Analysis Strategies for Constrained Mixture and Mixture Process Experiments Using JMP Pro 14

Dr. Philip J. Ramsey
Principal Lecturer in Statistics
University of New Hampshire
Durham, NH, USA
philip.ramsey@unh.edu

North Haven Group
Owner and Chief Consultant
Brookline, New Hampshire, USA
pjrstats@gmail.com

Introduction

Mixture experiments have a history dating back well over 200 years; e.g., George Pearson, 1784, *Observations and Experiments for investigating the chymical history*.

A small sub-discipline of experimental design: a few key researchers.

John A. Cornell (1941–2016): Mr. Mixtures

Greg F. Piepel: Student of John and prolific researcher and author on mixture designs.

Henri Scheffe (1907–1977): Pioneered modern mixture designs (published key paper 1958)

Introduction

Mixture experiments consist of experiments where the experimental factors are components of a formulation or recipe.

The components can be either a liquid, a gas, or even a solid.

The impact of the mixture components on the responses of interest **depends only on the proportions of the components and not on the amounts.**

If there is a dependency on the amount of the components present, then classical mixture designs are inappropriate – more on this later.

Because the experimental factors are components of a formulation, there are inherent constraints on the settings of these factors.

Introduction

Factorial experimental designs allow independent settings for each of the experimental factors - no explicit constraints exist on the mutual factor settings.

Mixture experimental designs impose constraints on the possible settings for each factor since they are components of a formulation.

Mixture factors can not be independently manipulated.

Mixture experimental designs, with *q* factors, impose the constraints

$$X_1 + X_2 + \ldots + X_q = 1.0 \quad \text{and for} \quad 0 \le X_i \le 1.0 \;\; i=1,\ldots,q$$

Introduction

A type of important effect in mixture experiments is referred to as ***Nonlinear Blending***, or ***NLB*** .

NLB effects represent the departure of the observed response from what one would predict the response to be as a linear, additive combination of pure component blends.

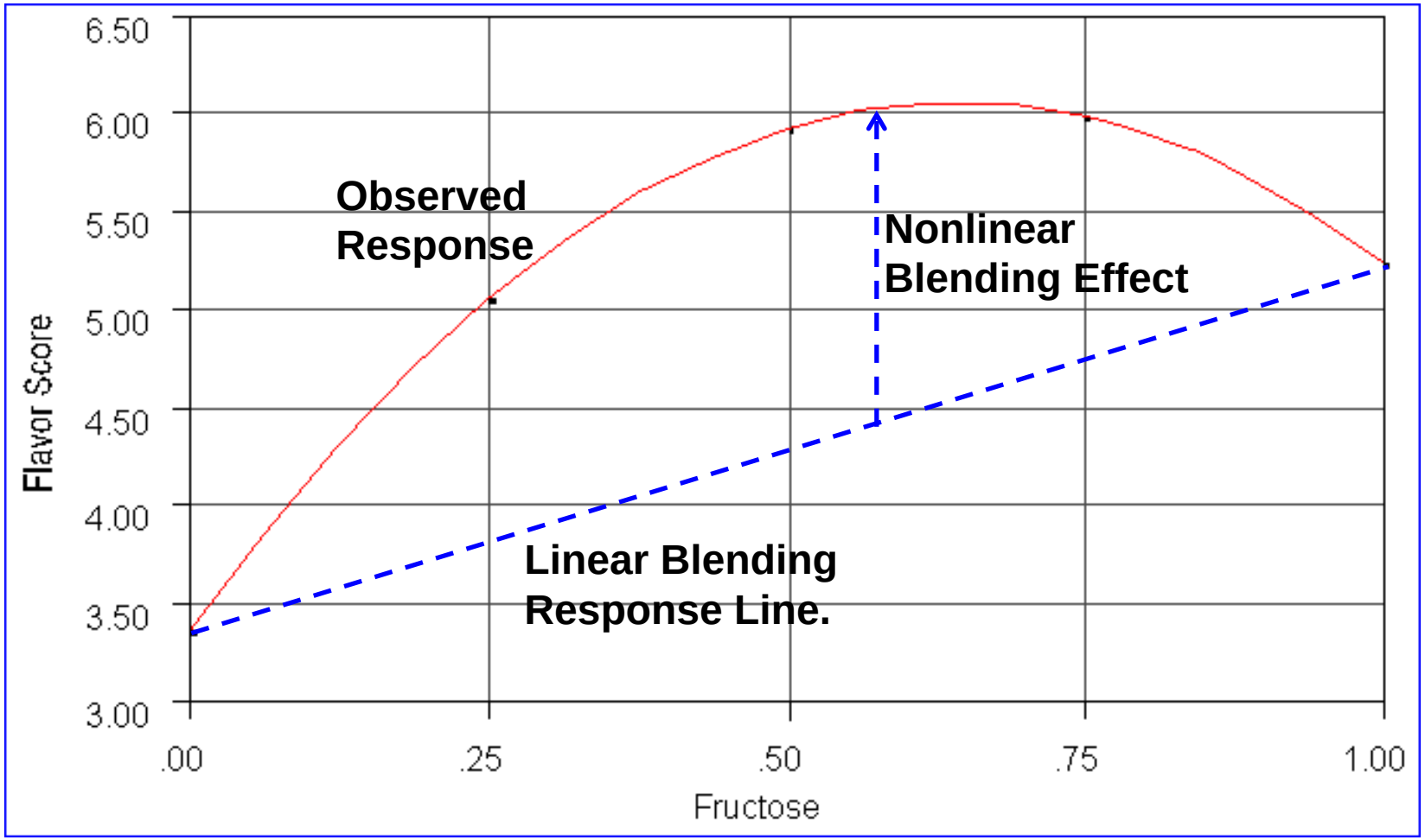
NLB is common, e.g., mixtures of water and alcohol vs boiling point of the mixture.

NLB typically has two forms: **Synergistic** where the response is higher than expected and **Antagonistic** where the response is lower than expected from the pure component blend responses.

Large NLB **effects of second and third order are commonly encountered** in mixture experiments.

Introduction

Example of synergistic **NLB** for a two component mixture experiment on blends of sweeteners glucose and fructose.



Introduction

Additional constraints on mixture factor settings are common.

Regional or inequality constraints (upper and lower bounds)

$$Lower_i < X_i < Upper_i,\ L_i > 0, U_i < 1.0,\ \sum_i X_i = 1.0,\quad i = 1, \ldots, q$$

Example, one would not have cake mixes with 100% eggs or 100% flour, etc., a proportion of some or all ingredients must be present in every mixture trial performed.

Another type of constraint often exists in which bounds are placed on a linear combination of some of the components in the mixture.

Typically, **linear constraints** involve 2 and sometimes 3 components of the mixture and have the general form.

$$X_i + X_j \le a,\ i \ne j,\ a < 1.0$$

Introduction

When regional constraints exist, it is common to rescale the mixture components to help mitigate the multicollinearity issues.

Basically, the component settings are stretched to better fit the encompassing simplex region and reduce multicollinearity.

A common rescaling method is referred to as L-pseudocomponent scaling (can be used with both upper and lower constraints)

$$L_i \le X_i \le 1.0, \quad \sum_i X_i = 1.0, \quad L = 1 - \sum_i L_i, \quad \text{i=1,…,q}$$

$$X_i^* = \frac{X_i - L_i}{1 - \sum_{i=1}^{q} L_i} = \frac{X_i - L_i}{L} \text{ or } X_i = X_i^*(1 - \sum_{i=1}^{q} L_i) + L_i$$

Note, the scaling **method of Cornell and Gorman (JQT, 2002)** is more effective but is not implemented in JMP.

Analyzing Mixture Experiments

Mixture designs are really a type of response surface design and as such the primary goal of a mixture experiment is to develop a model that satisfactorily predicts future performance of the response(s).

The model is subsequently used to characterize the behavior of the response(s) over the constrained mixture space – the models are commonly used for optimization and sensitivity analysis.

Selecting satisfactory predictive models require that we divide the data into a **training** set to fit the model and a **validation** set, not used to fit the model, that assesses the prediction capability.

Predictive models are selected that have smallest prediction error on the validation set – should predict future performance the best.

Unfortunately, in designed experiments rarely is there sufficient data to form a training and validation set.

So how does one select a best predictive model in DOE?

Analyzing Mixture Experiments

The model selection problem is particularly difficult in mixtures.

- The existence of constraints results in a great deal of multicollinearity.
- The multicollinearity results in correlations among the estimated effects, especially NLB effects; standard errors of the estimates are generally quite large.
- Traditional model selection techniques using p-values are not useful for mixtures; p-values are distorted by even small amounts of multicollinearity and are misleading.
- The pure component terms must be forced into all models.
- Mixture models have 0 intercept – sometimes a problem.
- The designs are typically very efficient and even supersaturated depending upon the mixture model of interest.

Analyzing Mixture Experiments

Today's powerful computers, with double precision, and modern statistical software such as JMP provide us with better capabilities to analyze and build predictive models for mixtures.

We will investigate and illustrate a couple methods (many more are possible, users should try an array of methods).

- Traditional **Forward selection** using the pseudo factor method of Miller (1992) – interesting and unappreciated approach.
- Traditional **Forward selection** and **Pruned Forward** using fractionally weighted bootstrapping and auto-validation (Gotwalt and Ramsey, 2018).

The Generalized Regression platform in JMP Pro makes it straightforward to implement these methods for mixtures.

Note, I do not recommend **All Possible Models** for mixtures in part due to the need to force the pure component terms in every model

Miller's Pseudo Factor Method

Calculate the largest possible number of effects *p* that could occur in a mixture model; includes cross products, polynomials, etc.

