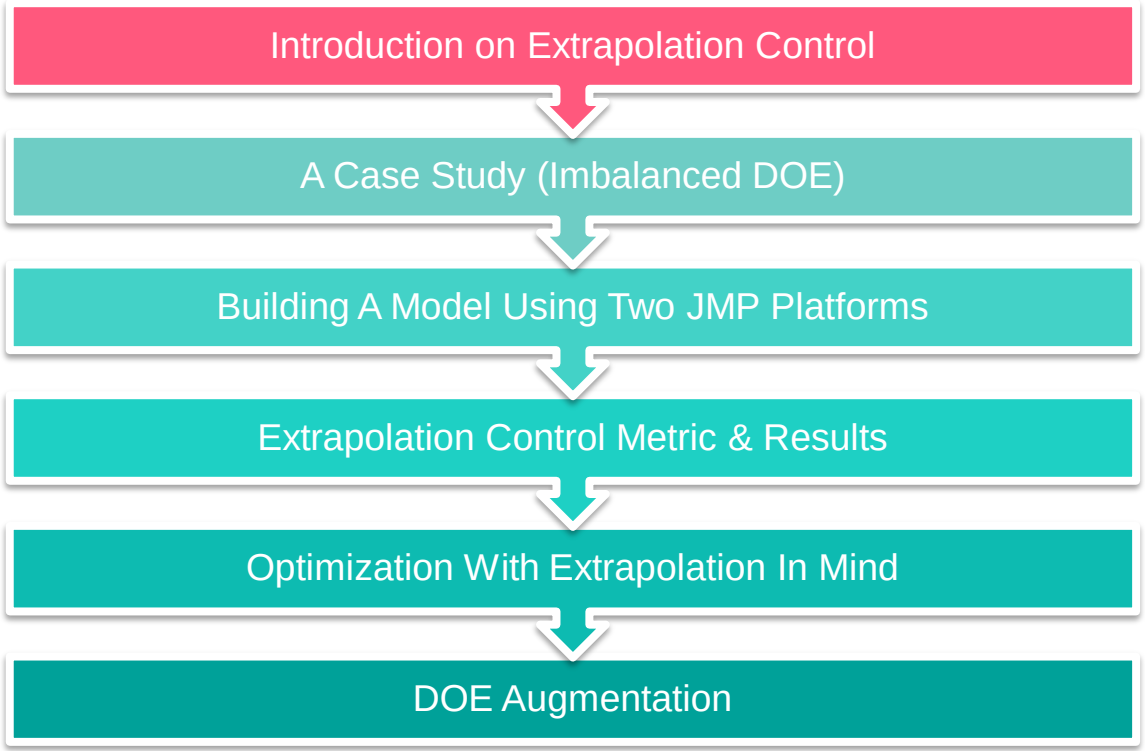
Utilizing Extrapolation Control Feature in JMP® Pro 16 to Optimize Real-World Mining Process

Xinjie Tong

08/31/2021

Present at Discovery Summit Americas 2021

OUTLINE


Introduction on Extrapolation Control
A Case Study (Imbalanced DOE)
Building A Model Using Two JMP Platforms
Extrapolation Control Metric & Results
Optimization With Extrapolation In Mind
DOE Augmentation

JMP 16 Pro - Extrapolation Control Criteria*

- Linear Model
  - Use leverage (hat) to detect extrapolation.
  - Available for Fit Least Squares.
  - $Leverage(x) = x^T(X^TX)^{-1}x$ where x is the prediction point and X is the design matrix of training data
  - For a linear model, leverage/hat is equivalent to a scaled prediction variance (for mean).

    $$standard\ error = RMSE * \sqrt{h_x}$$

  - Extrapolation happens when
    - ✓ $h > h_{max}$
    - ✓ $h > 3h_{avg} = 3p/n$
  - Hat is a function of model terms and design matrix.

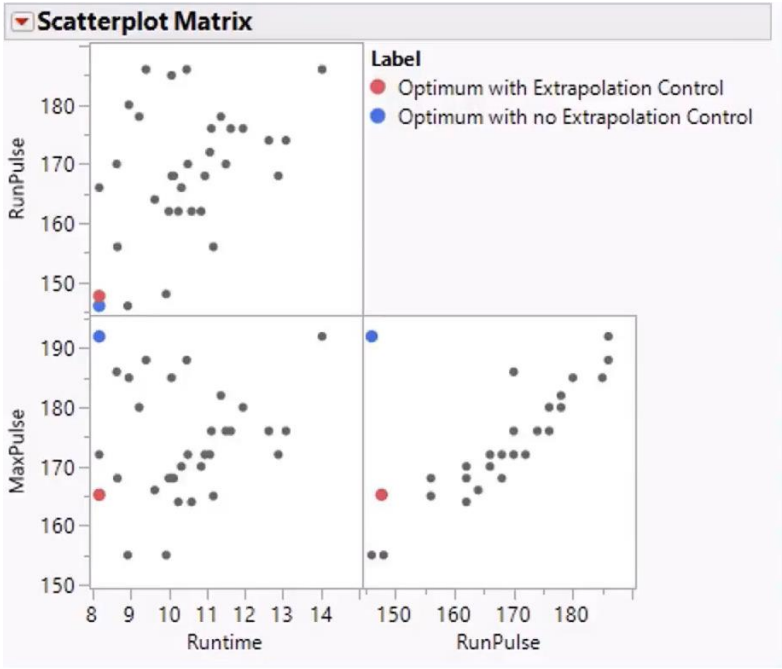
- Nonlinear Model
  - Use regularized Hotelling's T square to detect extrapolation (unsupervised). Not a function of model terms.
  - Extrapolation happens when > mean + 3 * standard deviation of the sample regularized T Squares
  - Available for Generalized Regression, Neural, Support Vector Machines, Partial Least Squares, Naïve Bayes

Hotelling's T^2 defined as:

$$T^2 = (x - \bar{x})^T \hat{\boldsymbol{\Sigma}}^{-1} (x - \bar{x})$$

*Lancaster, L.; Ash, J.; Gotwalt, C. Controlling Extrapolation in the Prediction Profiler in JMP Pro 16. SAS.

