predictum®

JMP Discovery Europe 2021

SVEM: A paradigm shift in Design and Analysis of Experiments

1

Agenda

predictum®

- Introductions
- What, Why & How of SVEM
  - Quick overview of DOE and Machine learning
  - Describe blending them into SVEM
  - Analyze real-world SVEM experiments
  - Review Current Research
  - Demonstrate JMP's new Candidate Set Designer
- Next Steps & Q&A

2

2

SVEM is a remarkable new method to extract more insights with fewer experimental cycles and build more accurate predictive models from small sets of data, including DOEs.

Less Cost / Faster to Market / Faster Problem-Solving

3

Speakers

Wayne Levin
President,
Predictum Inc.

levin@predictum.com

Chris Gotwalt, PhD
JMP Chief Data Scientist

JMP Division,
SAS Institute

christopher.gotwalt@jmp.com

Philip Ramsey, PhD
Sr. Data Scientist
Predictum Inc

Professor
University of New Hampshire

philip.ramsey@predictum.com

4

4

JMP's Contributions to Data Science

- DOE
  - Coordinate Exchange Algorithm
  - Definitive Screening Designs
  - SVEM

5

SVEM

- Self-Validated, Ensemble Models
  - Combining two well-practiced methods
    - Machine Learning
    - Design of Experiments
- Overcoming limitations of limited amounts of data

6

6

SVEM Field Trials: What We Have Learned

- SVEM works exceptionally well!
  - n < p
  - Higher order models
  - More accurate predictions
  - Recover from broken DOEs
    - DSDs that can't be run in Fit Definitive
    - Lacking power
    - Missing or questionable runs
- Does not help in Fractional Factorials

SVEM Field Trials: What We Have Learned

- SVEM add-in full potential not tested
  - Need to design experiments with SVEM in mind
  - More factors
  - Bayesian I-optimal designs
  - Mixture designs

SVEM: What, Why, and How

Chris Gotwalt, JMP
Phil Ramsey, Univ. New Hampshire
Trent Lemkus, Univ. New Hampshire

9

SVEM: What, Why, and How

10

MACHINE LEARNING 101

	Compaction Pressure	Sintering Time	Sintering Temp C
1	75	30.80	1122
2	80	27.48	1111
3	100	27.48	1127
4	100	27.48	1118
5	85	20.85	1221
6	100	27.48	1118
7	95	27.48	1138
8	65	30.80	1122
9	75	30.80	1113
10	70	30.80	1127
11	90	20.85	1221
12	70	30.80	1221

$$f(pressure, time, temp) \longrightarrow$$

	Shrinkage
1	-0.37
2	-0.58
3	-0.16
4	0.05
5	-1.50
6	-0.09
7	0.55
8	-1.43
9	-0.93
10	-0.79
11	-1.22
12	-1.99

jmp Statistical Discovery.™ From SAS.

sas THE POWER TO KNOW.

11

MACHINE LEARNING 101

	Compaction Pressure	Sintering Time	Sintering Temp C
1	75	30.80	1122
2	80	27.48	1111
3	100	27.48	1127
4	100	27.48	1118
5	85	20.85	1221
6	100	27.48	1118
7	95	27.48	1138
8	65	30.80	1122

Fit models to the training set

	Shrinkage
1	-0.37
2	-0.58
3	-0.16
4	0.05
5	-1.50
6	-0.09
7	0.55
8	-1.43

9	75	30.80	1113
10	70	30.80	1127
11	90	20.85	1221
12	70	30.80	1221

Assess models using the validation set

9	-0.93
10	-0.79
11	-1.22
12	-1.99

jmp Statistical Discovery.™ From SAS.

sas THE POWER TO KNOW.

