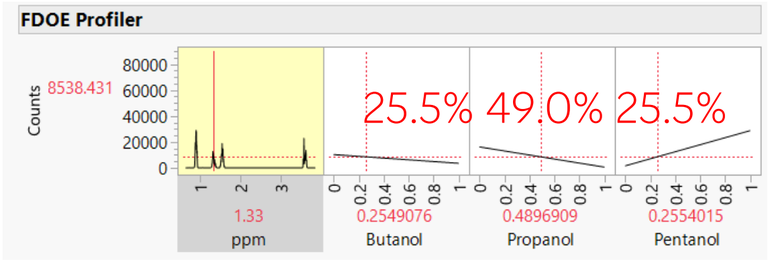
Predicted Proportions for Check Point Spectra from 6-Blend FDA Model



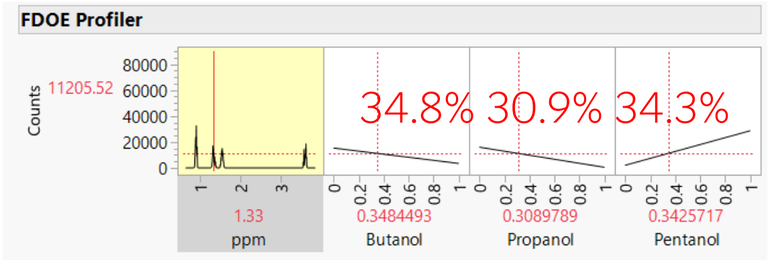
RMSE 0.67%



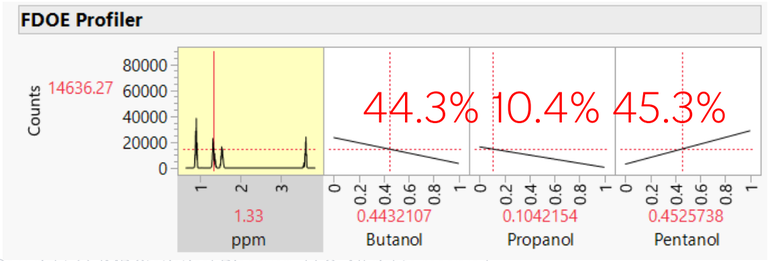
0.85%



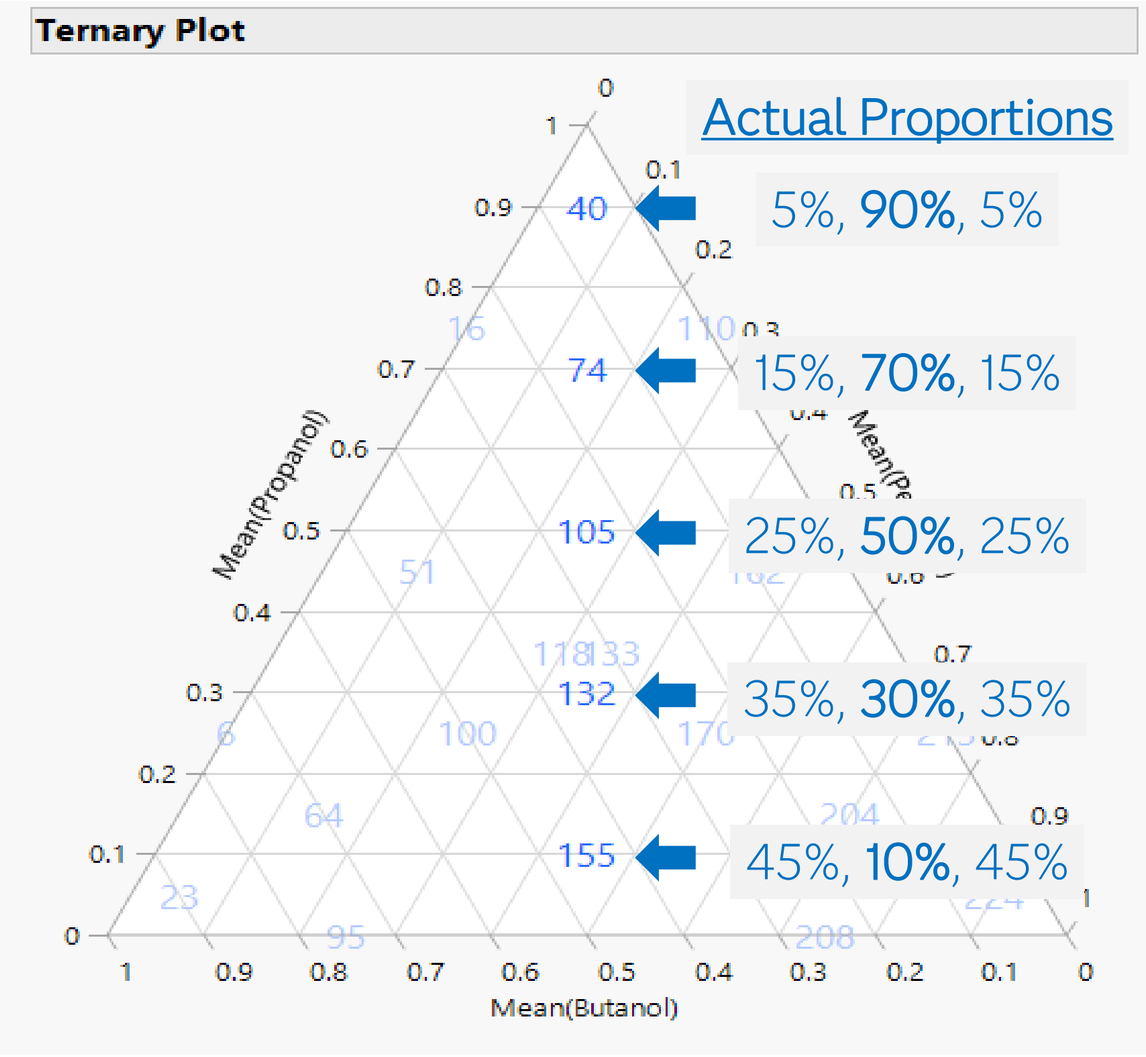
0.71%



0.67%



0.50%



Let's go to JMP…

- Perform FDA on the NMR spectra of 6 alcohol blends and identify the composition of the 7th ***target*** blend

How might this be used in industries like chemical, biopharma, food, & consumer products

1. Run a DOE with component and factor ranges you believe encompass the unknown competitor formulation settings

2. Use FDA to model the spectra of the DOE blends & conditions

3. Use FDA-DOE model and the target spectra of competitor's product to closely determine the actual formulation

How might this be used in industries like defense and aerospace?