Example of extrapolation control for linear model



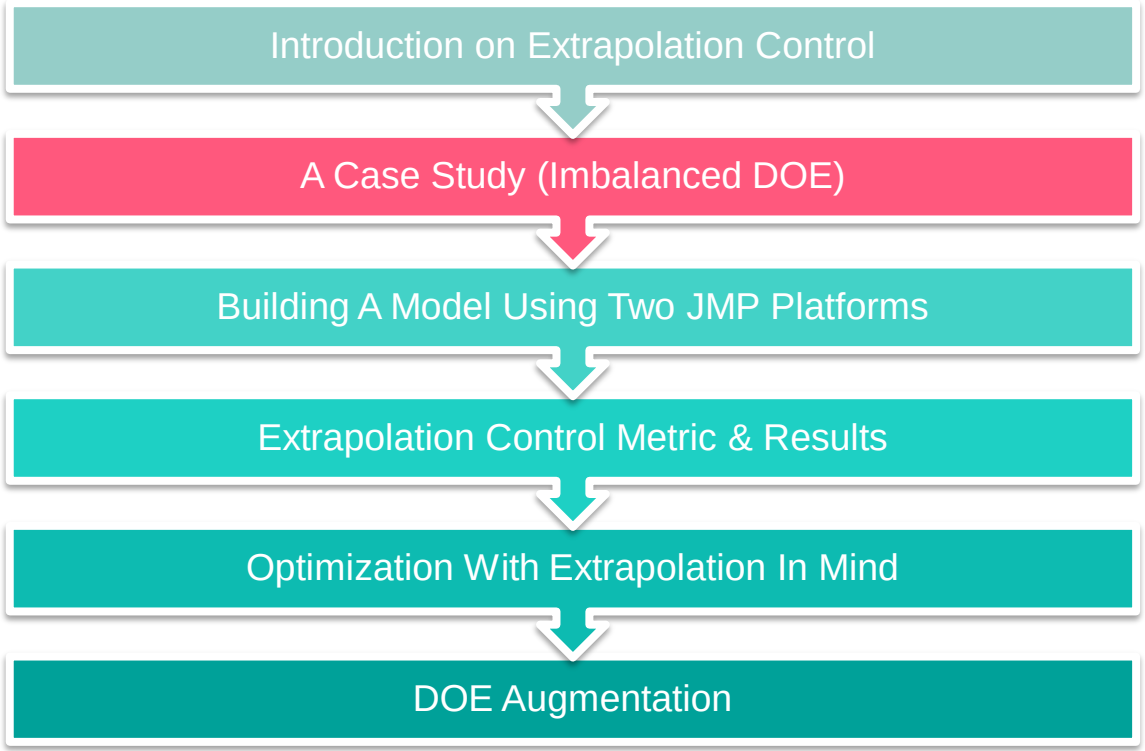
An example from SAS video*.

- Model = least squares
- Extrapolation metric = leverage
- Threshold = 3*average leverage

Extrapolation control detects the conditions that do not follow the relationship among input variables.

*Lancaster, L.; Ash, J.; Gotwalt, C. Controlling Extrapolation in the Prediction Profiler in JMP Pro 16. SAS.

OUTLINE


Introduction on Extrapolation Control
A Case Study (Imbalanced DOE)
Building A Model Using Two JMP Platforms
Extrapolation Control Metric & Results
Optimization With Extrapolation In Mind
DOE Augmentation

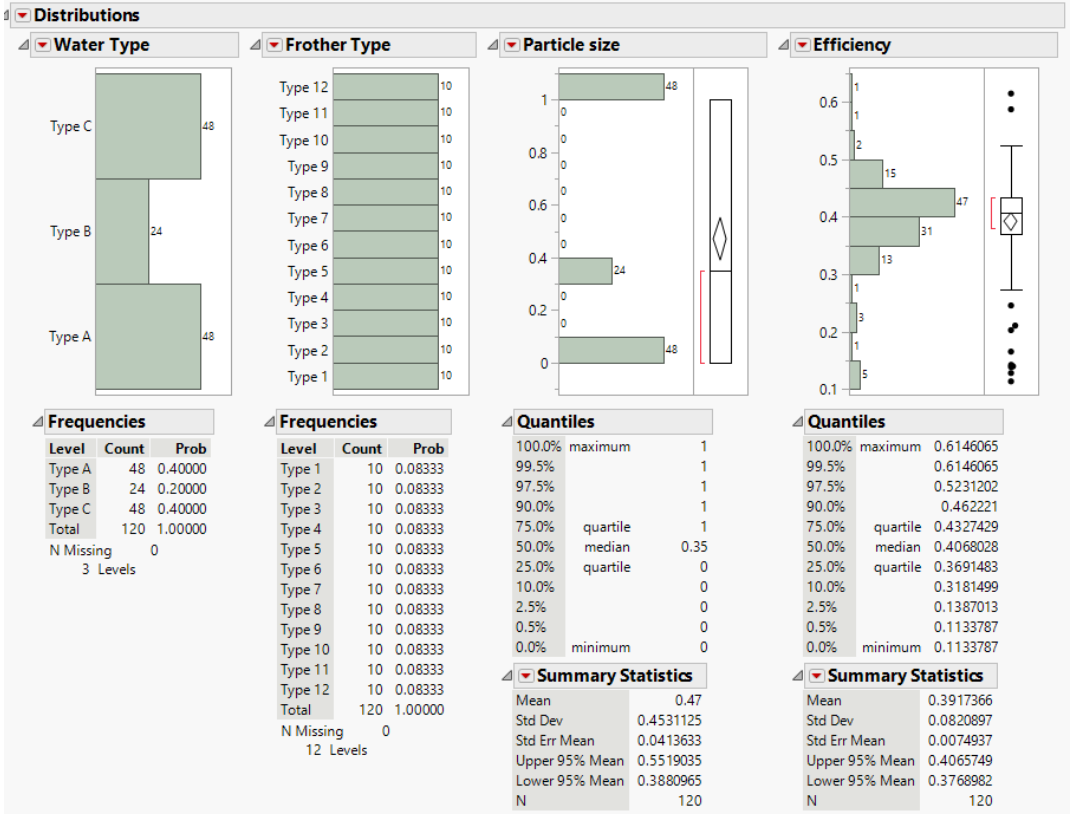
THE CASE STUDY

- Objective: optimize mining conditions for enhanced efficiency

- Input factors:
  1. Water type (categorical) – 3 levels
  2. Frother type (categorical) – 12 levels
  3. Particle size (continuous) – 3 levels