12

DOE MODELING 101

	Compaction Pressure	Sintering Time	Sintering Temp
1	60	15	1000
2	60	15	1000
3	60	15	1250
4	60	30	1000
5	60	30	1250
6	60	30	1250
7	115	15	1000
8	115	15	1250
9	115	15	1250
10	115	30	1000
11	115	30	1000
12	115	30	1250

$$f(pressure, time, temp) \longrightarrow$$

	Shrinkage
1	-1.85
2	-2.01
3	-3.75
4	-1.74
5	-3.8
6	-3.74
7	-1.95
8	-3.32
9	-2.79
10	-1.23
11	-1.31
12	-2.35

jmp Statistical Discovery.™ From SAS.

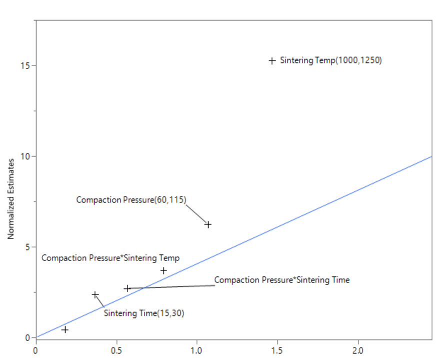
13

DOE MODELING 101

Effect Tests

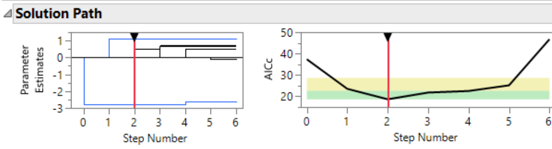
Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Compaction Pressure(60,115)	1	1	1.1009611	33.0087	0.0022*
Sintering Time(15,30)	1	1	0.4005065	12.0079	0.0179*
Sintering Temp(1000,1250)	1	1	6.0741458	182.1132	<.0001*
Compaction Pressure*Sintering Time	1	1	0.2442070	7.3217	0.0425*
Compaction Pressure*Sintering Temp	1	1	0.4566092	13.6899	0.0140*
Sintering Time*Sintering Temp	1	1	0.0064027	0.1920	0.6796

Half-Normal Plot



Forward Selection via AICc

Solution Path



Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F	
Sintering Temp(1000,1250)	1	1	7.777386	65.978133	<.0001*	
Compaction Pressure(60,115)	1	1	1.3023508	11.048271	0.0089*	
Sintering Time(15,30)	1	0	0	0	1.0000	Removed
Compaction Pressure*Sintering Time	1	0	0	0	1.0000	Removed
Compaction Pressure*Sintering Temp	1	0	0	0	1.0000	Removed
Sintering Time*Sintering Temp	1	0	0	0	1.0000	Removed

jmp Statistical Discovery.™ From SAS.

14

WHY NOT APPLY MACHINE LEARNING TO DOEs?

Training Set

	Compaction Pressure	Sintering Time	Sintering Temp
1	60	15	1000
2	60	15	1000
3	60	15	1250
4	60	30	1000
5	60	30	1250
6	60	30	1250
7	115	15	1000
8	115	15	1250

	Shrinkage
1	-1.85
2	-2.01
3	-3.75
4	-1.74
5	-3.8
6	-3.74
7	-1.95
8	-3.32

Validation Set

9	115	15	1250
10	115	30	1000
11	115	30	1000
12	115	30	1250

9	-2.79
10	-1.23
11	-1.31
12	-2.35

jmp Statistical Discovery.™ From SAS.

sas THE POWER TO KNOW.

15

WHY NOT APPLY MACHINE LEARNING TO DOEs?

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F	
Sintering Temp(1000,1250)	1	1	3.2802443	240.26823	0.0041*	
Compaction Pressure*Sintering Temp	1	1	0.0464416	3.4017103	0.2064	
Sintering Time*Sintering Temp	1	1	0.002799	0.205015	0.6951	
Compaction Pressure(60,115)	1	0	0	0	1.0000	LostDFs
Sintering Time(15,30)	1	0	0	0	1.0000	LostDFs
Compaction Pressure*Sintering Time	1	0	0	0	1.0000	LostDFs

jmp Statistical Discovery.™ From SAS.

sas THE POWER TO KNOW.