1. Help formulate chemical composition of a decoy flare, so its spectra matches that of a particular aircraft engine

2. By analyzing
   - spectra of sensors
   - sonar signatures
   - radar signatures

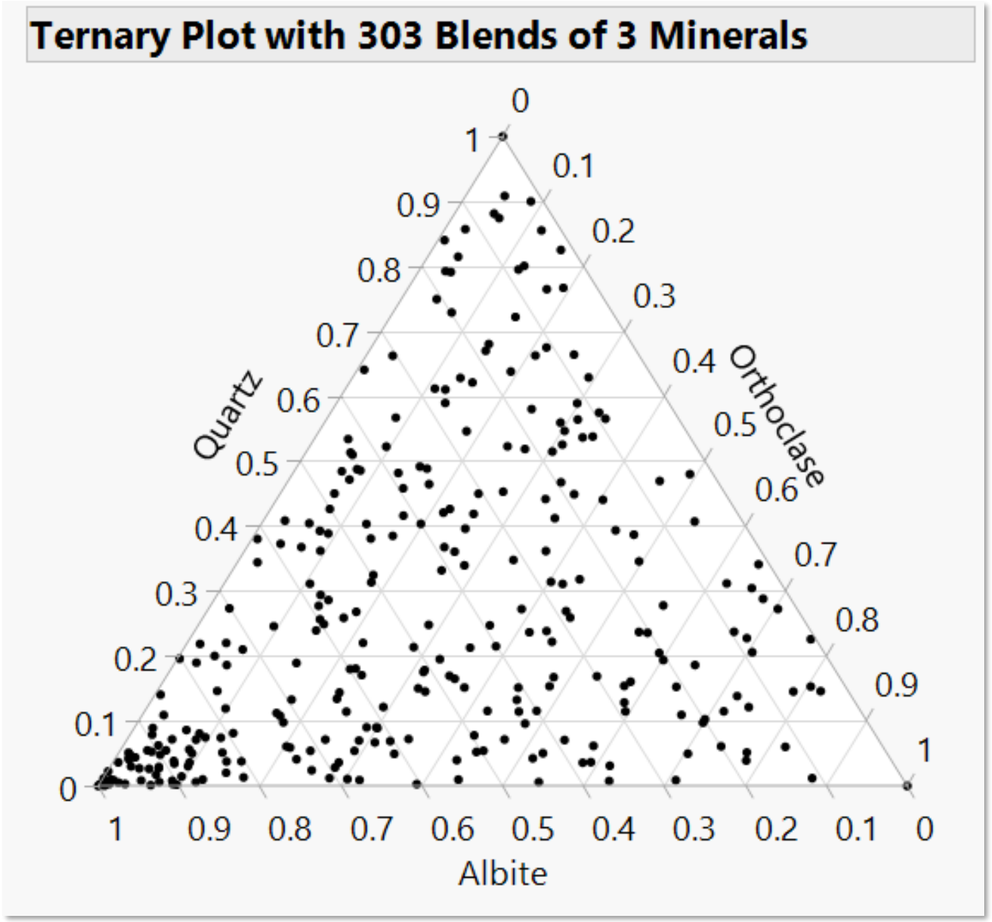
   } All as functions of factors like shape, speed, angle, distance, vibration, etc.

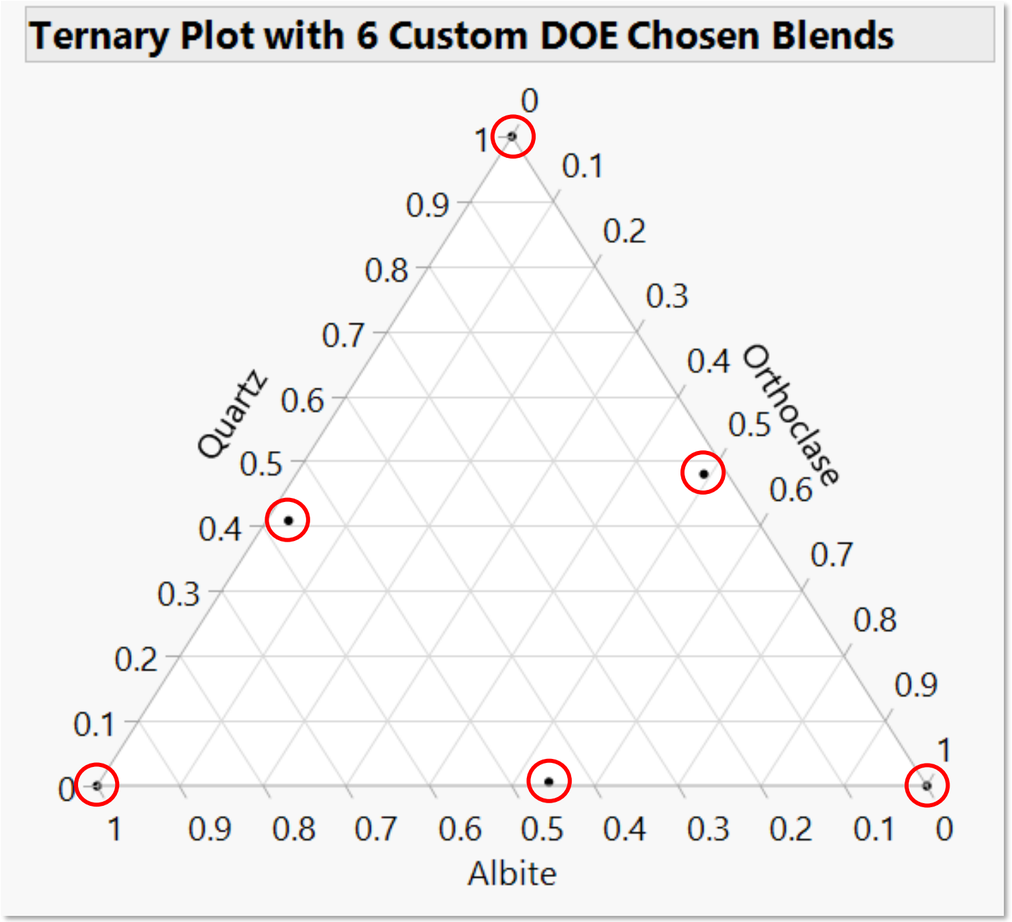
Better predict remaining useful life of engines and improve identification of detected submersibles and aircraft