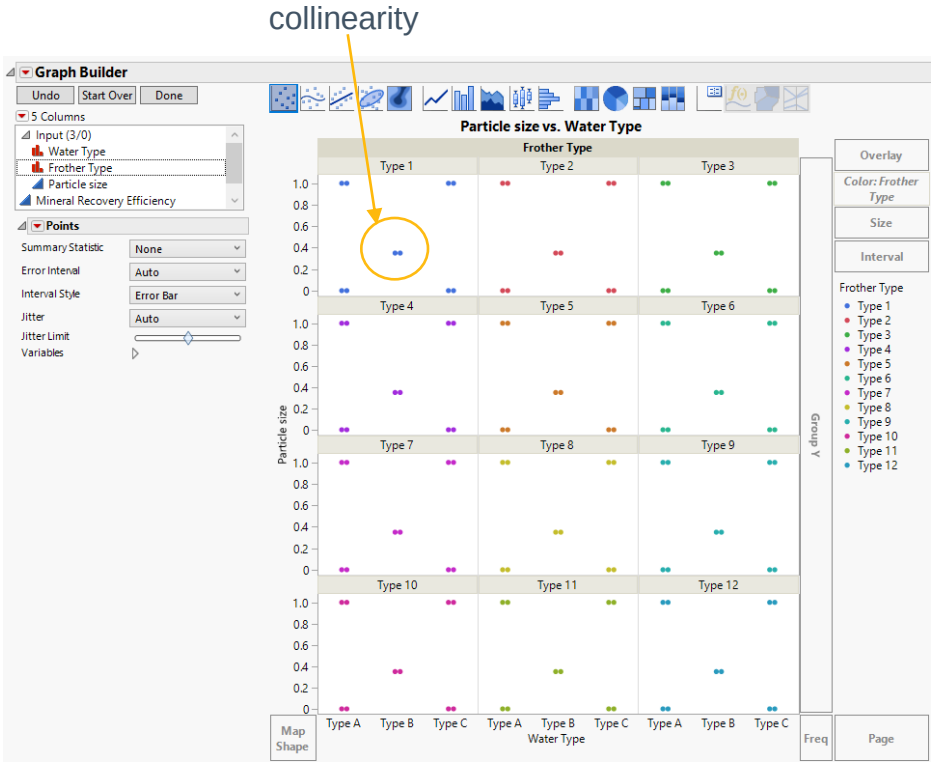
- Output factor:
  1. Efficiency (a value between 0 and 1)

All values of the input variables and the response variable were coded.

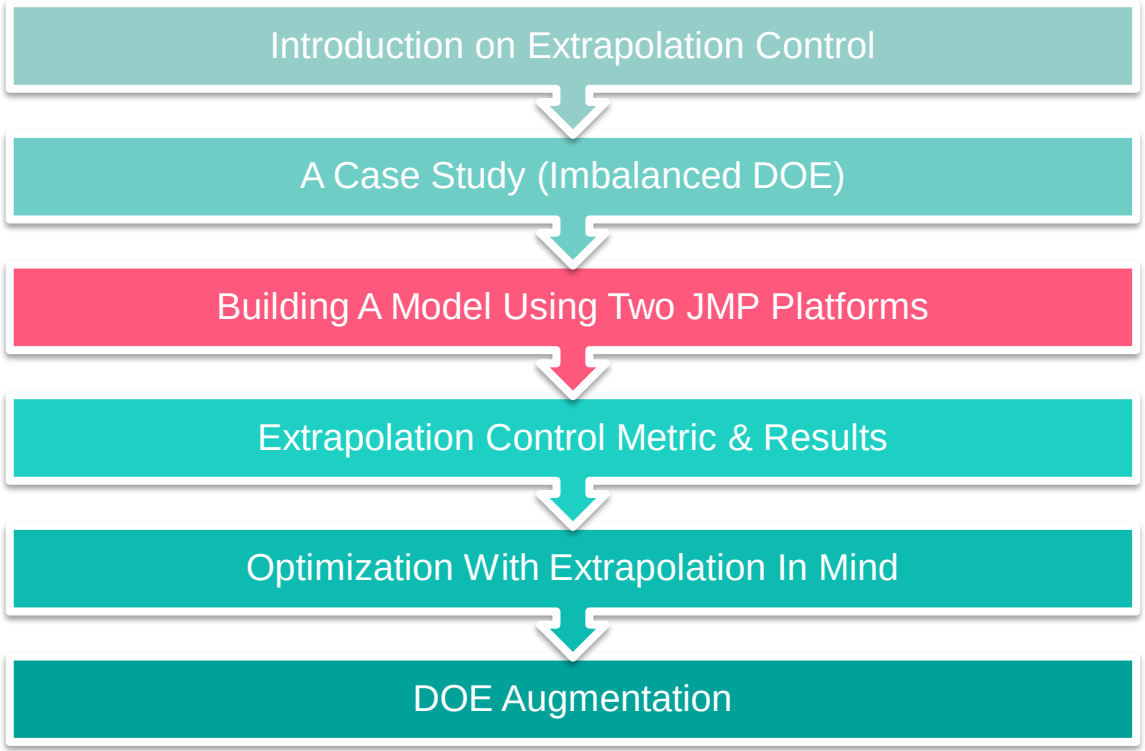


THE DATA IMBALANCE

- Collinearity exists for Water Type B.
- Extrapolation control can help prevent the risks associated with imbalanced data.