Create an additional set of *p* random factors or pseudo factors – Miller used random uniformly distributed factors.

Define the largest possible model in the mixture factors and then add the additional set of pseudo factors.

Under the assumption that all effects are 0, the probability of a mixture effect or pseudo factor entering the model is the same, hence the use of *p* pseudo factors – Miller never explained his rationale.

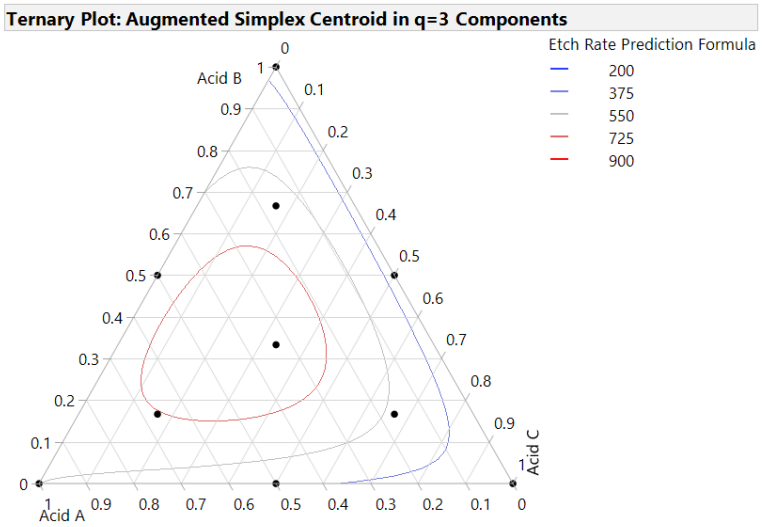
Use Forward Selection with say AICc or BIC as the objective function and stop the selection process when either the first pseudo factor enters the model or the fitting criterion is minimized.

For large mixture problems, this approach can become unwieldy.

Miller's Pseudo Factor Method

Example (due to Montgomery): In semiconductor manufacture, silicon wafers are acid etched prior to a metalization step in the process.

An experiment has been performed by research scientists to find the optimum composition of etchant – the composition yielding the highest etching rate.



The etchant is composed of three acids:

A = nitric,

B = hydrochloric, and

C = Phosphoric.

A simplex centroid design was used and the largest possible model contains 7 effects: 3 pure component terms, 3 two-way NLB terms, and 1 three-way NLB term.

Miller's Pseudo Factor Method

Below is the design with pseudo factors added and the response.

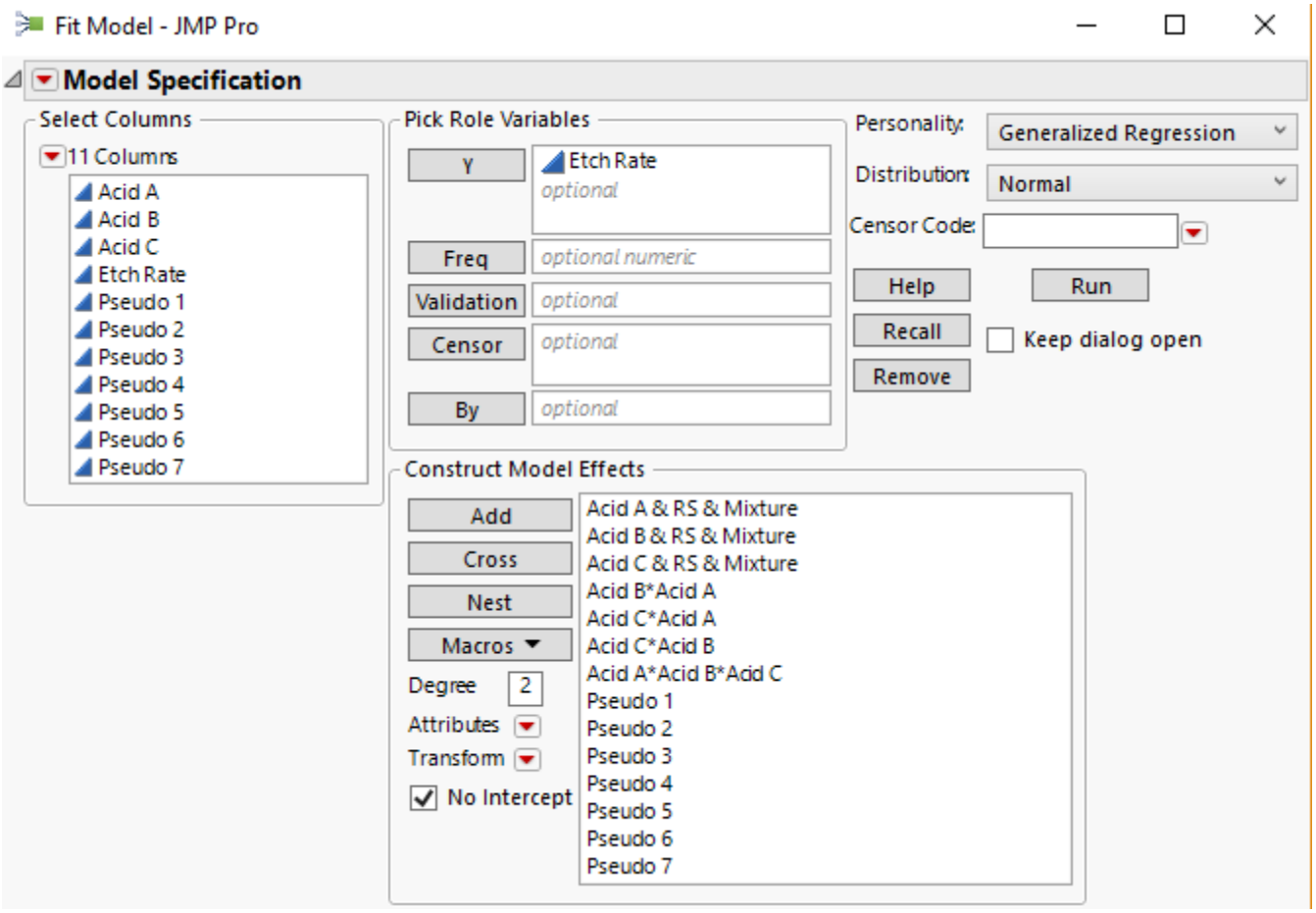
	Acid A	Acid B	Acid C	Etch Rate	Pseudo 1	Pseudo 2	Pseudo 3	Pseudo 4	Pseudo 5	Pseudo 6	Pseudo 7
1	1	0	0	540	0.7773	0.3166	0.3299	0.1219	0.9820	0.6059	0.5861
2	1	0	0	560	0.4977	0.5328	0.6055	0.2501	0.5196	0.6867	0.4053
3	0	1	0	330	0.4836	0.9084	0.5460	0.4851	0.9114	0.2204	0.8571
4	0	1	0	350	0.4890	0.6580	0.2410	0.5191	0.2306	0.7201	0.0553
5	0	0	1	295	0.7780	0.5340	0.2090	0.2676	0.1940	0.2137	0.7248
6	0	0	1	260	0.4901	0.8901	0.8114	0.8635	0.9882	0.2655	0.0061
7	0.5	0.5	0	610	0.1218	0.3392	0.2024	0.5710	0.6999	0.4234	0.9175
8	0.5	0	0.5	425	0.7579	0.4642	0.8581	0.8342	0.5390	0.5561	0.2514
9	0	0.5	0.5	330	0.4038	0.4675	0.6753	0.5720	0.5669	0.3338	0.9574
10	0.33333	0.33333	0.33333	800	0.6825	0.3021	0.8554	0.5710	0.7260	0.4933	0.6600
11	0.33333	0.33333	0.33333	850	0.5378	0.5762	0.0210	0.6334	0.8094	0.5212	0.9652
12	0.66667	0.16667	0.16667	710	0.2096	0.9733	0.8266	0.3688	0.9245	0.9714	0.8881
13	0.16667	0.66667	0.16667	640	0.3233	0.8396	0.5685	0.4570	0.3524	0.3550	0.9580
14	0.16667	0.16667	0.66667	460	0.7787	0.5607	0.1269	0.9727	0.6522	0.2195	0.0290

In Fit Model define a mixture response surface model and add the three-way NLB term, then add the 7 pseudo factors.

We will use the Generalized Regression platform to do the forward selection analysis; if you do not have JMP Pro use the Stepwise platform.

Miller's Pseudo Factor Method

Below is a screenshot of the Fit Model launch dialog for the analysis.



Miller's Pseudo Factor Method

To the right is a screenshot of the Gen Reg Model Launch configuration.