Case 2 - Reanalysis of Raman Spectral Data for 3-Mineral Mixture DOE using FDA

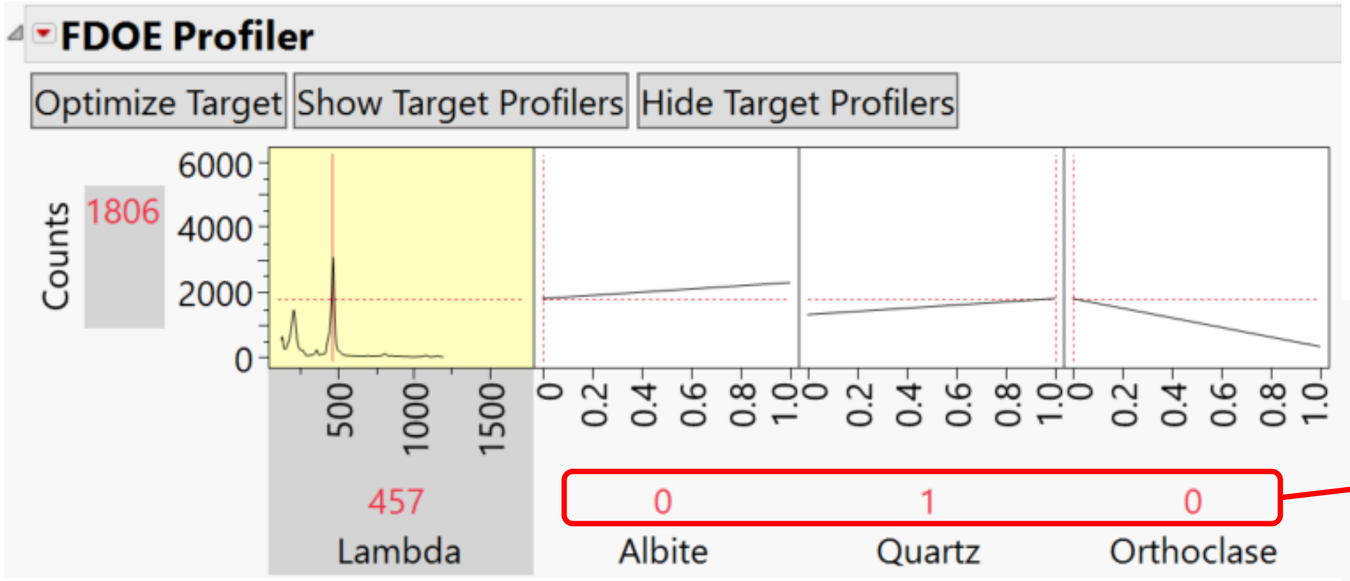
All 303 blends used as candidate trials

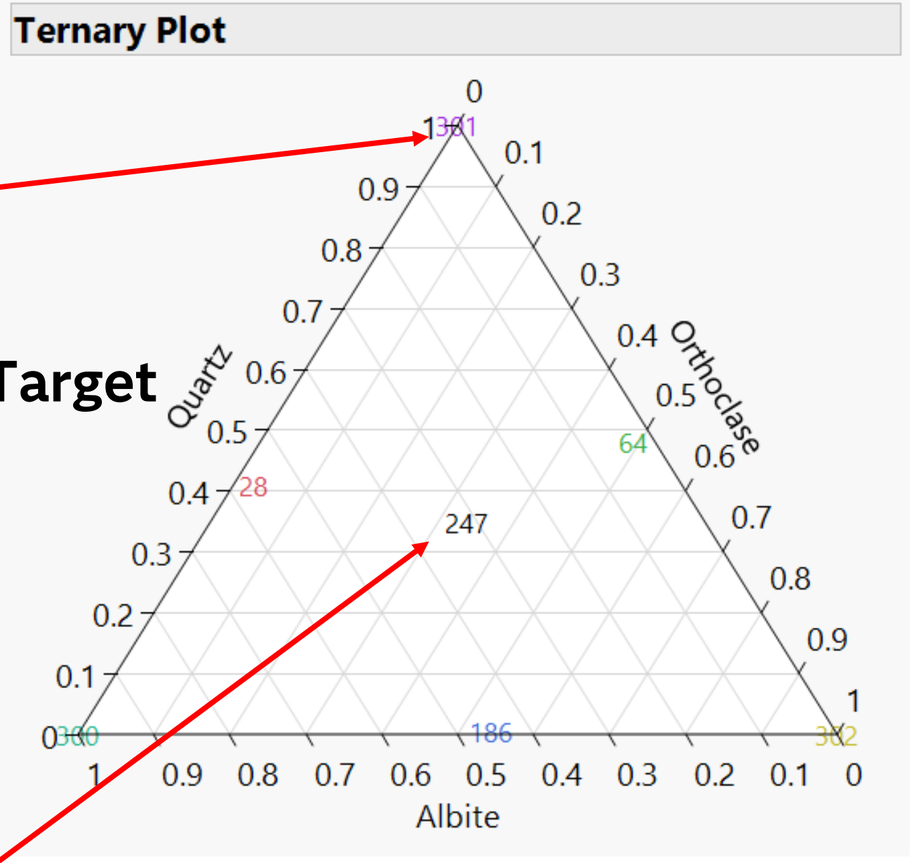
6 on right are resulting subset DOE



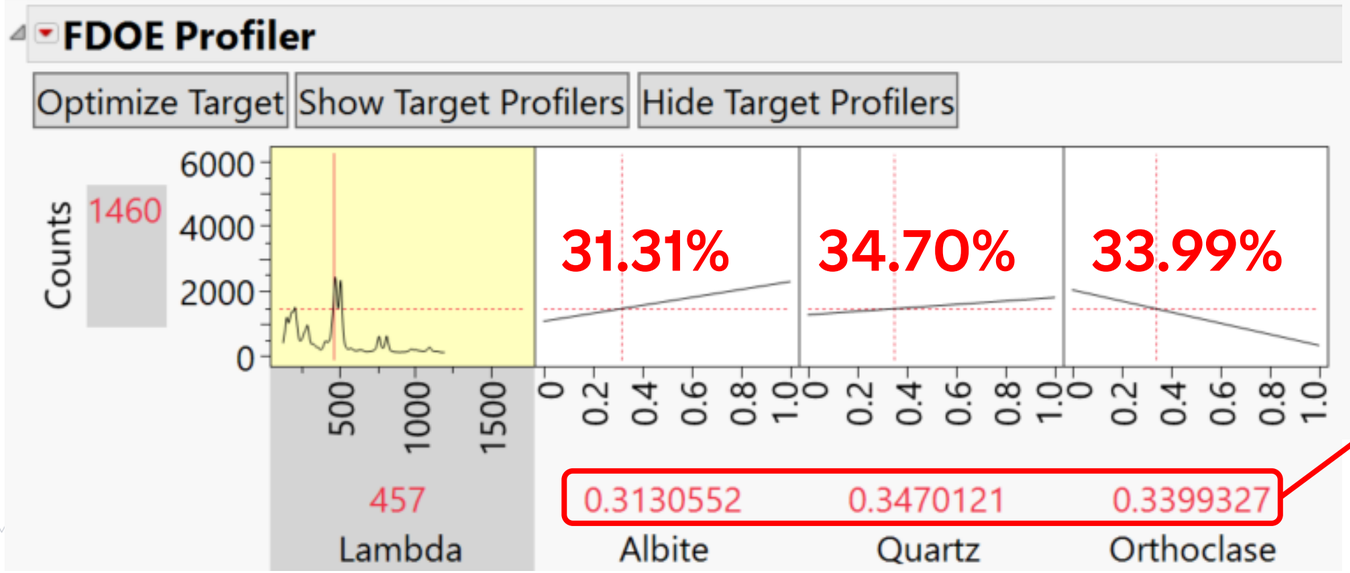