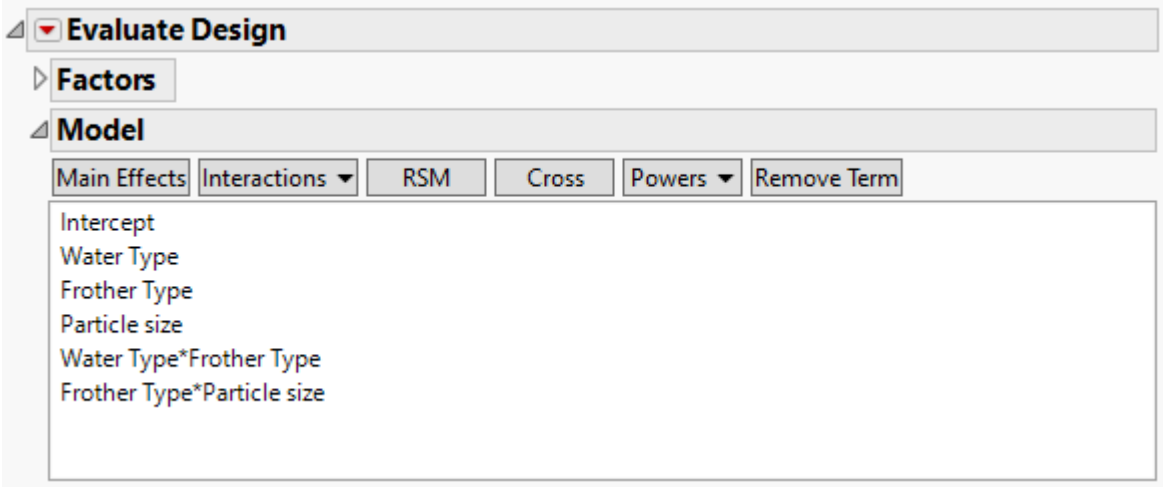


OUTLINE


Introduction on Extrapolation Control
A Case Study (Imbalanced DOE)
Building A Model Using Two JMP Platforms
Extrapolation Control Metric & Results
Optimization With Extrapolation In Mind
DOE Augmentation

Evaluate Design

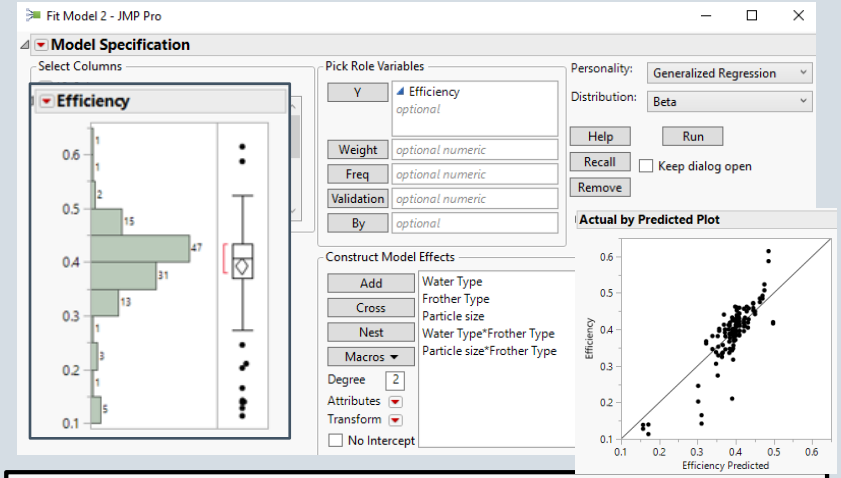
Demo

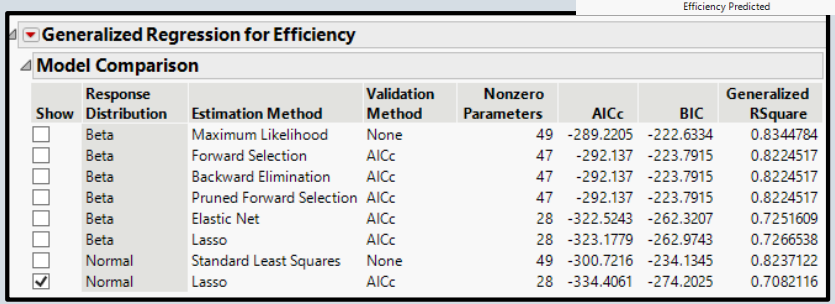


Cannot estimate Particle Size & Water Type interaction

Fit Model Using Two Platforms in JMP

- Generalized Regression (nonlinear)