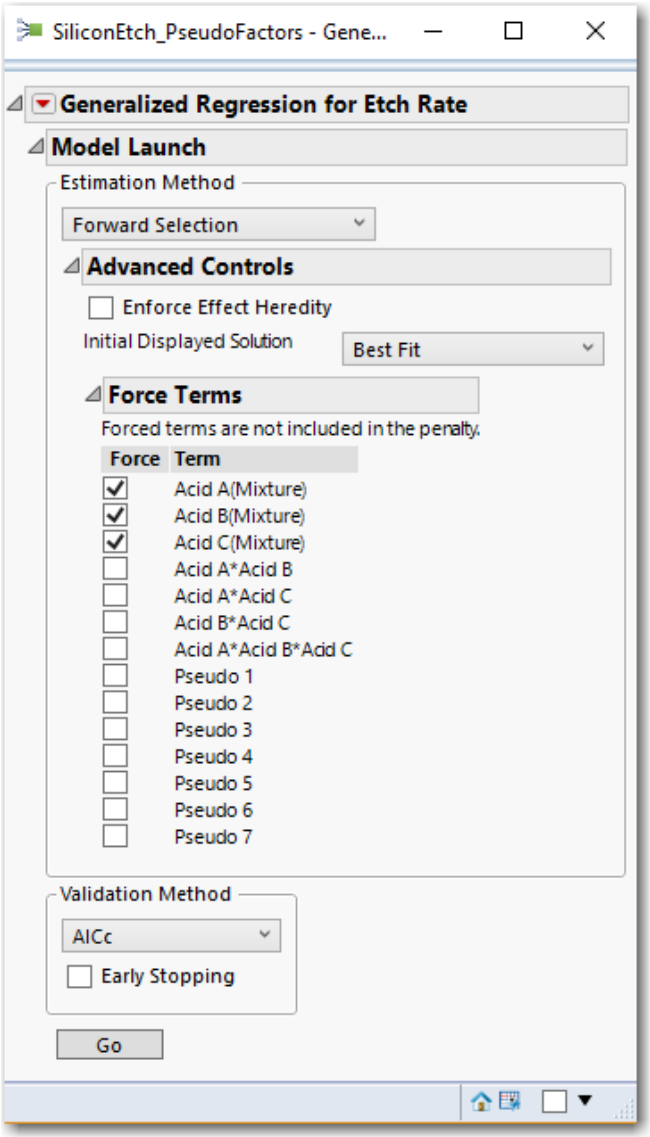
Of course other options are possible.

Important to **force in the pure component terms!**

I **do not recommend enforcing heredity in predictive model building** whether a mixture model or not; can lead to over fitting.

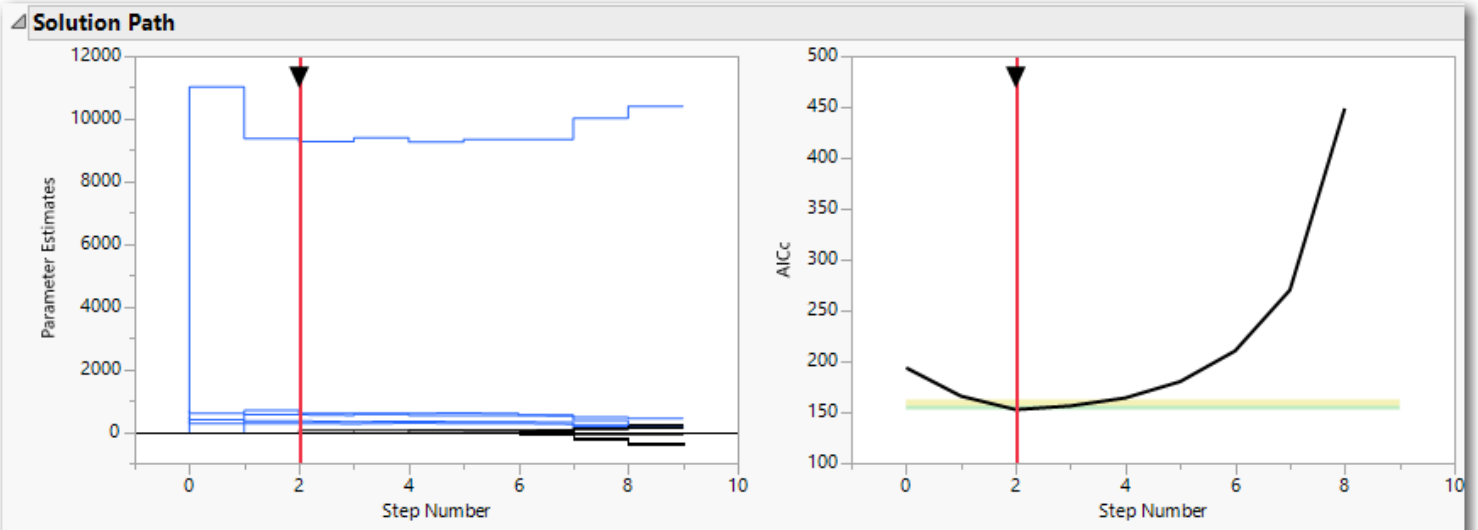
This is controversial.

Here we use **AICc** as the objective function for model selection; again you can use other options.



Miller's Pseudo Factor Method

The algorithm stops after two steps, the three pure component effects the **A*B** NLB term and the **A*B*C** NLB terms define the model.

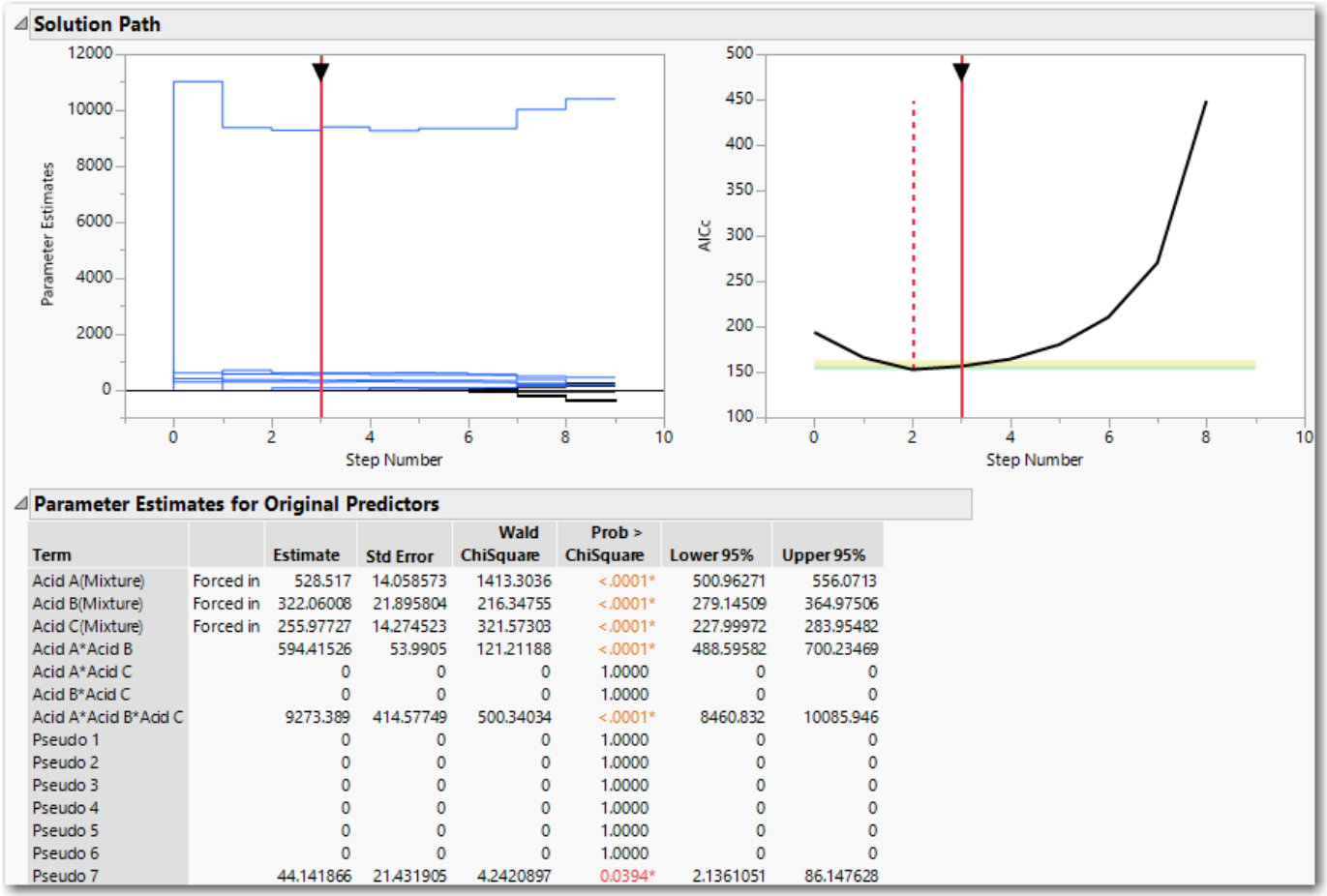


Parameter Estimates for Original Predictors

Term		Estimate	Std Error	Wald ChiSquare	Prob > ChiSquare	Lower 95%	Upper 95%
Acid A(Mixture)	Forced in	549.10103	7.0280392	6104.3041	<.0001*	535.32632	562.87573
Acid B(Mixture)	Forced in	348.10103	9.300977	1400.7269	<.0001*	329.87145	366.33061
Acid C(Mixture)	Forced in	270.75231	15.048573	323.70822	<.0001*	241.25765	300.24697
Acid A*Acid B		684.7659	54.012093	160.73196	<.0001*	578.90414	790.62766
Acid A*Acid C		0	0	0	1.0000	0	0
Acid B*Acid C		0	0	0	1.0000	0	0
Acid A*Acid B*Acid C		9372.7367	562.10762	278.03142	<.0001*	8271.026	10474.447
Pseudo 1		0	0	0	1.0000	0	0
Pseudo 2		0	0	0	1.0000	0	0
Pseudo 3		0	0	0	1.0000	0	0
Pseudo 4		0	0	0	1.0000	0	0
Pseudo 5		0	0	0	1.0000	0	0
Pseudo 6		0	0	0	1.0000	0	0
Pseudo 7		0	0	0	1.0000	0	0

Miller's Pseudo Factor Method

Using the model selection slider (red vertical line) we manually force a third selection step. Notice a pseudo factor enters the model, so we stop at two steps.