16

WHY NOT APPLY MACHINE LEARNING TO DOEs?

	Compaction Pressure	Sintering Time	Sintering Temp	Shrinkage
1	60	15	1000	-1.85
2	60	15	1000	-2.01
3	60	15	1250	-3.75
4	60	30	1000	-1.74
5	60	30	1250	-3.8
6	60	30	1250	-3.74
7	115	15	1000	-1.95
8	115	15	1250	-3.32
9	115	15	1250	-2.79
10	115	30	1000	-1.23
11	115	30	1000	-1.31
12	115	30	1250	-2.35

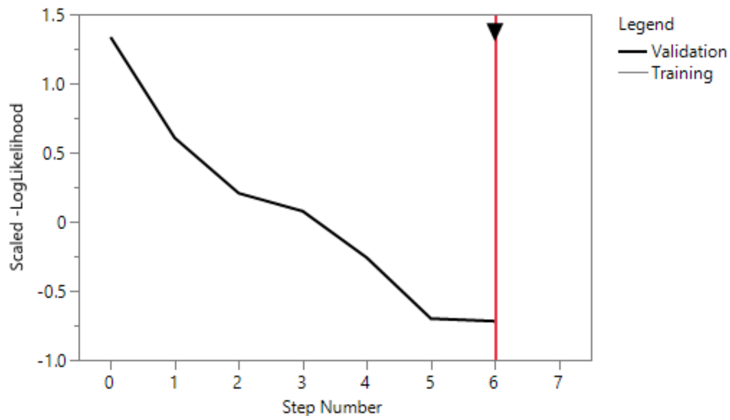
17

WHY NOT APPLY MACHINE LEARNING TO DOEs?

	Compaction Pressure	Sintering Time	Sintering Temp	Shrinkage	Validation
1	60	15	1000	-1.85	Training
2	60	15	1000	-1.85	Validation
3	60	15	1000	-2.01	Training
4	60	15	1000	-2.01	Validation
5	60	15	1250	-3.75	Training
6	60	15	1250	-3.75	Validation
7	60	30	1000	-1.74	Training
8	60	30	1000	-1.74	Validation
9	60	30	1250	-3.8	Training
10	60	30	1250	-3.8	Validation
11	60	30	1250	-3.74	Training
12	60	30	1250	-3.74	Validation
13	115	15	1000	-1.95	Training
14	115	15	1000	-1.95	Validation
15	115	15	1250	-3.32	Training
16	115	15	1250	-3.32	Validation
17	115	15	1250	-2.79	Training
18	115	15	1250	-2.79	Validation
19	115	30	1000	-1.23	Training
20	115	30	1000	-1.23	Validation
21	115	30	1000	-1.31	Training
22	115	30	1000	-1.31	Validation
23	115	30	1250	-2.35	Training
24	115	30	1250	-2.35	Validation

18

WHY NOT APPLY MACHINE LEARNING TO DOEs?



jmp Statistical Discovery.™ From SAS.