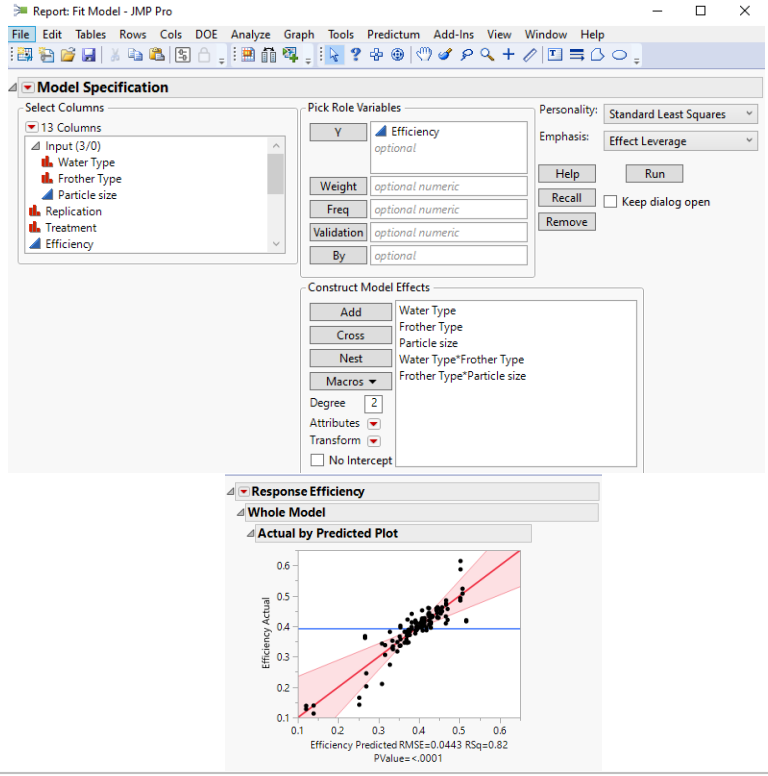


Generalized Regression for Efficiency

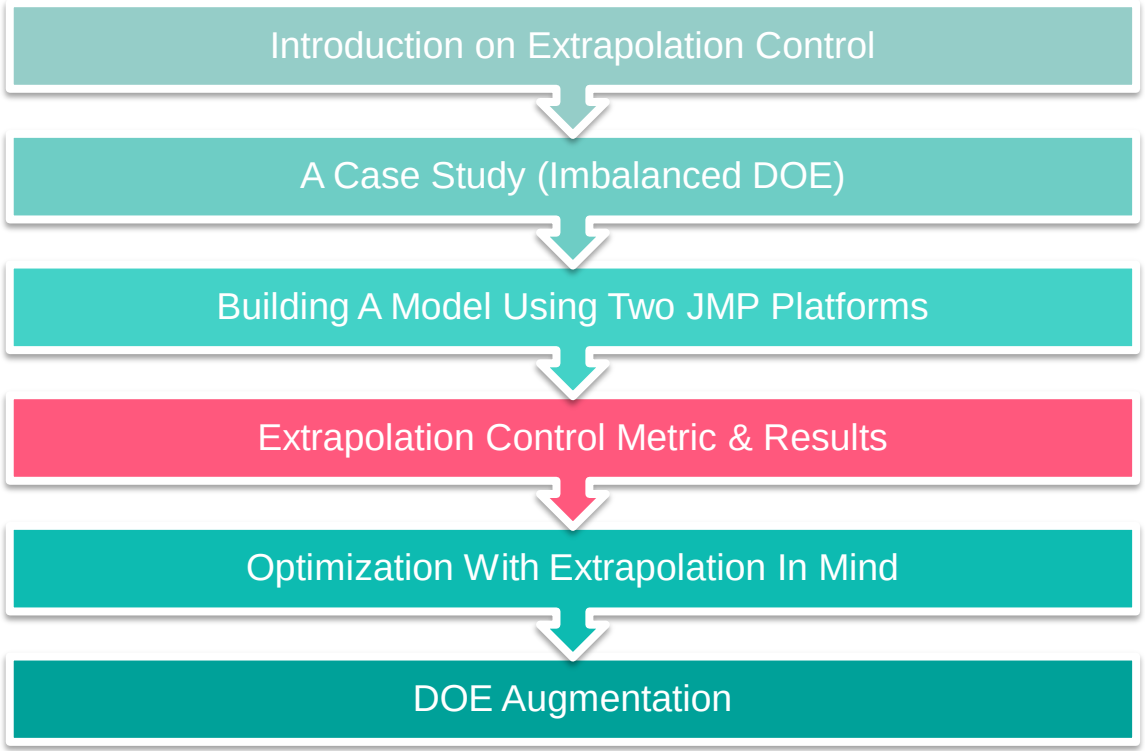
Model Comparison

Show	Response Distribution	Estimation Method	Validation Method	Nonzero Parameters	AICc	BIC	Generalized RSquare
☐	Beta	Maximum Likelihood	None	49	-289.2205	-222.6334	0.8344784
☐	Beta	Forward Selection	AICc	47	-292.137	-223.7915	0.8224517
☐	Beta	Backward Elimination	AICc	47	-292.137	-223.7915	0.8224517
☐	Beta	Pruned Forward Selection	AICc	47	-292.137	-223.7915	0.8224517
☐	Beta	Elastic Net	AICc	28	-322.5243	-262.3207	0.7251609
☐	Beta	Lasso	AICc	28	-323.1779	-262.9743	0.7266538
☐	Normal	Standard Least Squares	None	49	-300.7216	-234.1345	0.8237122
☑	Normal	Lasso	AICc	28	-334.4061	-274.2025	0.7082116

- Fit Least Square (linear)

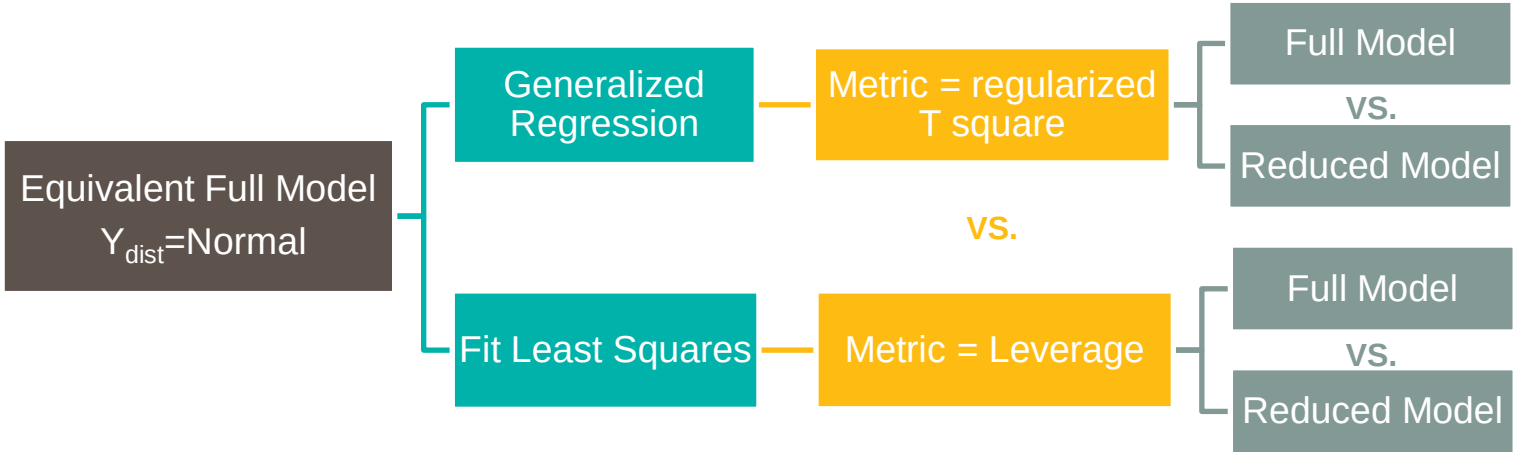


OUTLINE


Introduction on Extrapolation Control
A Case Study (Imbalanced DOE)
Building A Model Using Two JMP Platforms
Extrapolation Control Metric & Results
Optimization With Extrapolation In Mind
DOE Augmentation