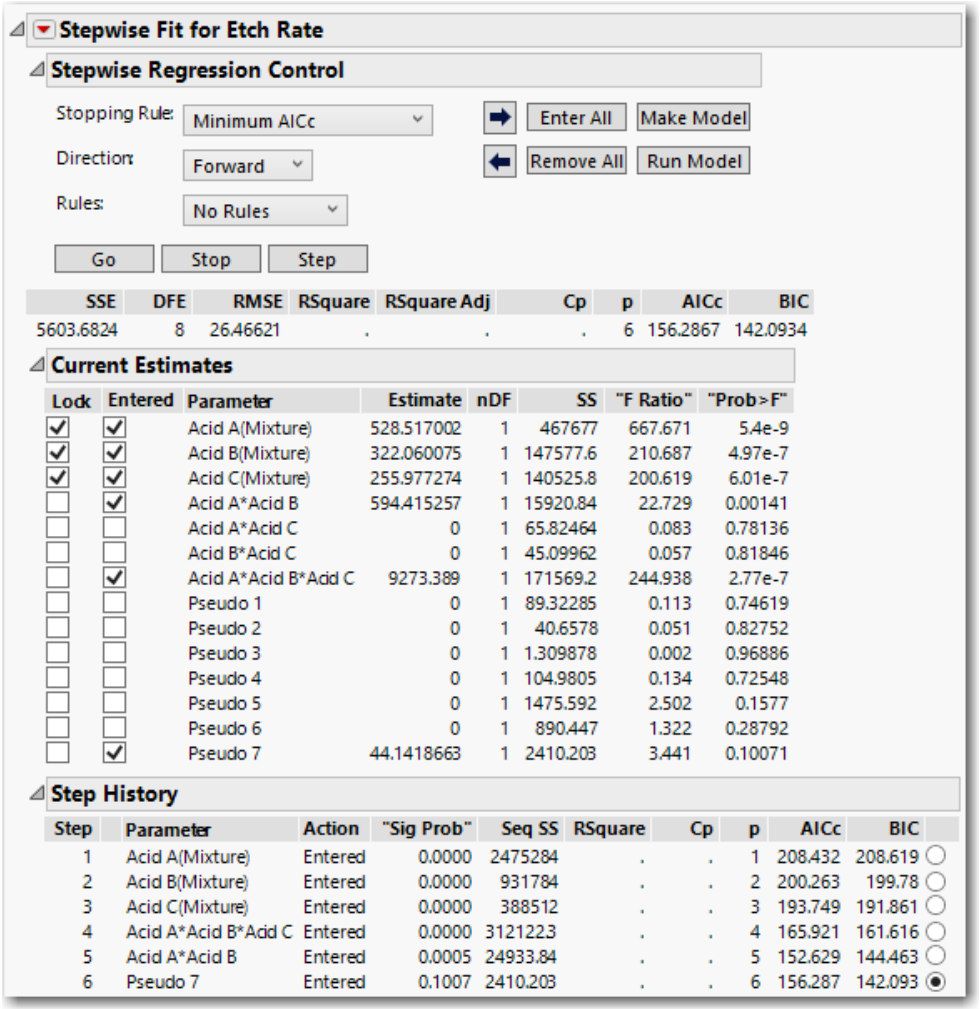


Parameter Estimates for Original Predictors

Term		Estimate	Std Error	Wald ChiSquare	Prob > ChiSquare	Lower 95%	Upper 95%
Acid A(Mixture)	Forced in	528.517	14.058573	1413.3036	<.0001*	500.96271	556.0713
Acid B(Mixture)	Forced in	322.06008	21.895804	216.34755	<.0001*	279.14509	364.97506
Acid C(Mixture)	Forced in	255.97727	14.274523	321.57303	<.0001*	227.99972	283.95482
Acid A*Acid B		594.41526	53.9905	121.21188	<.0001*	488.59582	700.23469
Acid A*Acid C		0	0	0	1.0000	0	0
Acid B*Acid C		0	0	0	1.0000	0	0
Acid A*Acid B*Acid C		9273.389	414.57749	500.34034	<.0001*	8460.832	10085.946
Pseudo 1		0	0	0	1.0000	0	0
Pseudo 2		0	0	0	1.0000	0	0
Pseudo 3		0	0	0	1.0000	0	0
Pseudo 4		0	0	0	1.0000	0	0
Pseudo 5		0	0	0	1.0000	0	0
Pseudo 6		0	0	0	1.0000	0	0
Pseudo 7		44.141866	21.431905	4.2420897	0.0394*	2.1361051	86.147628

Miller's Pseudo Factor Method

For reference below is a screenshot of the same analysis performed in the **Stepwise** platform; the results are identical to Gen Reg.



Stepwise Fit for Etch Rate

Stepwise Regression Control

Stopping Rule: Minimum AICc — Enter All — Make Model

Direction: Forward — Remove All — Run Model

Rules: No Rules

Go — Stop — Step

SSE	DFE	RMSE	RSquare	RSquare Adj	Cp	p	AICc	BIC
5603.6824	8	26.46621	.	.	.	6	156.2867	142.0934

Current Estimates

Lock	Entered	Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
☑	☑	Acid A(Mixture)	528.517002	1	467677	667.671	5.4e-9
☑	☑	Acid B(Mixture)	322.060075	1	147577.6	210.687	4.97e-7
☑	☑	Acid C(Mixture)	255.977274	1	140525.8	200.619	6.01e-7
☐	☑	Acid A*Acid B	594.415257	1	15920.84	22.729	0.00141
☐	☐	Acid A*Acid C	0	1	65.82464	0.083	0.78136
☐	☐	Acid B*Acid C	0	1	45.09962	0.057	0.81846
☐	☑	Acid A*Acid B*Acid C	9273.389	1	171569.2	244.938	2.77e-7
☐	☐	Pseudo 1	0	1	89.32285	0.113	0.74619
☐	☐	Pseudo 2	0	1	40.6578	0.051	0.82752
☐	☐	Pseudo 3	0	1	1.309878	0.002	0.96886
☐	☐	Pseudo 4	0	1	104.9805	0.134	0.72548
☐	☐	Pseudo 5	0	1	1475.592	2.502	0.1577
☐	☐	Pseudo 6	0	1	890.447	1.322	0.28792
☐	☑	Pseudo 7	44.1418663	1	2410.203	3.441	0.10071

Step History

Step	Parameter	Action	"Sig Prob"	Seq SS	RSquare	Cp	p	AICc	BIC
1	Acid A(Mixture)	Entered	0.0000	2475284	.	.	1	208.432	208.619
2	Acid B(Mixture)	Entered	0.0000	931784	.	.	2	200.263	199.78
3	Acid C(Mixture)	Entered	0.0000	388512	.	.	3	193.749	191.861
4	Acid A*Acid B*Acid C	Entered	0.0000	3121223	.	.	4	165.921	161.616
5	Acid A*Acid B	Entered	0.0005	24933.84	.	.	5	152.629	144.463
6	Pseudo 7	Entered	0.1007	2410.203	.	.	6	156.287	142.093

FWB and Autovalidation

Gotwalt and Ramsey (2018 Discovery Frankfurt, Cary) propose a new method of model selection based upon fractionally weighted bootstrapping and autovalidation; slides on JMP Communities.

The method also makes use of a pseudo or null factor.

Recall, a problem in building predictive models in DOE is the lack of a validation set to assess prediction capability.

Autovalidation uses a copy of the original data as a validation set; sounds crazy but give it a chance.

The key is that the observations in the original or training set and validation set (copy) are differentially weighted such that the original or training set is anticorrelated with the validation copy set.

Thousands of bootstrap repetitions of model selection are performed and on each bootstrap trial new weights are generated randomly.

FWB and Autovalidation

The method is straightforward.

- In the Fit Model launch dialog define the largest model of interest and then add a null or pseudo factor to the model.
- Use Generalized Regression, pick the desired model selection algorithm, force the pure components into the model, do not enforce heredity (same as Miller's method).
- Perform the model selection, then use the **Simulate** function in JMP Pro to perform thousands of fractionally weighted bootstrap repetitions of the model selection.
- Gen Reg keeps track of which effects enter the model on each repetition.
- All effects entering the model with higher frequency then the null factor are considered potential model terms.

Case Study: The Waste Glass Experiment

Redgate, Piepel, et. al., 1992, discuss a 10 component (oxides), highly constrained mixture experiment to study a vitrification process to transform high level radioactive waste to a borosilicate glass form for subsequent permanent sequester.

The response of interest for this case study is **Viscosity** @1150^{o}C.

The design consisted of 81 trials (68 unique settings and 13 replicate trials). The authors do not discuss how the design was generated; most likely an early form of D optimal design.

Both regional and linear (multicomponent) constraints exist.

The design was created to **estimate pure component effects and two-way NLB effects**, however assuming some amount of effect sparsity, higher order effects could be considered.

For $q = 10$ components the full 2nd order Scheffe model has **55 terms**.

Case Study: The Waste Glass Experiment

The oxide components and **regional constraints** are given in the table below.

Oxide	Lower Bound	Upper Bound
SiO_2	0.42	0.57
B_2O_3	0.05	0.20
Na_2O	0.05	0.20
Li_2O	0.01	0.07
CaO	0.00	0.10
MgO	0.00	0.08
Fe_2O_3	0.02	0.15
Al_2O_3	0.00	0.15
ZrO_2	0.00	0.13
Remainder	0.01	0.10

Case Study: The Waste Glass Experiment

The linear constraints involving the oxides are as follows:

$$SiO_2 - 3Al_2O_3 \geq 0.0$$
$$MgO + CaO \leq 0.08$$
$$Fe_2O_3 + Al_2O_3 + ZrO_2 + \text{Remainder} \leq 0.21$$
$$Al_2O_3 + ZrO_2 \leq 0.14$$
$$MgO + CaO + ZrO_2 \leq 0.18$$

The response limits for Viscosity are [1.5, 12].