19

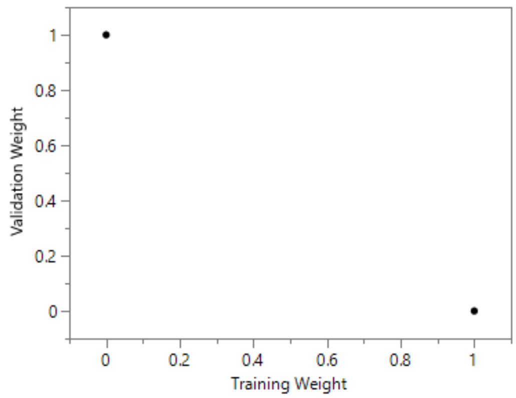
HOLDBACK AS A WEIGHTING SCHEME

	Compaction Pressure	Sintering Time	Sintering Temp	Shrinkage	Training Weight	Validation Weight
1	60	15	1000	-1.85	1	0
2	60	15	1000	-2.01	1	0
3	60	15	1250	-3.75	1	0
4	60	30	1000	-1.74	1	0
5	60	30	1250	-3.8	1	0
6	60	30	1250	-3.74	1	0
7	115	15	1000	-1.95	1	0
8	115	15	1250	-3.32	1	0
9	115	15	1250	-2.79	0	1
10	115	30	1000	-1.23	0	1
11	115	30	1000	-1.31	0	1
12	115	30	1250	-2.35	0	1

jmp Statistical Discovery.™ From SAS.

20

HOLDBACK AS A WEIGHTING SCHEME

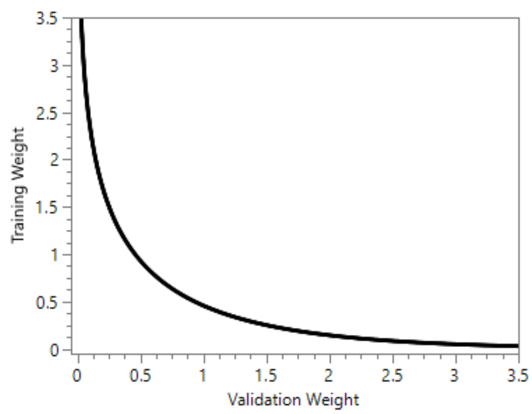


jmp Statistical Discovery.™ From SAS.

sas THE POWER TO KNOW.

21

HOLDBACK AS A WEIGHTING SCHEME



jmp Statistical Discovery.™ From SAS.

sas THE POWER TO KNOW.

22

AUTOVALIDATION WEIGHTS

Create exponentially distributed weights by the *probability integral transform:*

u_i ~ Uniform(0,1)
w_i^T = *Exponential Quantile*(u_i , 1)

$w_i^{Training}$ has an Exponential distribution with mean and variance 1.0.

$1 - u_i$ is also Uniform(0,1) *and perfectly anticorrelated with* u_i

w_i^V = *Exponential Quantile*($1 - u_i$, 1)
w_i^V will be highly anticorrelated with w_i^T and yet has the exact same distribution!

jmp Statistical Discovery.™ From SAS.

SAS THE POWER TO KNOW.

23

AUTOVALIDATION WEIGHTS

	Compaction Pressure	Sintering Time	Sintering Temp	Shrinkage	U	Training Weight	Validation Weight
1	60	15	1000	-1.85	0.077	2.565	0.08
2	60	15	1000	-2.01	0.154	1.872	0.167
3	60	15	1250	-3.75	0.231	1.466	0.262
4	60	30	1000	-1.74	0.308	1.179	0.368
5	60	30	1250	-3.8	0.385	0.956	0.486
6	60	30	1250	-3.74	0.462	0.773	0.619
7	115	15	1000	-1.95	0.538	0.619	0.773
8	115	15	1250	-3.32	0.615	0.486	0.956
9	115	15	1250	-2.79	0.692	0.368	1.179
10	115	30	1000	-1.23	0.769	0.262	1.466
11	115	30	1000	-1.31	0.846	0.167	1.872
12	115	30	1250	-2.35	0.923	0.08	2.565

jmp Statistical Discovery.™ From SAS.

Copyright © 2018, SAS Institute, Inc. All rights reserved.

SAS THE POWER TO KNOW.