Comparison of Extrapolation Matrics

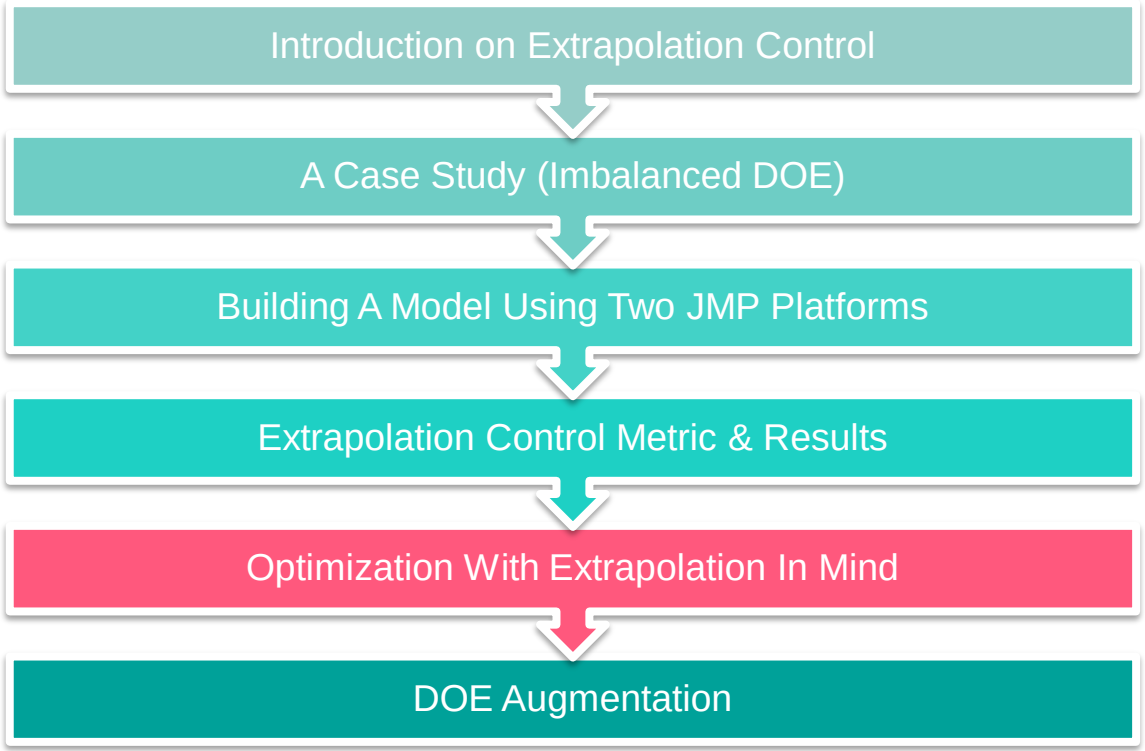
Demo



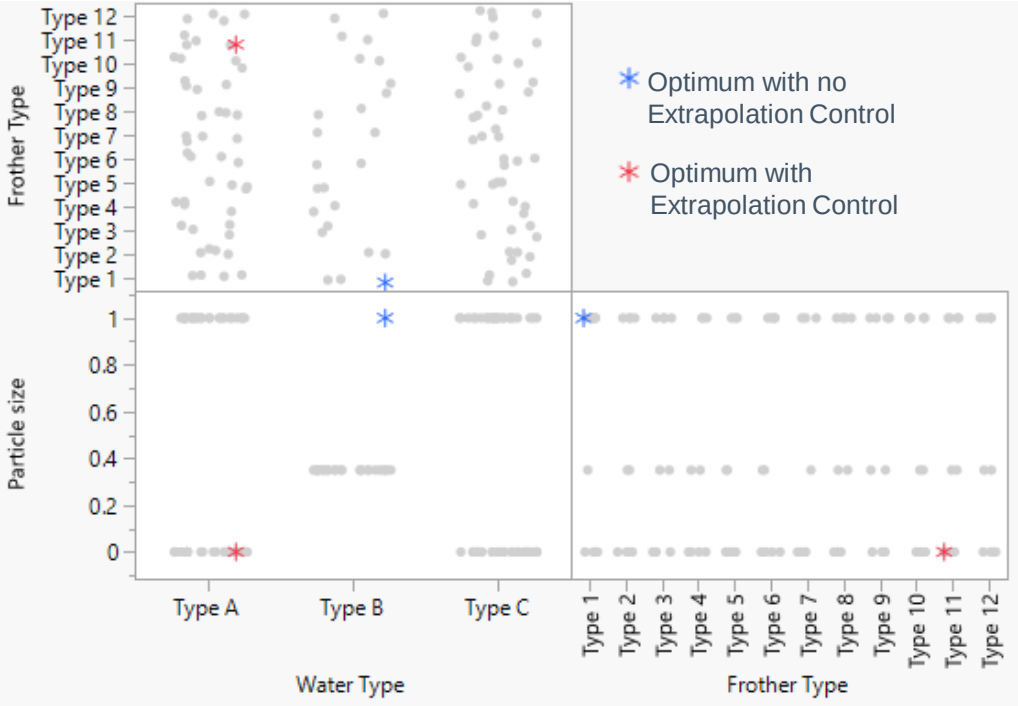
COMPARISON

	Extrapolation Control Metric			
	Leverage		Regularized T Square	
Model	Full	Reduced	Full	Reduced
Threshold	0.50	0.13	16.3	16.3
Dependency of model form & design matrix	Yes		No	
Strictness	Strict		Flexible	

OUTLINE


Introduction on Extrapolation Control
A Case Study (Imbalanced DOE)
Building A Model Using Two JMP Platforms
Extrapolation Control Metric & Results
Optimization With Extrapolation In Mind
DOE Augmentation