We begin by analyzing the full second order Scheffe model using autovalidation and the Generalized Regression platform.

In the original paper the authors used Miller's method (55 pseudo factors added to the model) as well as traditional Stepwise selection without pseudo factors.

Case Study: The Waste Glass Experiment

The traditonal 2nd order Scheffe model (Scheffe, 1958) has the form

$$Y_{ijk} = \sum_{i=1}^{q} \beta_i X_i + \sum_{i<j} \sum \beta_{ij} X_i X_j$$

Notice, that no pure quadratic terms are included. It can be shown (Smith, 2005) that a pure quadratic mixture term can be re-expressed as combination of a linear term and string of cross product terms.

Therefore, if we add all of the pure quadratics to the mixture model, the model matrix is singular; Scheffe elected to drop the quadratics.

However, a valid mixture model can contain a combination of cross product terms and pure quadratic terms, just not all of them, and actually fit better.

Redgate and Piepel included cross products and pure quadratics in their model selection strategy in order to improve the fit.

Case Study: The Waste Glass Experiment

The quadratic Scheffe model had considerable lack of fit, as did the models selected by Redgate and Piepel, so we next tried fitting Scheffe special cubic models.

$$Y_{ijk} = \sum_{i=1}^{q} \beta_i X_i + \sum\sum_{i<j} \beta_{ij} X_i X_j + \sum\sum\sum_{i<j<k} \beta_{ijk} X_i X_j X_k + \varepsilon_{ijk}$$

The original design was not generated to fit special cubic terms, however using a forward selection strategy and assuming some effect sparsity it is possible to fit special cubic models.

The special cubic model in 10 components has **n = 166** terms apart from the pure component terms. Obvioiusly the design is supersatured for this model.

The autovalidation method was repeated, but this time using a special cubic model in the ten components.

Case Study: The Waste Glass Experiment

Quick diversion: How might we display the results of the simulation?

One nice approach is odds ratios. If p_i is the proportion of times the i^{th} effect entered the model during the simulation, then $p_i/(1-p_i)$ is the estimated odds of entering the model.

Under a null hypothesis that all effects are independent and inactive p_0 is the probability of entering the model purely at random and $p_0/(1-p_0)$ is the odds of randomly entering the model.

The odds ratio for the i^{th} effect is then

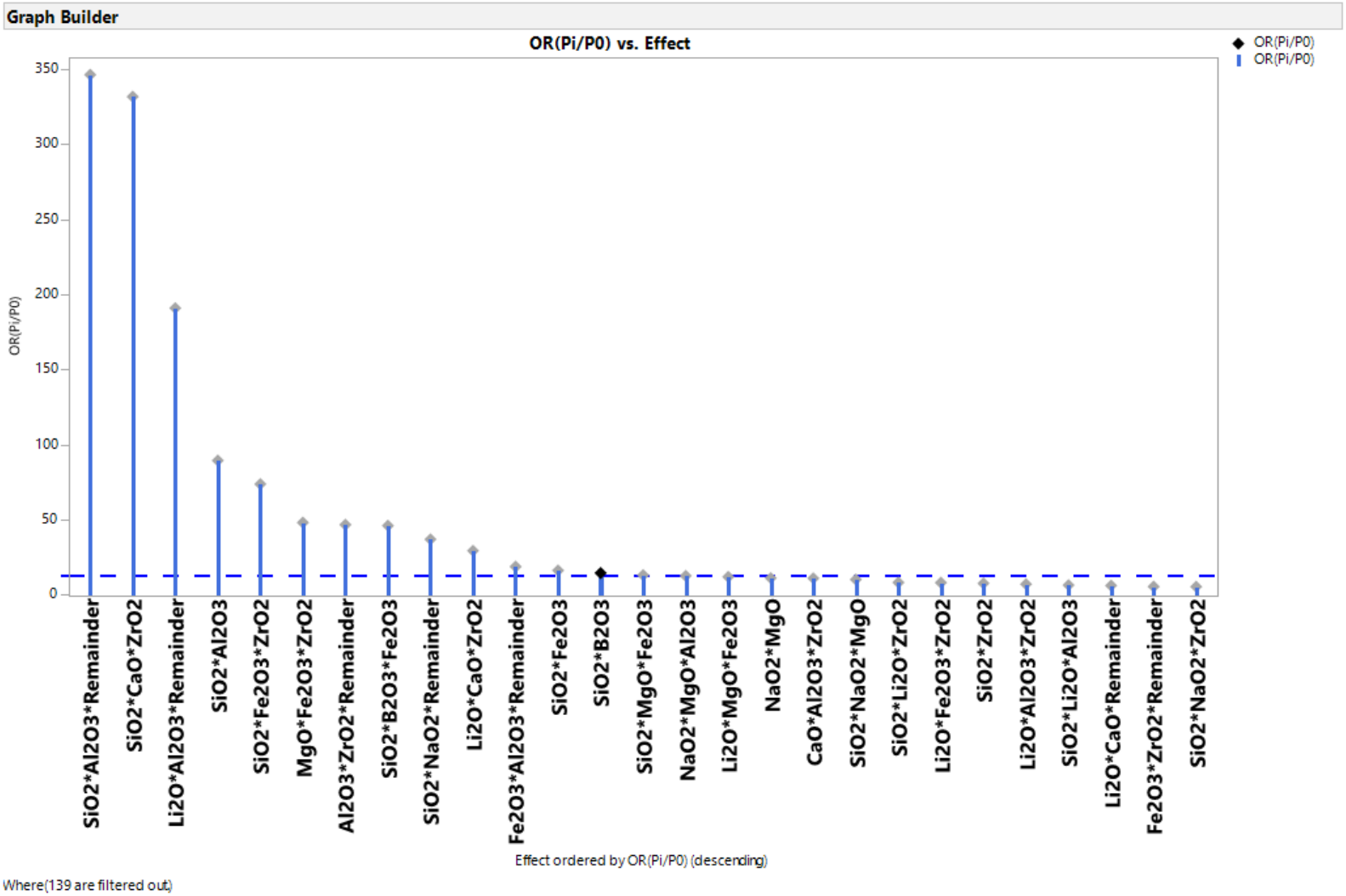
$$OR_i = \frac{\frac{p_i}{(1-p_i)}}{\frac{p_0}{(1-p_0)}}$$

As an alternative the proportion for the null factor could be used in the denominator.

ORs have nice statistical properties that could be exploited for further inference.

Case Study: The Waste Glass Experiment

Below is a screenshot of the simulation results ordered descending by odds ratio of entering the model vs the null hypothesis.



Case Study: The Waste Glass Experiment

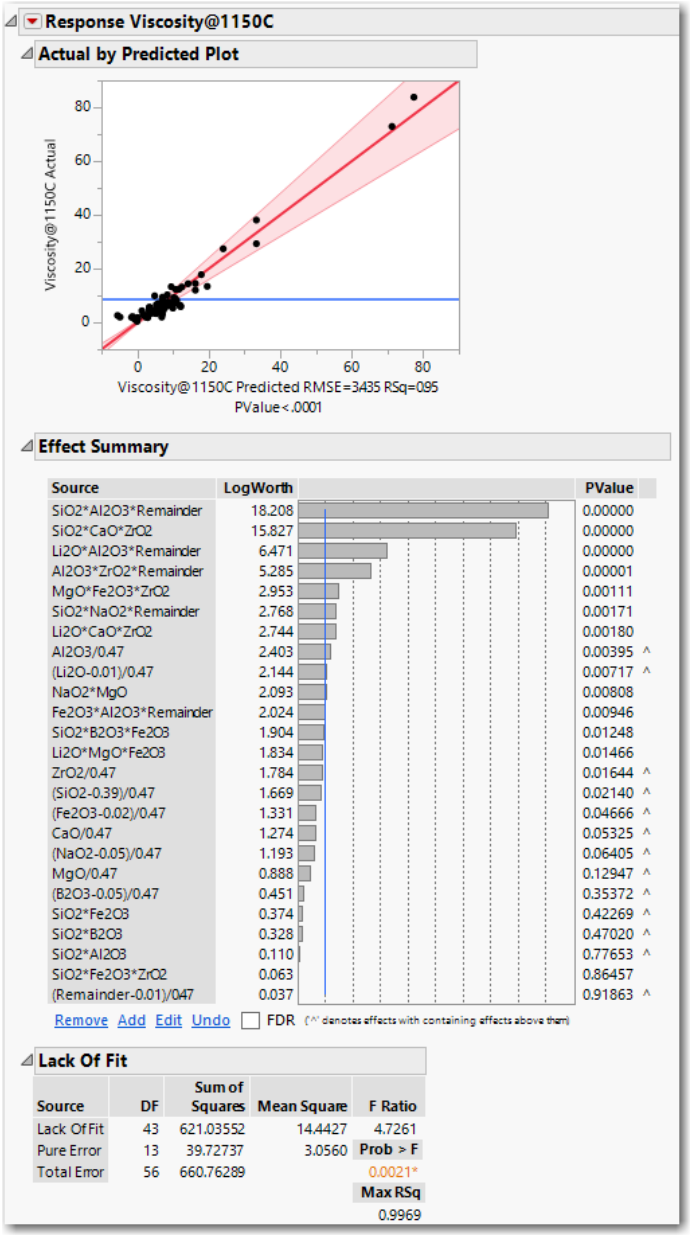
To the right is a screenshot of the final selected special cubic model.