24

ROAD TO SVEM

	Compaction Pressure	Sintering Time	Sintering Temp	Shrinkage	U	Training Weight	Validation Weight
1	60	15	1000	-1.85	0.804	0.218	1.63
2	60	15	1000	-2.01	0.033	3.426	0.033
3	60	15	1250	-3.75	0.624	0.472	0.978
4	60	30	1000	-1.74	0.972	0.028	3.585
5	60	30	1250	-3.8	0.992	0.008	4.85
6	60	30	1250	-3.74	0.146	1.922	0.158
7	115	15	1000	-1.95	0.046	3.082	0.047
8	115	15	1250	-3.32	0.362	1.017	0.449
9	115	15	1250	-2.79	0.154	1.871	0.167
10	115	30	1000	-1.23	0.052	2.955	0.053
11	115	30	1000	-1.31	0.019	3.962	0.019
12	115	30	1250	-2.35	0.587	0.533	0.884

jmp Statistical Discovery.™ From SAS.

sas THE POWER TO KNOW.

25

ROAD TO SVEM

	Compaction Pressure	Sintering Time	Sintering Temp	Shrinkage	U	Training Weight	Validation Weight	Shrinkage Prediction Formula 1
1	60	15	1000	-1.85	0.804	0.218	1.63	-2.01
2	60	15	1000	-2.01	0.033	3.426	0.033	-2.01
3	60	15	1250	-3.75	0.624	0.472	0.978	-3.62
4	60	30	1000	-1.74	0.972	0.028	3.585	-2.01
5	60	30	1250	-3.8	0.992	0.008	4.85	-3.62
6	60	30	1250	-3.74	0.146	1.922	0.158	-3.62
7	115	15	1000	-1.95	0.046	3.082	0.047	-1.351
8	115	15	1250	-3.32	0.362	1.017	0.449	-2.962
9	115	15	1250	-2.79	0.154	1.871	0.167	-2.962
10	115	30	1000	-1.23	0.052	2.955	0.053	-1.351
11	115	30	1000	-1.31	0.019	3.962	0.019	-1.351
12	115	30	1250	-2.35	0.587	0.533	0.884	-2.962

jmp Statistical Discovery.™ From SAS.

sas THE POWER TO KNOW.

26

ROAD TO SVEM

$$-2.485909079 + 0.3294377473 \cdot \left(\frac{(\text{Compaction Pressure} - 87.5)}{27.5} \right) + -0.80505621 \cdot \left(\frac{(\text{Sintering Temp} - 1125)}{125} \right)$$

27

ROAD TO SVEM

	Compaction Pressure	Sintering Time	Sintering Temp	Shrinkage	U	Training Weight	Validation Weight	Shrinkage Prediction Formula 1
1	60	15	1000	-1.85	0.804	0.218	1.63	-2.01
2	60	15	1000	-2.01	0.033	3.426	0.033	-2.01
3	60	15	1250	-3.75	0.624	0.472	0.978	-3.62
4	60	30	1000	-1.74	0.972	0.028	3.585	-2.01
5	60	30	1250	-3.8	0.992	0.008	4.85	-3.62
6	60	30	1250	-3.74	0.146	1.922	0.158	-3.62
7	115	15	1000	-1.95	0.046	3.082	0.047	-1.351
8	115	15	1250	-3.32	0.362	1.017	0.449	-2.962
9	115	15	1250	-2.79	0.154	1.871	0.167	-2.962
10	115	30	1000	-1.23	0.052	2.955	0.053	-1.351
11	115	30	1000	-1.31	0.019	3.962	0.019	-1.351
12	115	30	1250	-2.35	0.587	0.533	0.884	-2.962