OPTIMIZATION



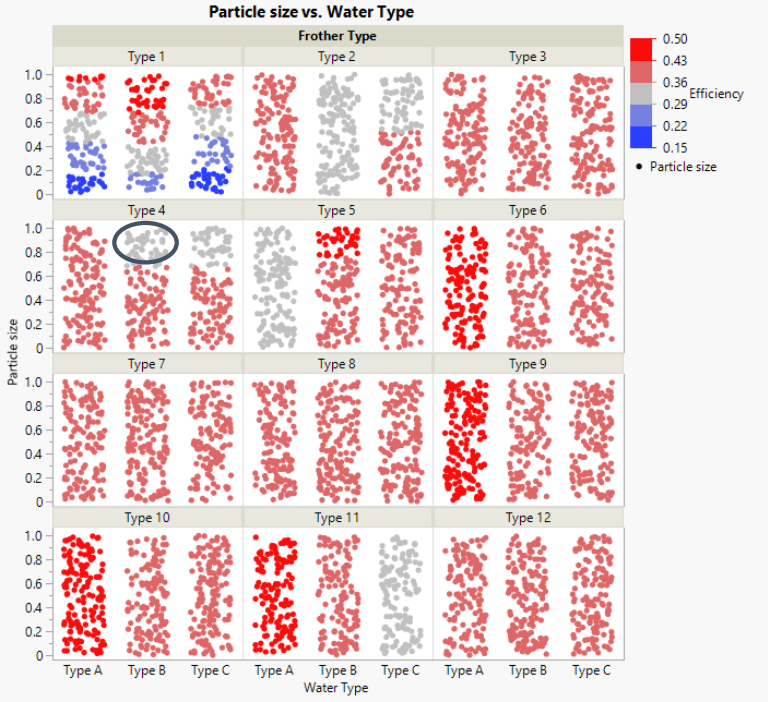
- Model = least squares
- Extrapolation metric = leverage
- Threshold = 3*average leverage

APPLICATION OF THE MODEL

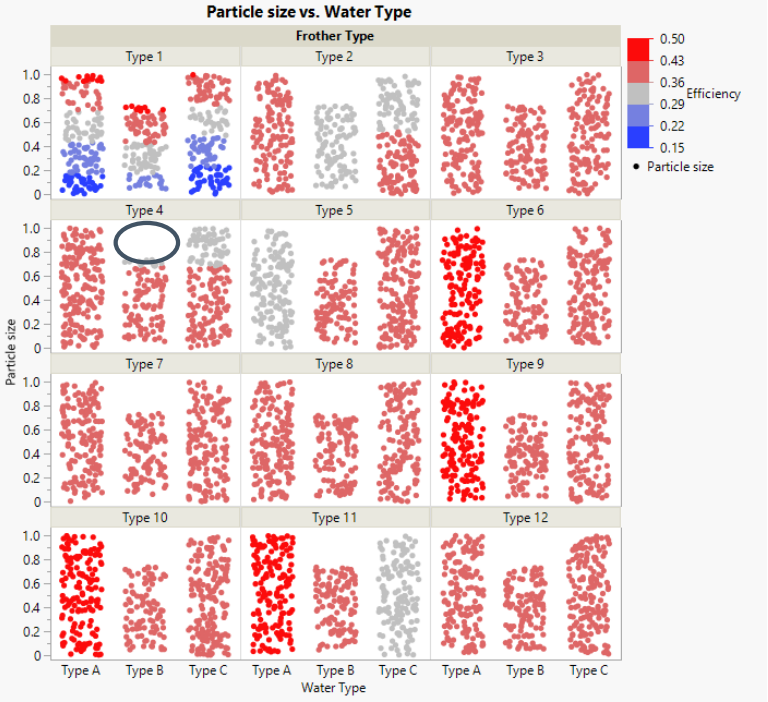
(GenReg; Y_{dist}=Normal, Lasso with AICc)

Demo

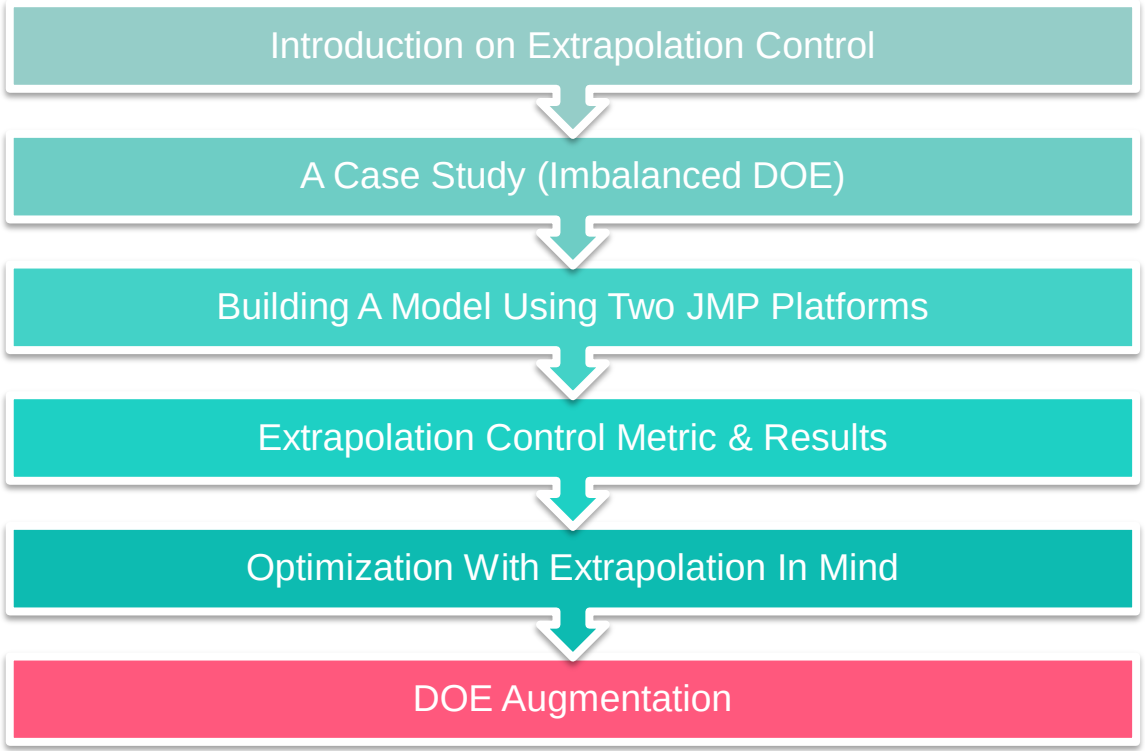
With No Extrapolation Control



With Extrapolation Control



OUTLINE


Introduction on Extrapolation Control
A Case Study (Imbalanced DOE)
Building A Model Using Two JMP Platforms
Extrapolation Control Metric & Results
Optimization With Extrapolation In Mind
DOE Augmentation