Although some lack of fit still exists, overall the model fits much better than the quadratic model.

It is common to have some level of lack of fit in mixture models that cannot be completely accommodated.

At this point it is wise to consult with subject matter experts as to whether the amount of LOF is of concern.

Overall, the LOF does not appear large and may not be of concern.



Case Study: The Waste Glass Experiment

Given there were 13 replicate runs, it is possible that some of them could be used as a validation set for model selection and the remainder as the training set.

There were also some extra degrees of freedom in the experiment that could also be used for validation – this is not the usual case.

To investigate the approach, a randomly selected subset of 10 runs were used as a validation set (mostly from the replicate runs) and the remainder for training.

The special cubic model was defined and Generalized Regression platform used for model selection.

As before pure component terms are forced into the model and Forward selection was used as the model selection algorithm.

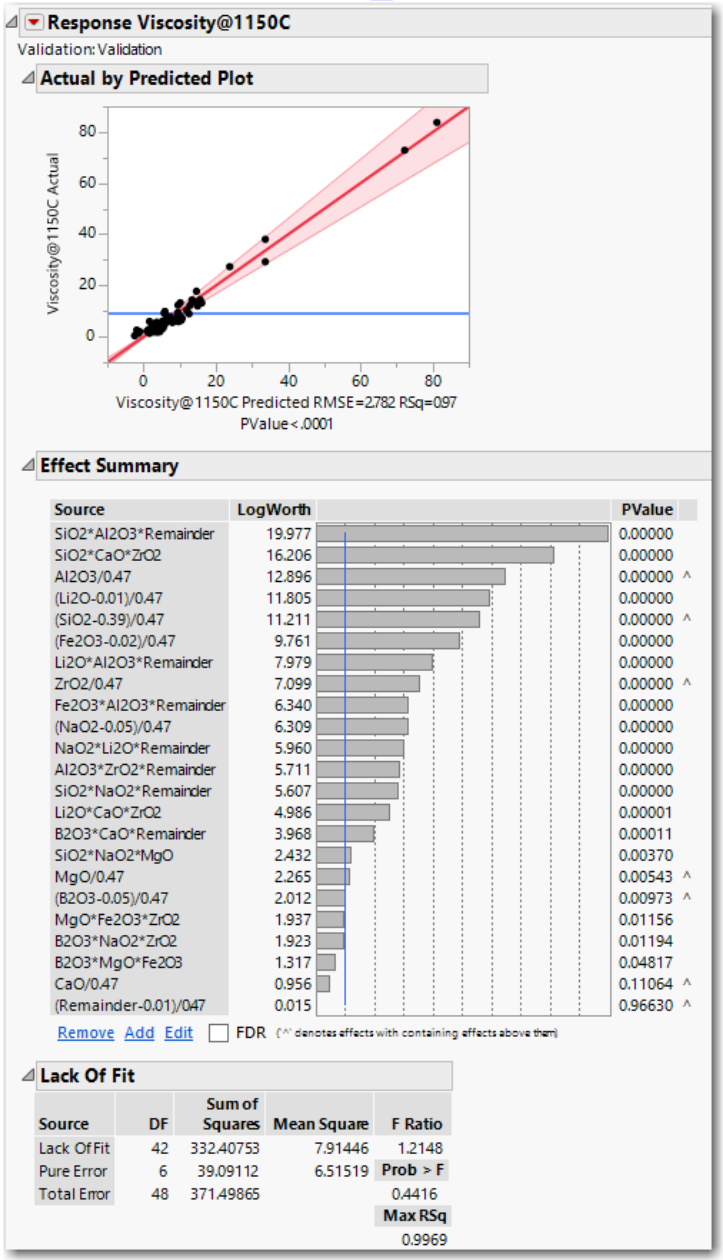
Case Study: The Waste Glass Experiment

To the right is a screenshot of the final fitted model.

Overall this model has the best fit with no significant LOF.

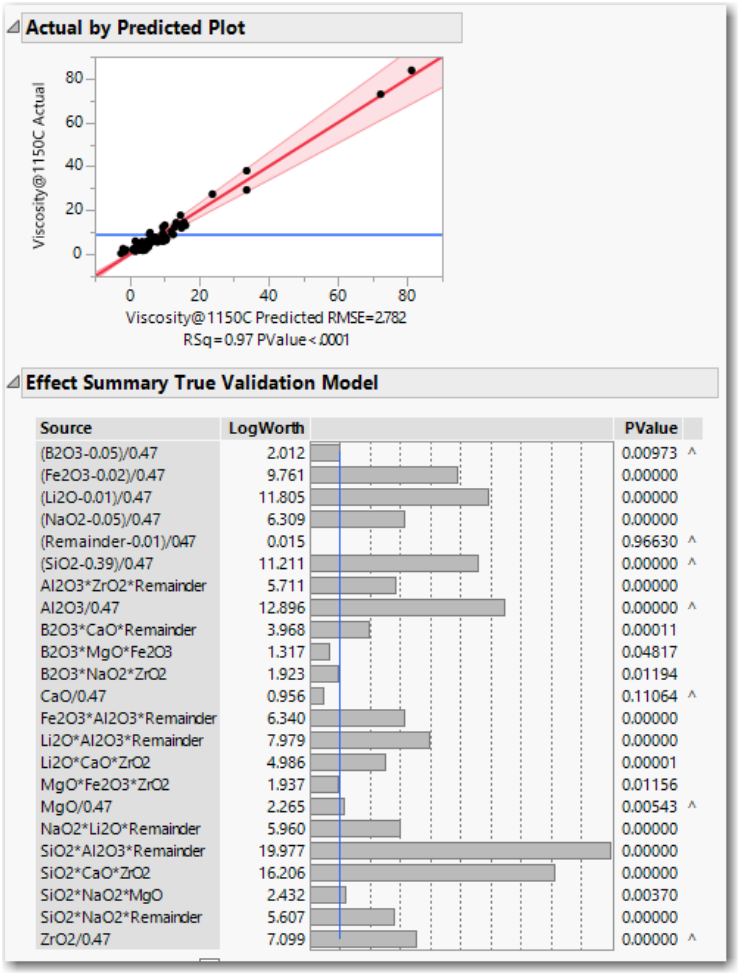
However, the autovalidation models had 13 df for the LOF test while this model has 6 due to the use of some trials for the validation set.

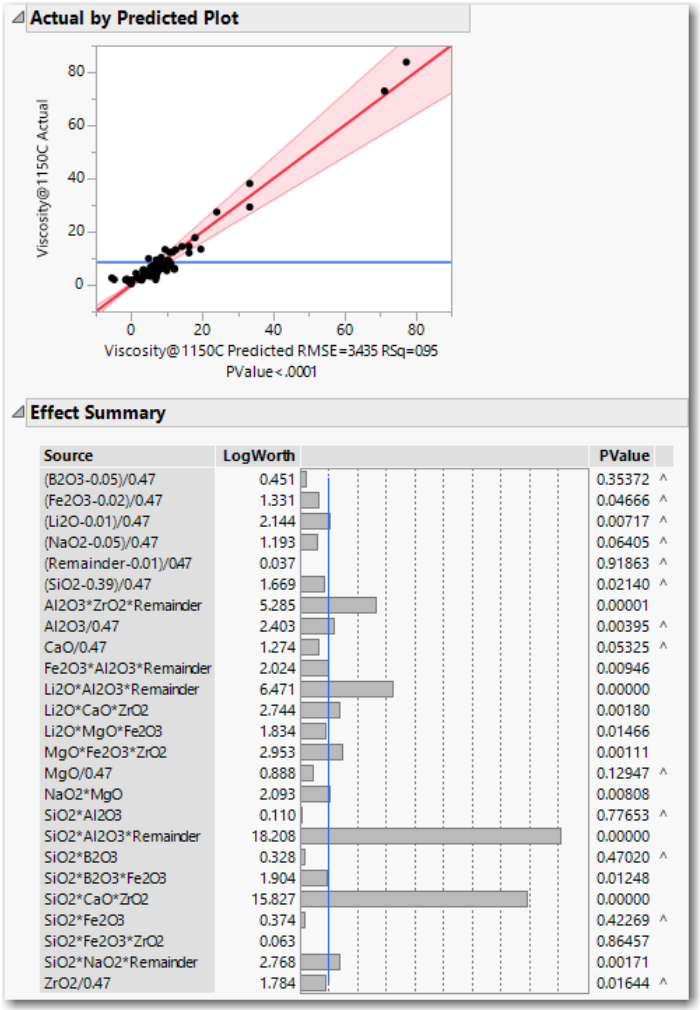
The autovalidation model and true validation model are actually similar.



Case Study: The Waste Glass Experiment

Below is a comparison of the two models, they are not identical but similar and both fit the data quite well.





Mixture Process Factor Experiments

A much underutilized class of experimental designs for process development and improvement are designs with combinations of mixture and process (factorial) factors.

In general, formulations developed under very controlled process conditions (often only one setting) do not work well in practice.

The reason is that the blending behavior of the mixture factors changes as a function of the settings of the process factors.

In order to develop robust and optimal processes it is important to simultaneously experiment with both the mixture and process factors.

Such experimental designs are commonly called **mixture process factor designs**.

Mixture Process Factor Experiments

Mixture process factor designs have a long history, however the classical designs were often infeasibly large.