Settings for Raman ID# 301 (0%, 100%, 0%) – before Optimize Target

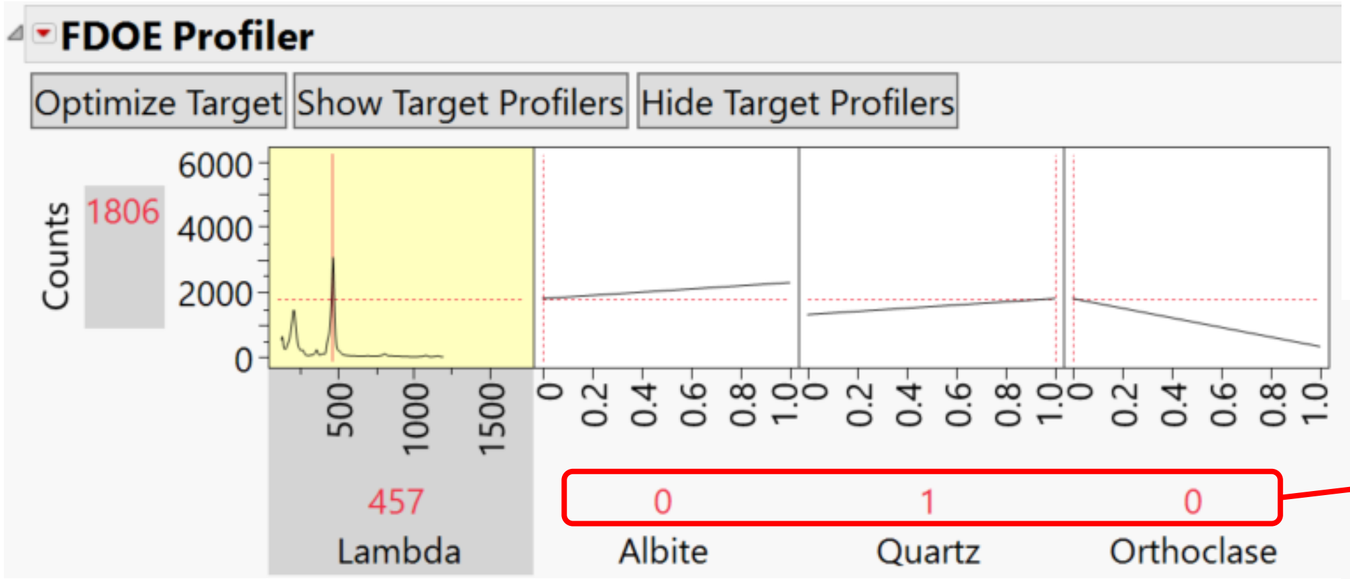


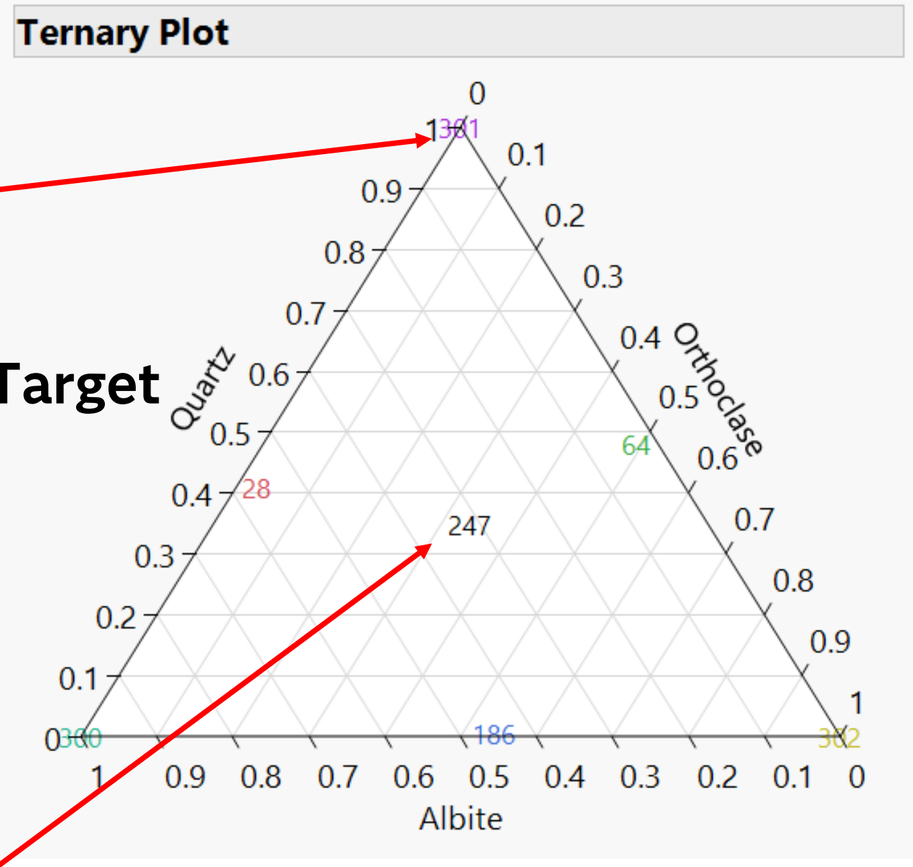


Predicted settings for Raman ID #247 – after Optimize Target

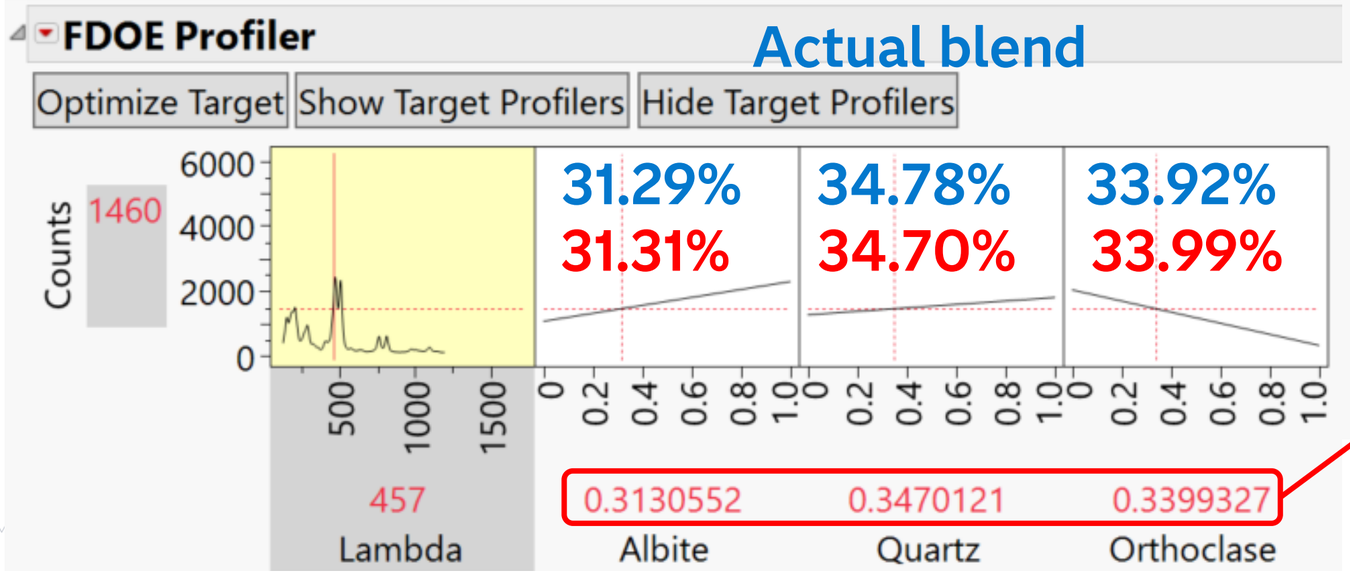


Settings for Raman ID# 301 (0%, 100%, 0%) – before Optimize Target