DOE AUGMENTATION

Augment Design

Factors

Name	Role	Changes	Values
Water Type	Categorical	Easy	Type A, Type B, Type C
Frother Type	Categorical	Easy	Type, Type, Type, Type, Type, Type, Type, Type, Type, Type, Type, Type
Particle size	Continuous	Easy	0, 1
Block	Blocking	Easy	1, 2

☑ Group new runs into separate block

Define Factor Constraints

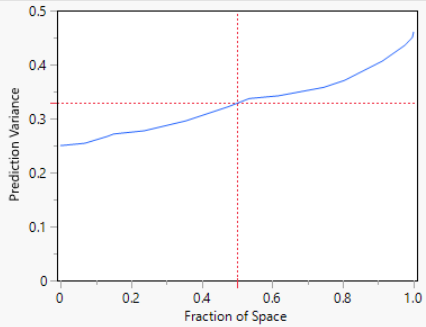
Model

Main Effects | Interactions | RSM | Cross | Powers | Remove Term

Name	Estimability
Intercept	Necessary
Water Type	Necessary
Frother Type	Necessary
Particle size	Necessary
Block	Necessary
Water Type*Frother Type	Necessary
Water Type*Particle size	Necessary
Frother Type*Particle size	Necessary

120 → 144 runs

Fraction of Design Space Plot

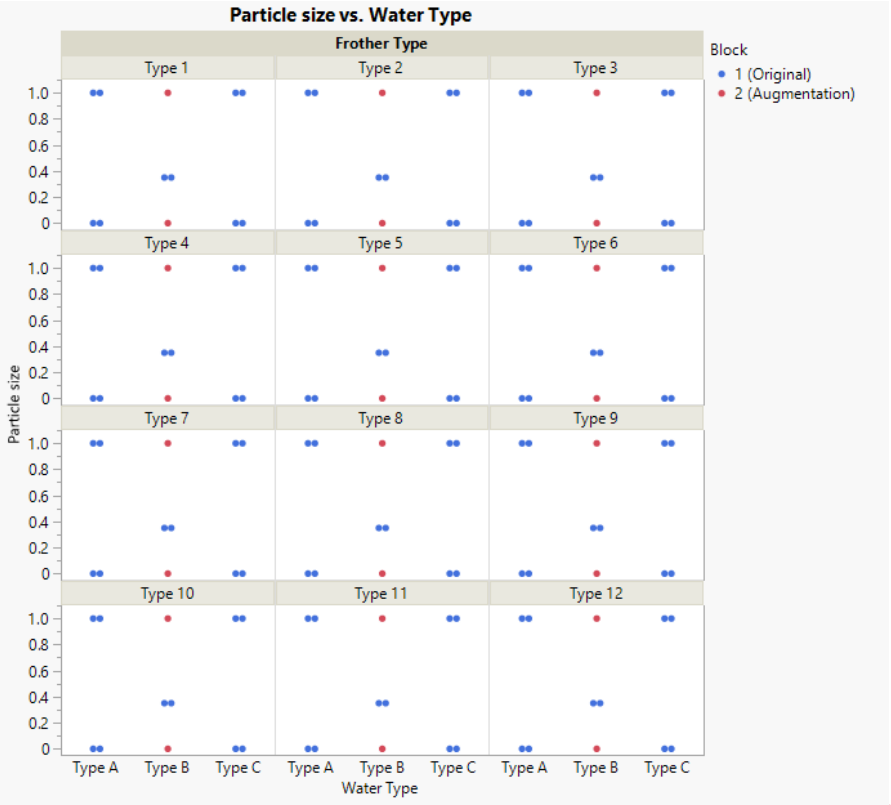


Design Diagnostics

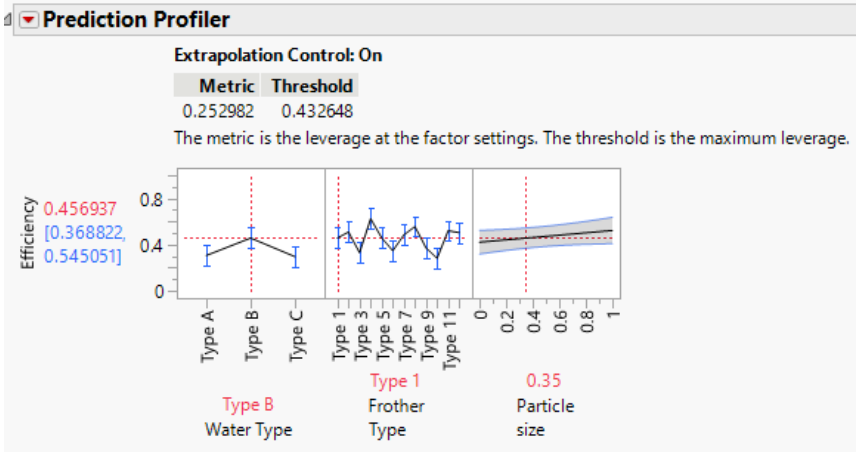
D Optimal Design

D Efficiency	93.00431
G Efficiency	76.82095
A Efficiency	87.2535
Average Variance of Prediction	0.326821
Design Creation Time (seconds)	21

THE AUGMENTED DOE – NO EXTRAPOLATION ISSUE



Leverage
Threshold = 0.43
DOE size = 144



The augmented DOE does not have extrapolation issue.

CONCLUSION

- Sources of extrapolation:
  - Correlation among input variables
  - Inestimable model terms
- Extrapolation control is the “sanity check” for the use of predictive model.
- The extrapolation control feature in JMP 16 Pro made it simple to practice, and provide easy understanding on the effectiveness of augmenting a dataset.

Reference

- Lancaster, L.; Ash, J.; Gotwalt, C. Controlling Extrapolation in the Prediction Profiler in JMP Pro 16. SAS. https://community.jmp.com/t5/Discovery-Summit-Europe-2021/Controlling-Extrapolation-in-the-Prediction-Profiler-in-JMP-Pro/ta-p/349218

- This presentation is provided in good faith for informational purposes only. Dow assumes no obligation or liability.

- Contact: Xinjie Tong, xtong1@dow.com, Statistician at Dow Inc.