With modern optimal designs (Custom Design) mixture process factor experiments can be generated that are reasonable in size.

The most common type of mixture process factor experiment is the mixture amount experiment, where the total amount of the mixture used is a process factor.

We will illustrate with a case study.

Mixture process factors still cannot have an intercept and are a bit tricky to create in JMP, but not so hard with a little practice.

Historically, mixture process factor experiments have been very difficult to analyze for reasons given earlier and often badly over fit models were selected.

Mixture-Process Factor Designs

A special notation is often used to write MPV models in a compact form. Each coefficient has a subscript defining the mixture factors and a superscript defining the process factors that are associated with that coefficient.

Here is a full MPV model for three mixture factors X_i and two process factors Z. Note the 2nd order terms in Z_j are often necessary.

$$
\begin{aligned}
Y_i = {} & \beta_1^0 X_1 + \beta_2^0 X_2 + \beta_3^0 X_3 + \beta_{12}^0 X_1 X_2 + \beta_{13}^0 X_1 X_3 + \beta_{23}^0 X_2 X_3 + \beta_{123}^0 X_1 X_2 X_3 + \\
& \left[\beta_1^1 X_1 + \beta_2^1 X_2 + \beta_3^1 X_3 + \beta_{12}^1 X_1 X_2 + \beta_{13}^1 X_1 X_3 + \beta_{23}^1 X_2 X_3 + \beta_{123}^1 X_1 X_2 X_3 \right] Z_1 + \\
& \left[\beta_1^2 X_1 + \beta_2^2 X_2 + \beta_3^2 X_3 + \beta_{12}^2 X_1 X_2 + \beta_{13}^2 X_1 X_3 + \beta_{23}^2 X_2 X_3 + \beta_{123}^2 X_1 X_2 X_3 \right] Z_2 + \\
& \left[\beta_1^{12} X_1 + \beta_2^{12} X_2 + \beta_3^{12} X_3 + \beta_{12}^{12} X_1 X_2 + \beta_{13}^{12} X_1 X_3 + \beta_{23}^{12} X_2 X_3 + \beta_{123}^{12} X_1 X_2 X_3 \right] Z_{12} + \\
& \left[\beta_1^{11} X_1 + \beta_2^{11} X_2 + \beta_3^{11} X_3 + \beta_{12}^{11} X_1 X_2 + \beta_{13}^{11} X_1 X_3 + \beta_{23}^{11} X_2 X_3 + \beta_{123}^{11} X_1 X_2 X_3 \right] Z_1^2 + \\
& \left[\beta_1^{22} X_1 + \beta_2^{22} X_2 + \beta_3^{22} X_3 + \beta_{12}^{22} X_1 X_2 + \beta_{13}^{22} X_1 X_3 + \beta_{23}^{22} X_2 X_3 + \beta_{123}^{22} X_1 X_2 X_3 \right] Z_2^2
\end{aligned}
$$

Mixture-Process Factor Case Study

Chardon (1989) discusses the development of a finishing material for cotton-polyester fabrics. We will use this example (file **MixtureProcessFabric.JMP**) to demonstrate the analysis of a MPV experiment.

The finisher is comprised of three fabric softeners and a catalyzed resin.

We will designate the three mixture components as **A**, **B**, and **C** (the softeners).

There are two process factors. **D** is the level of the catalyzed resin used in the finisher, and **E** is the total amount of softener in the finisher.

Note factor **E** is a **total amount** factor for the three mixture factors.

Mixture-Process Factor Case Study

The response is **hydrophobicity** of the fabric (water repellent measure), which the researchers would like to maximize.

Due to heteroscedasticity the authors recommended analyzing the log of hydrophobicity.

The researchers ran a simplex centroid mixture design at each of 7 settings of the process factors.

The process factor design is known as an equiradial (Doelhert) response surface design, which is hexagonal in shape.

There are a total of 49 runs in the MPV and the full MPV model has 42 potential terms to estimate – similar to our example given earlier.

Mixture-Process Factor Case Study

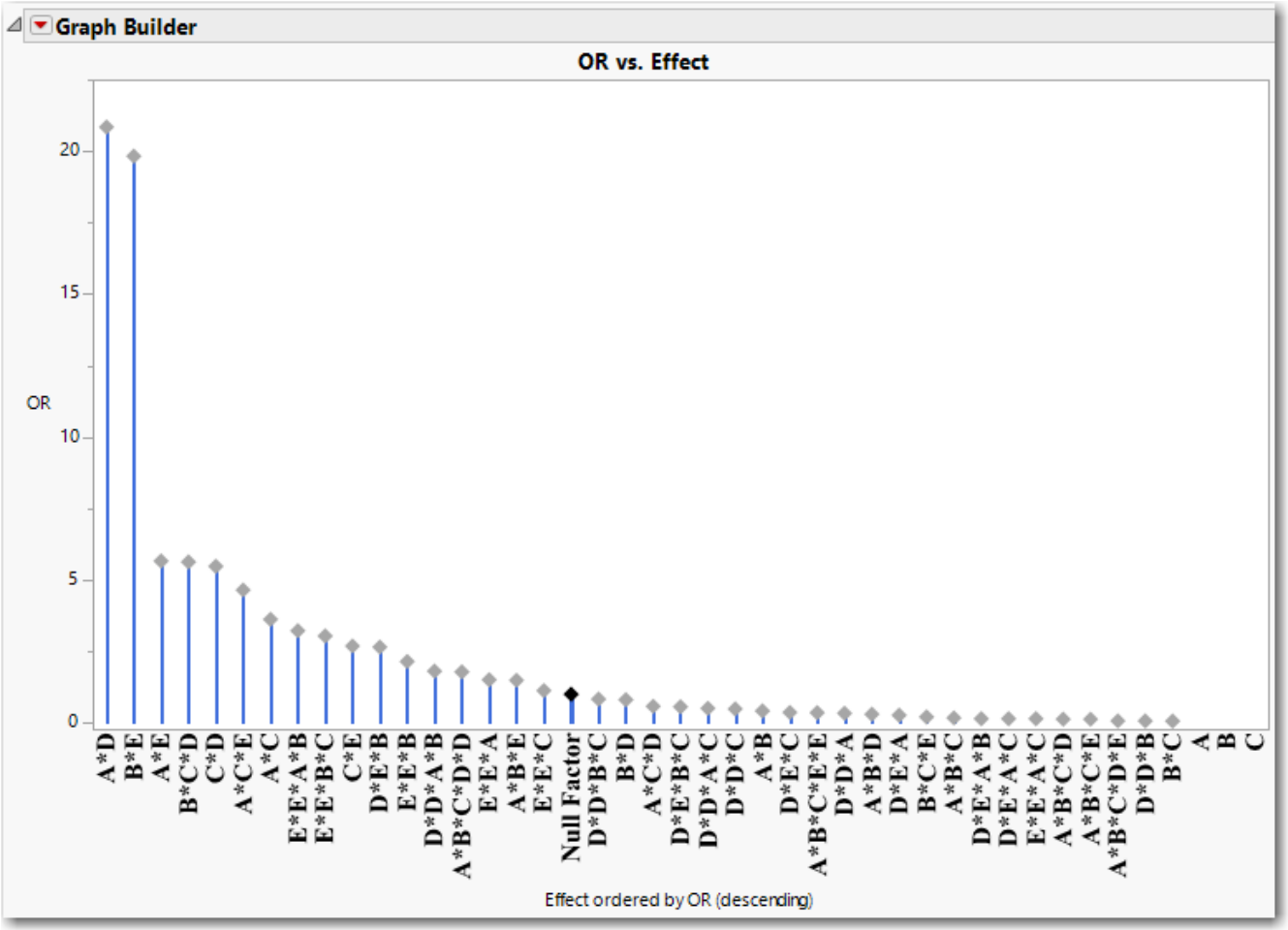
We will use autovalidation and Gen Reg to analyze the experiment.

Below is the full model, with 42 terms:

$$
\begin{aligned}
Y_i = {} & \beta_1^0 A + \beta_2^0 B + \beta_3^0 C + \beta_{12}^0 AB + \beta_{13}^0 AC + \beta_{23}^0 BC + \beta_{123}^0 ABC + \\
& \left[\beta_1^D A + \beta_2^D B + \beta_3^D C + \beta_{12}^D AB + \beta_{13}^D AC + \beta_{23}^D BC + \beta_{123}^D ABC \right] D + \\
& \left[\beta_1^E A + \beta_2^E B + \beta_3^E C + \beta_{12}^E AB + \beta_{13}^E AC + \beta_{23}^E BC + \beta_{123}^E ABC \right] E + \\
& \left[\beta_1^{DE} A + \beta_2^{DE} B + \beta_3^{DE} C + \beta_{12}^{DE} AB + \beta_{13}^{DE} AC + \beta_{23}^{DE} BC + \beta_{123}^{DE} ABC \right] DE + \\
& \left[\beta_1^{D^2} A + \beta_2^{D^2} B + \beta_3^{D^2} C + \beta_{12}^{D^2} AB + \beta_{13}^{D^2} AC + \beta_{23}^{D^2} BC + \beta_{123}^{D^2} ABC \right] D^2 + \\
& \left[\beta_1^{E^2} A + \beta_2^{E^2} B + \beta_3^{E^2} C + \beta_{12}^{E^2} AB + \beta_{13}^{E^2} AC + \beta_{23}^{E^2} BC + \beta_{123}^{E^2} ABC \right] E^2
\end{aligned}
$$

Mixture-Process Factor Case Study

Below is a plot of the effects by Odds Ratio for the autovalidation results. We will fit a model containing the pure components and the effects > null factor.