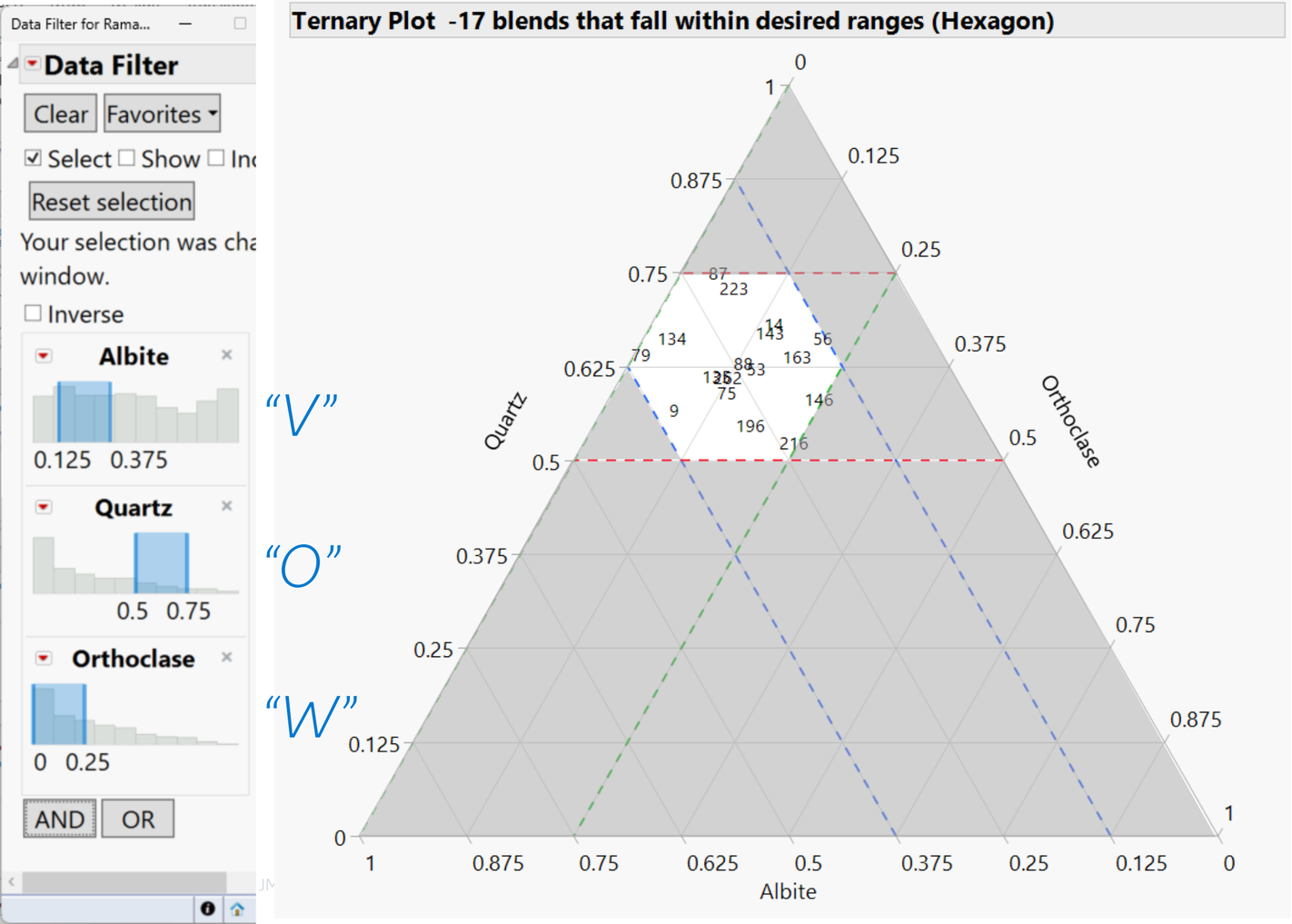
Predicted settings for Raman ID #247 – after Optimize Target



Case 2 - Reanalysis of Raman Spectral Data for 3-Mineral Mixture DOE using FDA

17 *filtered* blends used as candidate trials

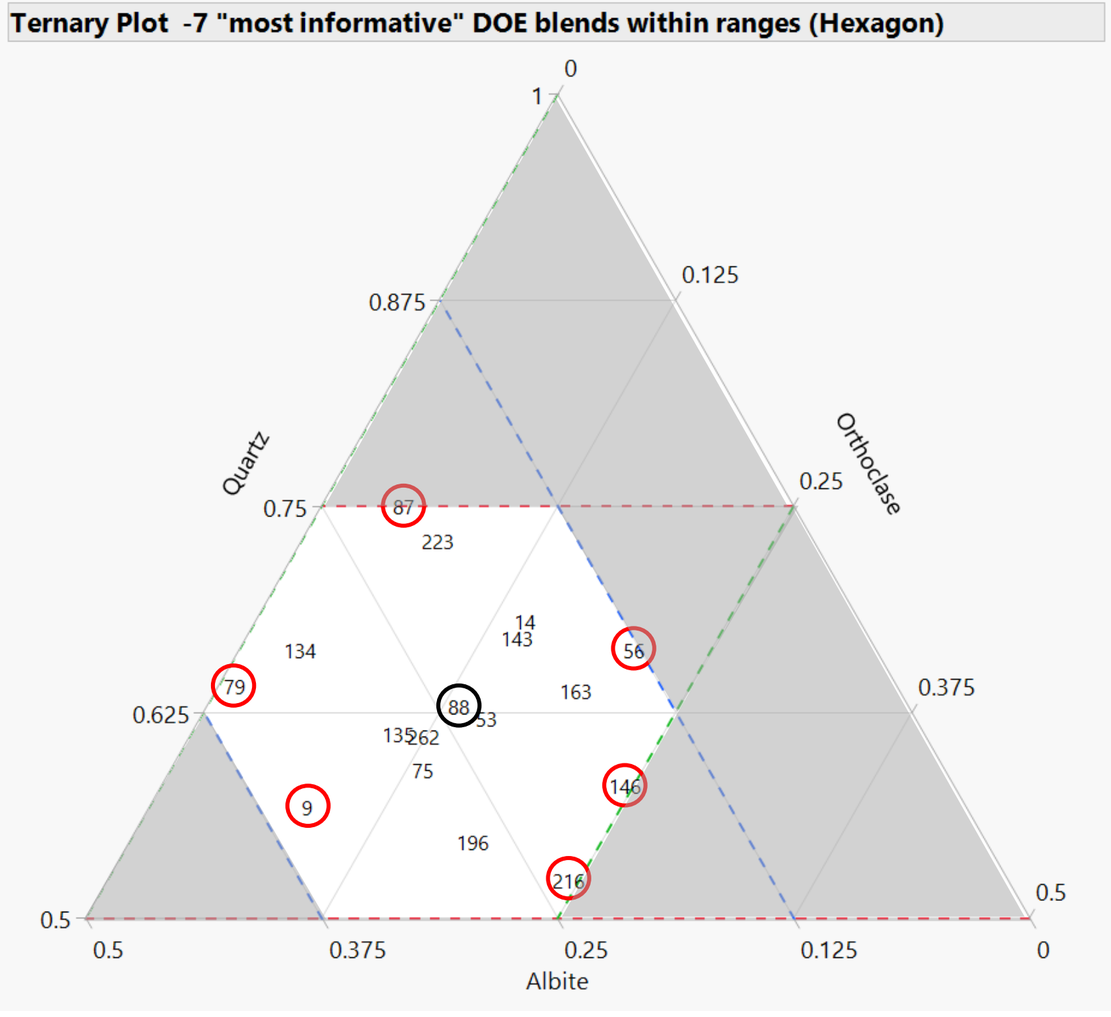
7 circled trials are DOE subset



"V"

"O"