28

ROAD TO SVEM

	Compaction Pressure	Sintering Time	Sintering Temp	Shrinkage	U	Training Weight	Validation Weight	Shrinkage Prediction Formula 1	Shrinkage Prediction Formula 2
1	60	15	1000	-1.85	0.529	0.637	0.753	-2.01	-1.902
2	60	15	1000	-2.01	0.055	2.9	0.057	-2.01	-1.902
3	60	15	1250	-3.75	0.637	0.451	1.014	-3.62	-3.82
4	60	30	1000	-1.74	0.714	0.337	1.251	-2.01	-1.826
5	60	30	1250	-3.8	0.93	0.073	2.654	-3.62	-3.745
6	60	30	1250	-3.74	0.721	0.326	1.278	-3.62	-3.745
7	115	15	1000	-1.95	0.466	0.763	0.628	-1.351	-1.951
8	115	15	1250	-3.32	0.254	1.372	0.293	-2.962	-3.082
9	115	15	1250	-2.79	0.644	0.44	1.033	-2.962	-3.082
10	115	30	1000	-1.23	0.694	0.366	1.183	-1.351	-1.229
11	115	30	1000	-1.31	0.627	0.467	0.985	-1.351	-1.229
12	115	30	1250	-2.35	0.231	1.464	0.263	-2.962	-2.36

Statistical Discovery.™ From SAS.

ROAD TO SVEM

$$
\begin{aligned}
&-2.489456343 \\
&+ 0.3340338147 \cdot \left(\frac{(\mathit{Compaction\ Pressure} - 87.5)}{27.5} \right) \\
&+ 0.1993742337 \cdot \left(\frac{(\mathit{Sintering\ Time} - 22.5)}{7.5} \right) \\
&+ -0.762447346 \cdot \left(\frac{(\mathit{Sintering\ Temp} - 1125)}{125} \right) \\
&+ \left(\frac{(\mathit{Compaction\ Pressure} - 87.5)}{27.5} \right) \cdot \left(\left(\frac{(\mathit{Sintering\ Time} - 22.5)}{7.5} \right) \cdot 0.1615679978 \right) \\
&+ \left(\frac{(\mathit{Compaction\ Pressure} - 87.5)}{27.5} \right) \cdot \left(\left(\frac{(\mathit{Sintering\ Temp} - 1125)}{125} \right) \cdot 0.1967542111 \right)
\end{aligned}
$$

Statistical Discovery.™ From SAS.

ROAD TO SVEM

	Compaction Pressure	Sintering Time	Sintering Temp	Shrinkage	U	Training Weight	Validation Weight	Shrinkage Prediction Formula 1	Shrinkage Prediction Formula 2
1	60	15	1000	-1.85	0.529	0.637	0.753	-2.01	-1.902
2	60	15	1000	-2.01	0.055	2.9	0.057	-2.01	-1.902
3	60	15	1250	-3.75	0.637	0.451	1.014	-3.62	-3.82
4	60	30	1000	-1.74	0.714	0.337	1.251	-2.01	-1.826
5	60	30	1250	-3.8	0.93	0.073	2.654	-3.62	-3.745
6	60	30	1250	-3.74	0.721	0.326	1.278	-3.62	-3.745
7	115	15	1000	-1.95	0.466	0.763	0.628	-1.351	-1.951
8	115	15	1250	-3.32	0.254	1.372	0.293	-2.962	-3.082
9	115	15	1250	-2.79	0.644	0.44	1.033	-2.962	-3.082
10	115	30	1000	-1.23	0.694	0.366	1.183	-1.351	-1.229
11	115	30	1000	-1.31	0.627	0.467	0.985	-1.351	-1.229
12	115	30	1250	-2.35	0.231	1.464	0.263	-2.962	-2.36

jmp Statistical Discovery.™ From SAS.

sas THE POWER TO KNOW.