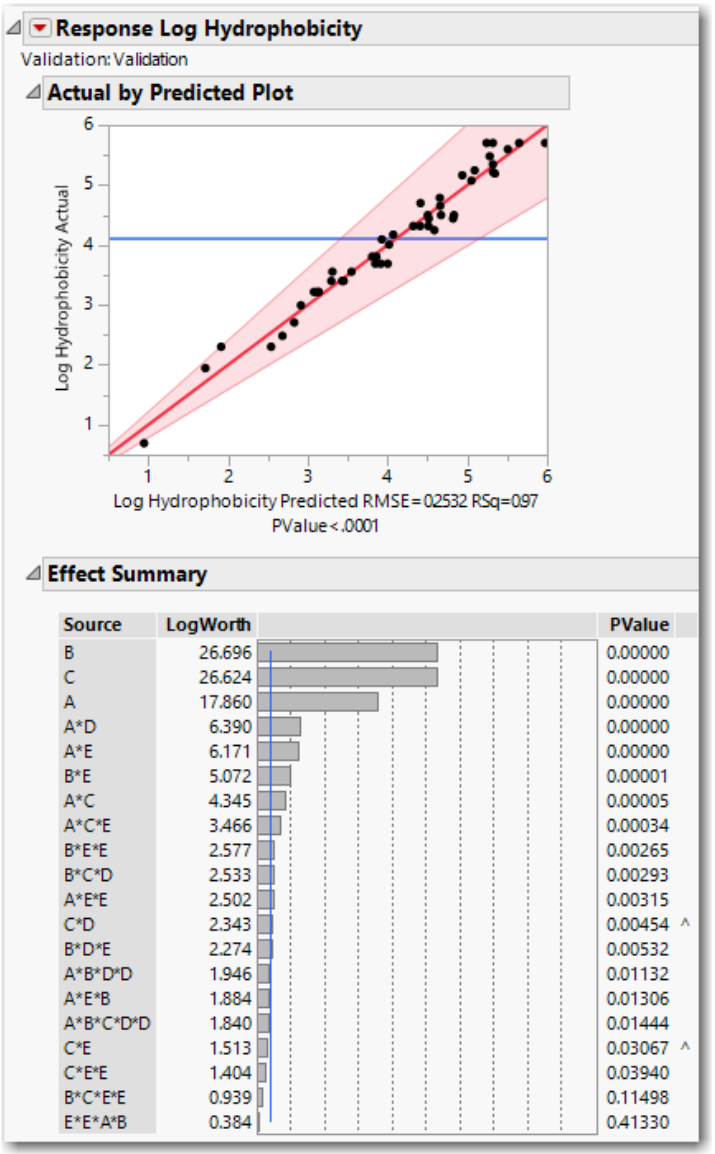


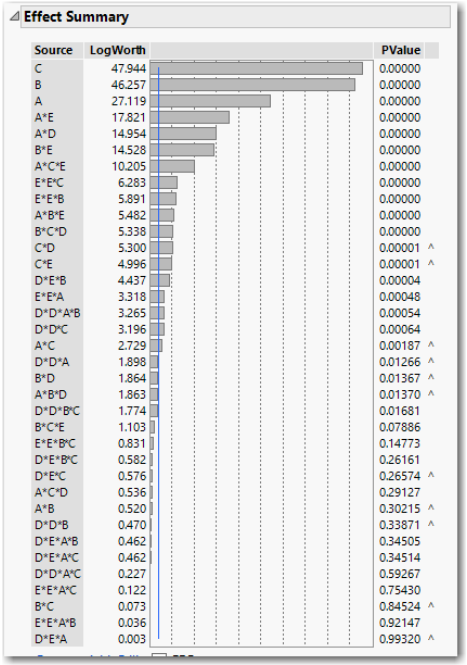
Mixture-Process Factor Case Study

To the right is a screenshot of fitted model based on the autovalidation results.

The fit appears quite good.

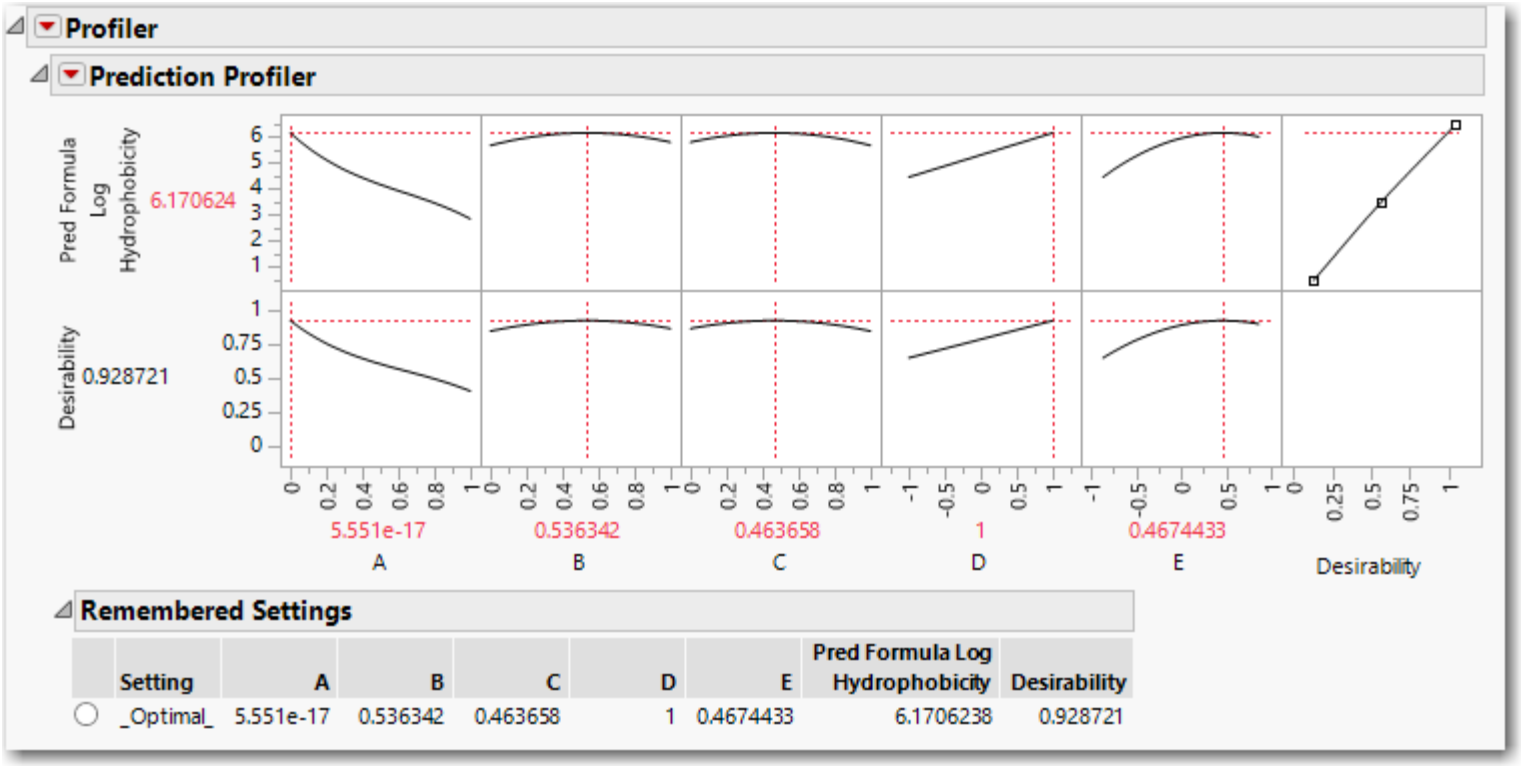
The published model is much larger and does not fit as well; see below.





Mixture-Process Factor Case Study

The goal of the experiment was to find settings of the factors that maximizes hydrophobicity. Using the autovalidation model and the Prediction Profiler we find the optimized settings – quite different from the original published recommendation.



Summary and Conclusions

Traditionally building predictive models from mixture and mixture process DOE data has been limited by the lack of validation trials to control over fitting.

Typically a DOE budget and time does not make it feasible to perform a separate set of validation trials.

Autovalidation and Fractionally Weighted Bootstrapping are two new viable techniques that enable predictive modeling from DOE data without running a set of validation trials.

The techniques use the original training data and bootstrapping concepts to form validation sets.

Autovalidation and FWB should be considered a part of mixture DOE analysis where the goal is to build predictive models.

Bibliography

Cornell, J.A. (2002). Experiments with Mixtures 3rd Ed. John Wiley & Sons.

Cornell, J.A., and Gorman, J.W. (2002). "Two New Mixture Models: Living With Collinearity but Removing Its Influence." *Journal of Quality Technology, Vol 35, No 1.*

Miller, A.J. (2002). *Subset Selection in Regression, 2nd Edition*. Chapman Hall/CRC, 235 pp.

Redgate, P.E., Piepel, G.F., and Hrma, P. (1992). "Second-Order Model Selection in Mixture Experiments." ASA, *Proceedings of the Section on Physical and Engineering Sciences.*

Scheffe, Henri (1958). "Experiments with Mixtures." *Journal of the Royal Statistical Society, Ser. B, 20, 344-360.*

Bibliography

Smith, W.F. (2005). “Experimental Design for Formulation.” *ASA-SIAM Series on Statistics and Applied Probability*, ASA, Alexandria, VA, USA.

Wu, Y., Boos, D., and Stefanski, L. (2007), “Controlling Variable Selection by the Addition of Pseudovariables”, *Journal of the American Statistical Association.*