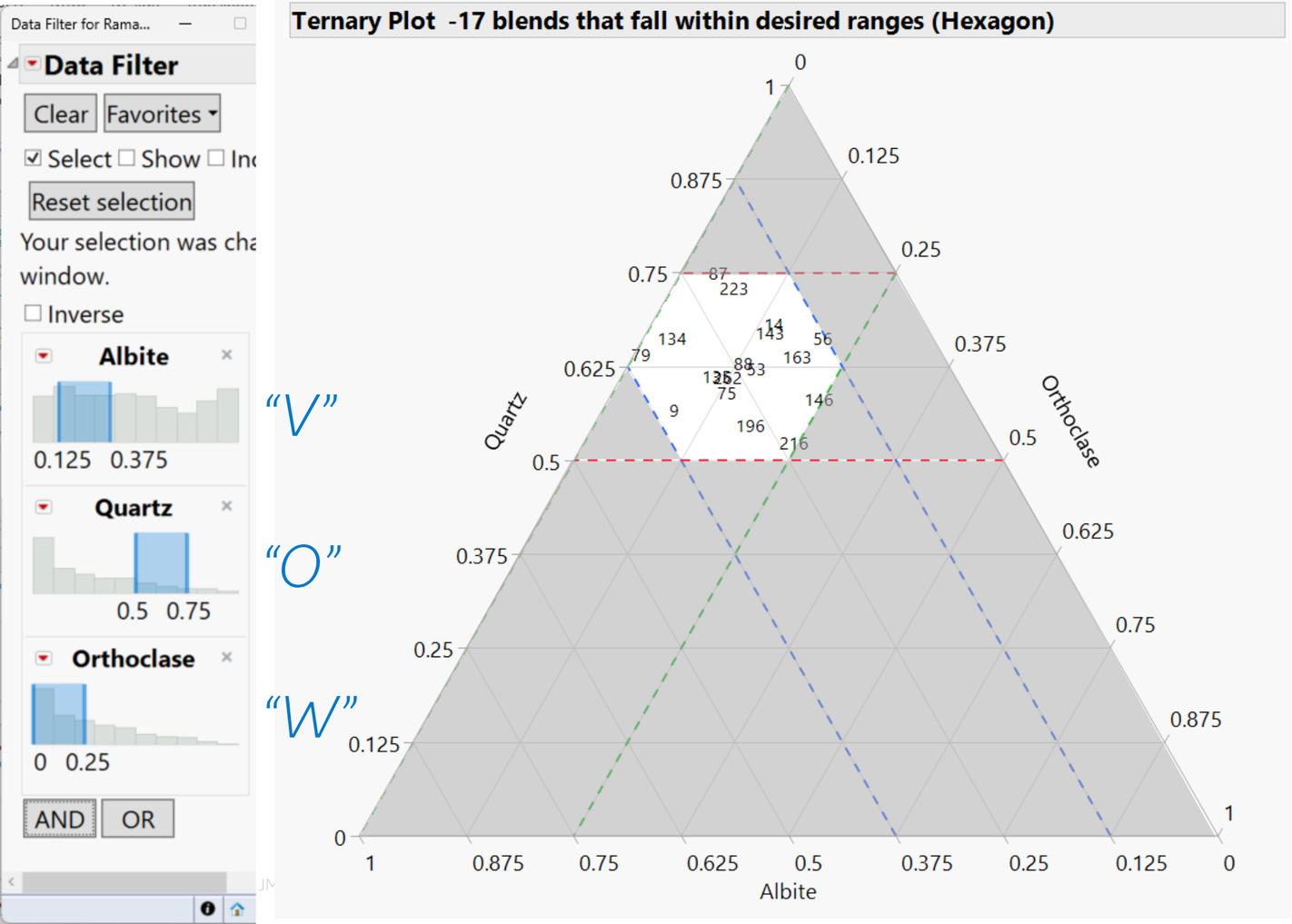
"W"

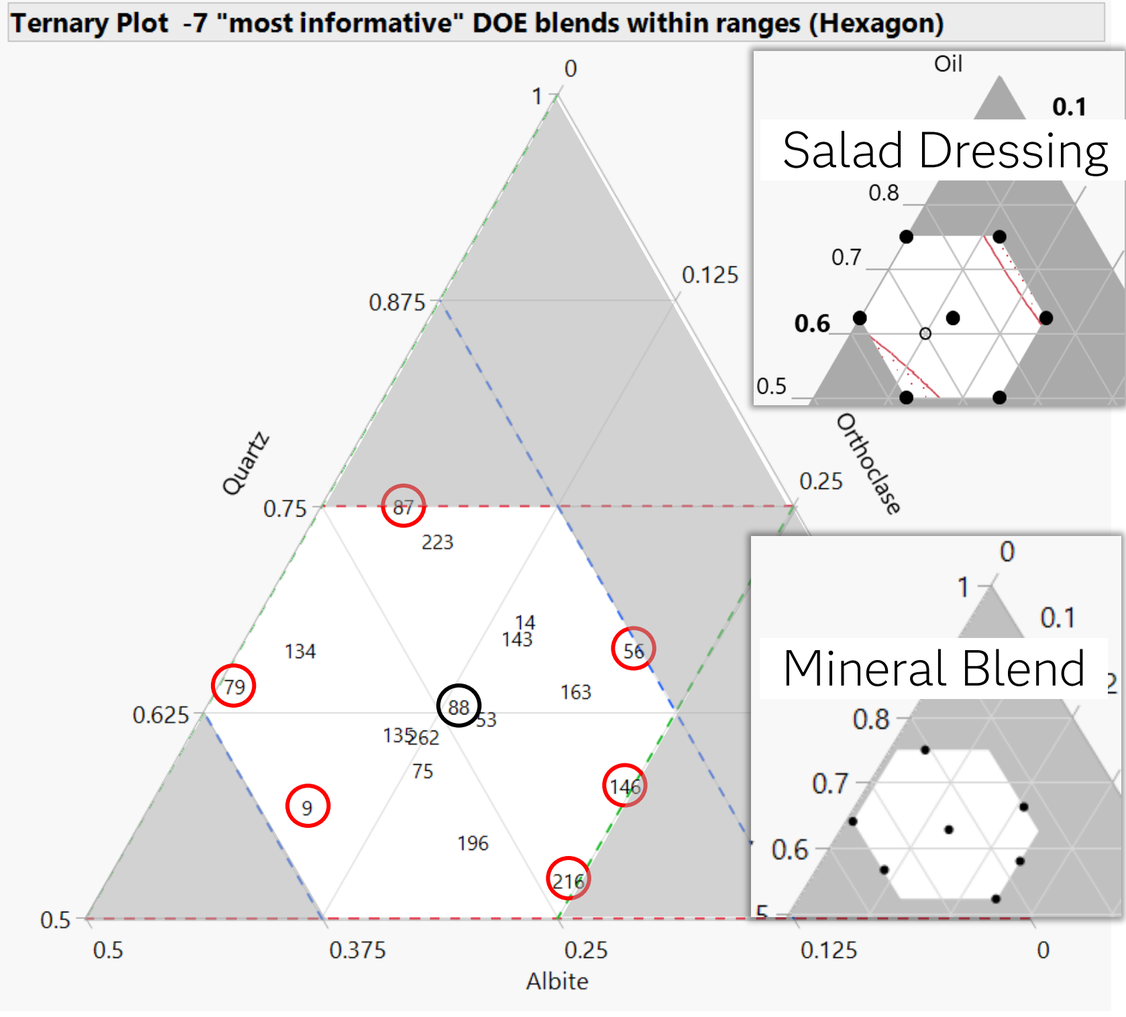


Case 2 - Reanalysis of Raman Spectral Data for 3-Mineral Mixture DOE using FDA

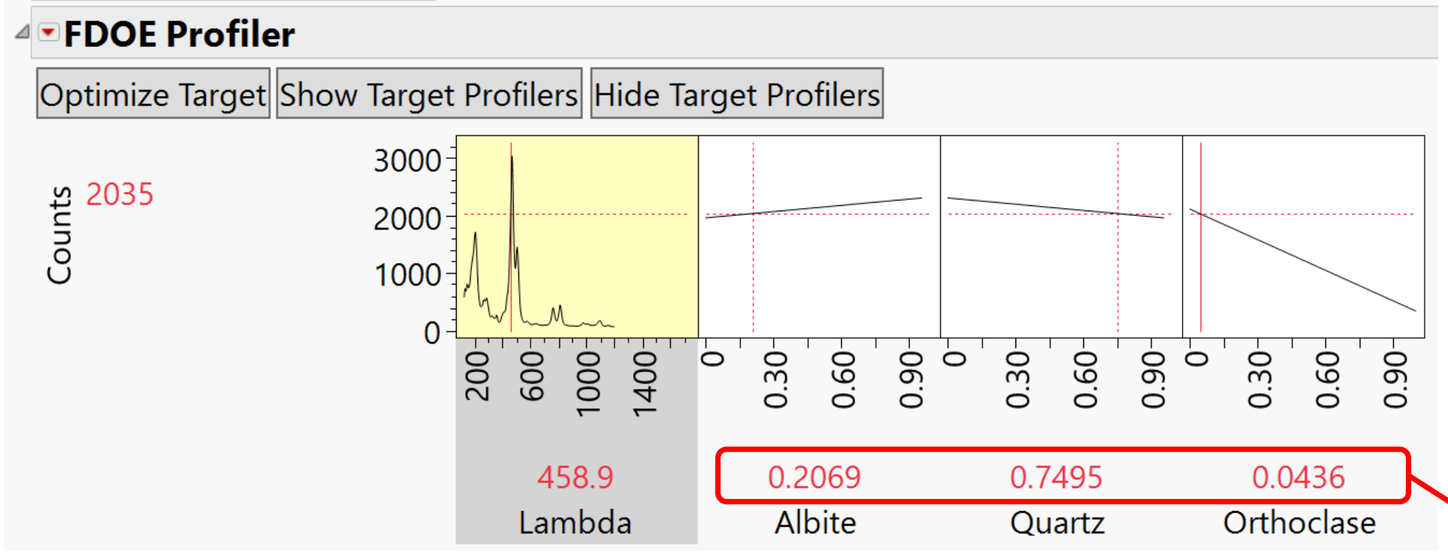
17 *filtered* blends used as candidate trials

7 circled trials are DOE subset

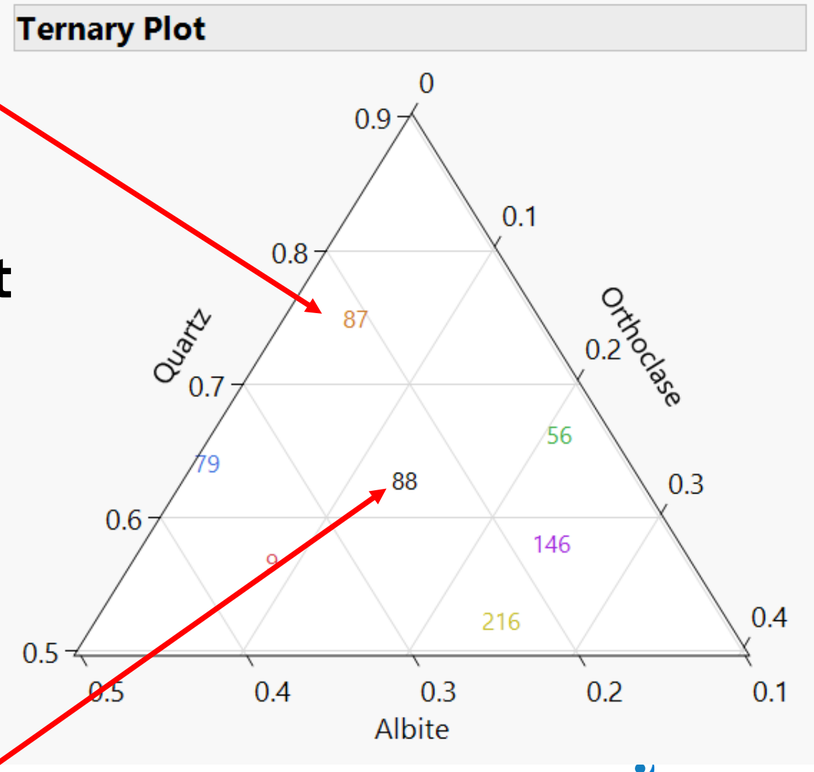




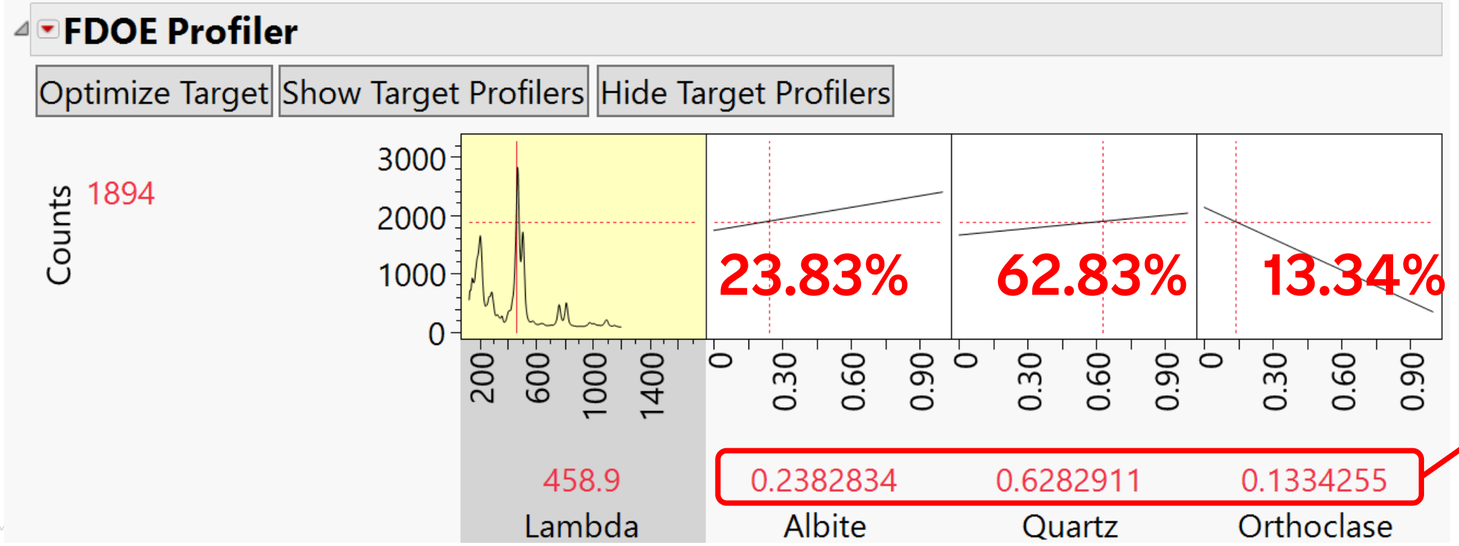
Settings for Raman ID# 87 (20.69%, 74.95%, 4.36%) – before Optimize Target



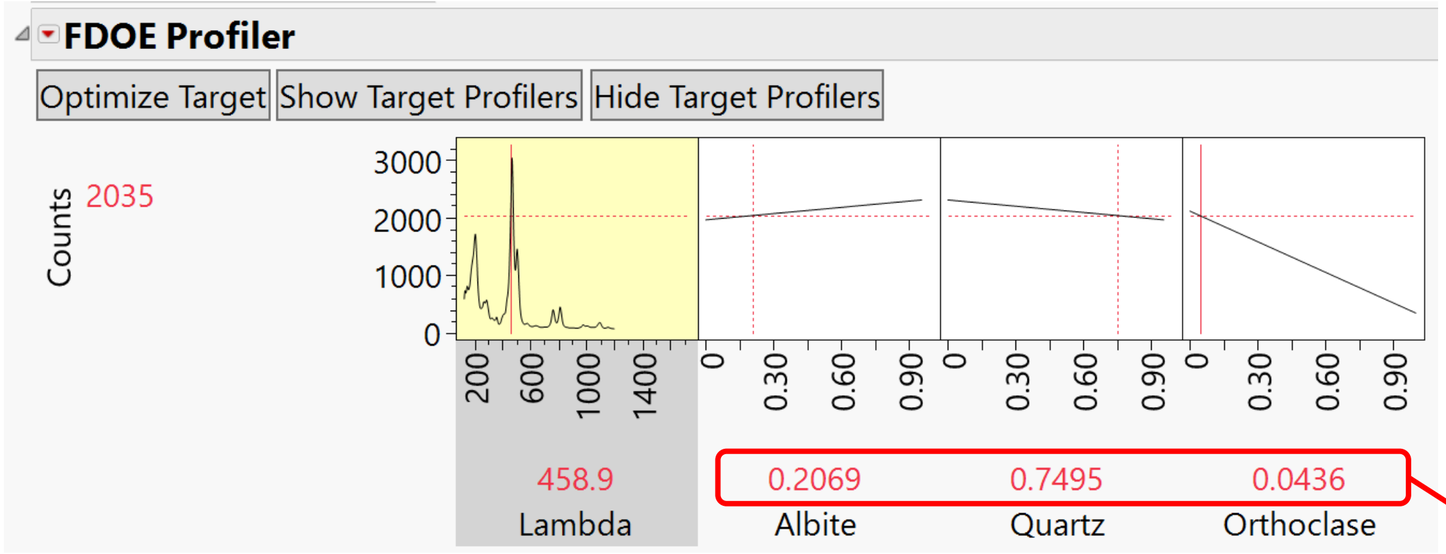
NOTE: Outer 6 points used to Train Model Raman ID# 88 used as Target spectra

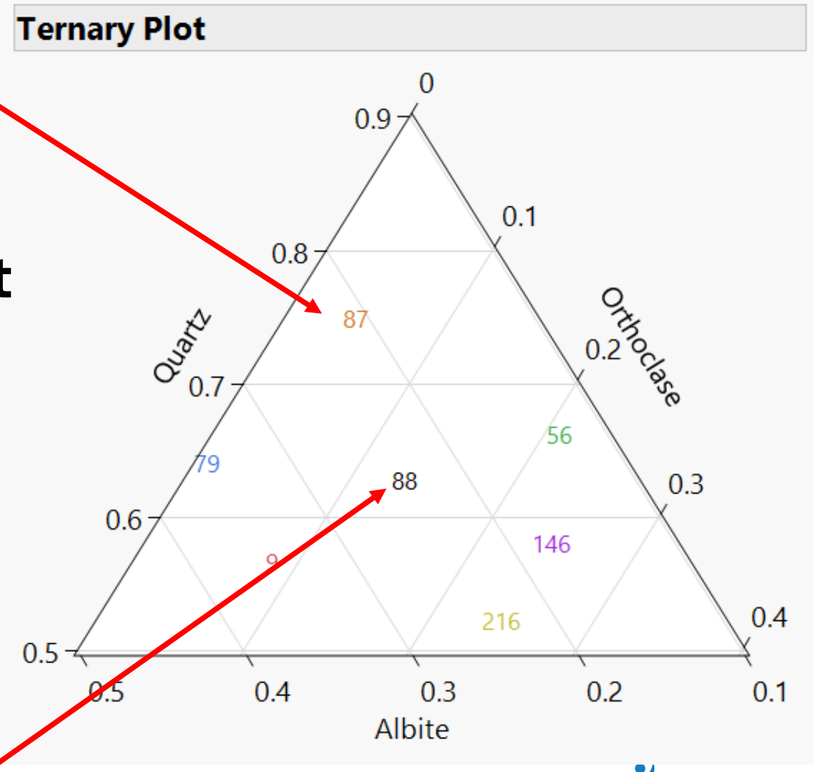


Predicted settings for Raman ID #88 – after Optimize Target

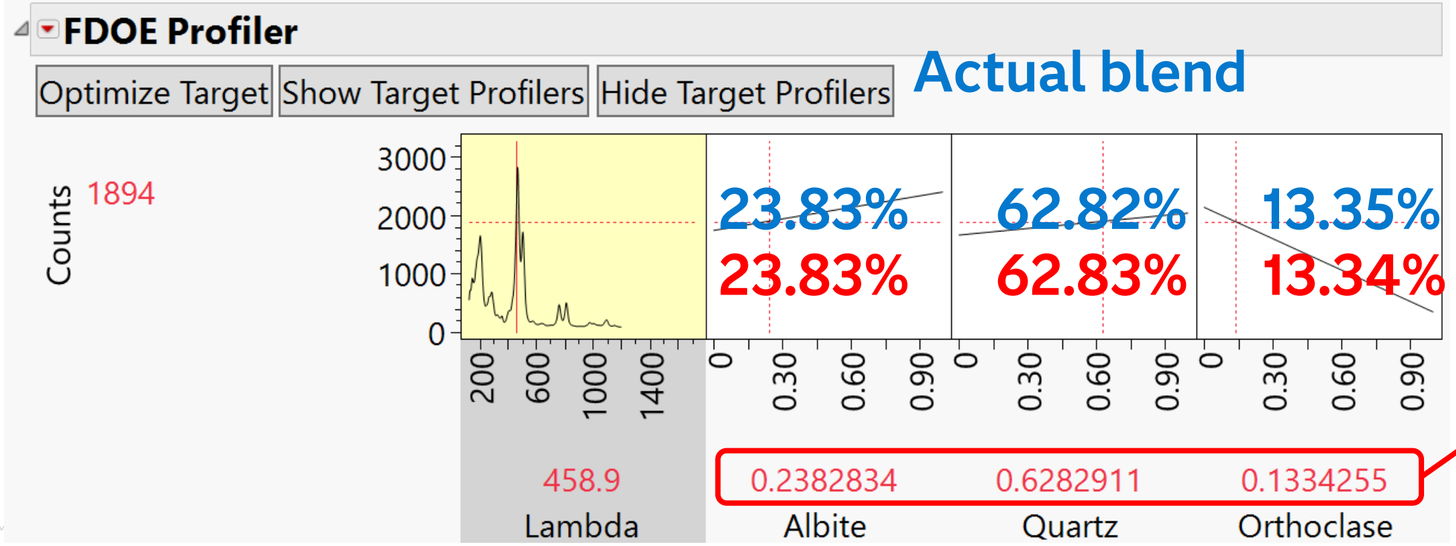


Settings for Raman ID# 87 (20.69%, 74.95%, 4.36%) – before Optimize Target





Predicted settings for Raman ID #88 – after Optimize Target



Let's go to JMP…

- Filter out candidates in an interesting subregion of the full mixture space
- Use Custom DOE to find the most informative subset of 7 trials

Takeaways

1. Use DOE to get most information from fewest trials
2. Use FDA to better model curves and spectra
3. Combine DOE with FDA to predict spectra
4. Combine DOE with FDA to predict formulation (factor settings)
5. Create a DOE by choosing informative subset of trials from candidates

Questions?

Tom Donnelly
Principal Systems Engineer
JMP Defense & Aerospace Team

tom.donnelly@jmp.com

Thanks to my JMP colleagues

Chris Gotwalt, Chief Data Scientist
Ryan Parker, Sr Research Statistician Developer

Developer Tutorial Video on Spectral Data