31

ROAD TO SVEM

	Compaction Pressure	Sintering Time	Sintering Temp	Shrinkage	U	Training Weight	Validation Weight	Shrinkage Prediction Formula 1	Shrinkage Prediction Formula 2	Shrinkage Prediction Formula 3
1	60	15	1000	-1.85	0.033	3.418	0.033	-2.01	-1.902	-1.885
2	60	15	1000	-2.01	0.146	1.925	0.158	-2.01	-1.902	-1.885
3	60	15	1250	-3.75	0.367	1.001	0.458	-3.62	-3.82	-3.885
4	60	30	1000	-1.74	0.093	2.379	0.097	-2.01	-1.826	-1.746
5	60	30	1250	-3.8	0.027	3.606	0.028	-3.62	-3.745	-3.747
6	60	30	1250	-3.74	0.425	0.857	0.553	-3.62	-3.745	-3.747
7	115	15	1000	-1.95	0.761	0.273	1.431	-1.351	-1.951	-1.935
8	115	15	1250	-3.32	0.189	1.667	0.209	-2.962	-3.082	-3.031
9	115	15	1250	-2.79	0.416	0.876	0.539	-2.962	-3.082	-3.031
10	115	30	1000	-1.23	0.687	0.375	1.162	-1.351	-1.229	-1.255
11	115	30	1000	-1.31	0.56	0.581	0.82	-1.351	-1.229	-1.255
12	115	30	1250	-2.35	0.982	0.018	4.026	-2.962	-2.36	-2.35

jmp Statistical Discovery.™ From SAS.

Copyright © 2018, SAS Institute, Inc. All rights reserved.

sas THE POWER TO KNOW.

32

ROAD TO SVEM

	Compaction Pressure	Sintering Time	Sintering Temp	Shrinkage	U	Training Weight	Validation Weight	Shrinkage Prediction Formula 1	Shrinkage Prediction Formula 2	Shrinkage Prediction Formula 3	Shrinkage Prediction Formula 4
1	60	15	1000	-1.85	0.746	0.293	1.372	-2.01	-1.902	-1.885	-1.921
2	60	15	1000	-2.01	0.315	1.157	0.378	-2.01	-1.902	-1.885	-1.921
3	60	15	1250	-3.75	0.163	1.814	0.178	-3.62	-3.82	-3.885	-3.797
4	60	30	1000	-1.74	0.474	0.747	0.642	-2.01	-1.826	-1.746	-1.856
5	60	30	1250	-3.8	0.286	1.253	0.336	-3.62	-3.745	-3.747	-3.731
6	60	30	1250	-3.74	0.896	0.11	2.265	-3.62	-3.745	-3.747	-3.731
7	115	15	1000	-1.95	0.184	1.69	0.204	-1.351	-1.951	-1.935	-2.025
8	115	15	1250	-3.32	0.075	2.596	0.078	-2.962	-3.082	-3.031	-3.236
9	115	15	1250	-2.79	0.838	0.177	1.819	-2.962	-3.082	-3.031	-3.236
10	115	30	1000	-1.23	0.452	0.793	0.602	-1.351	-1.229	-1.255	-1.237
11	115	30	1000	-1.31	0.133	2.014	0.143	-1.351	-1.229	-1.255	-1.237
12	115	30	1250	-2.35	0.252	1.38	0.29	-2.962	-2.36	-2.35	-2.448

jmp Statistical Discovery.™ From SAS.

sas THE POWER TO KNOW.

33

ROAD TO SVEM

	Compaction Pressure	Sintering Time	Sintering Temp	Shrinkage	U	Training Weight	Validation Weight	Shrinkage Prediction Formula 1	Shrinkage Prediction Formula 2	Shrinkage Prediction Formula 3	Shrinkage Prediction Formula 4	Shrinkage Prediction Formula 5
1	60	15	1000	-1.85	0.096	2.345	0.101	-2.01	-1.902	-1.885	-1.921	-1.873
2	60	15	1000	-2.01	0.671	0.399	1.112	-2.01	-1.902	-1.885	-1.921	-1.873
3	60	15	1250	-3.75	0.161	1.827	0.175	-3.62	-3.82	-3.885	-3.797	-3.749
4	60	30	1000	-1.74	0.987	0.013	4.322	-2.01	-1.826	-1.746	-1.856	-1.707
5	60	30	1250	-3.8	0.592	0.524	0.898	-3.62	-3.745	-3.747	-3.731	-3.778
6	60	30	1250	-3.74	0.736	0.307	1.331	-3.62	-3.745	-3.747	-3.731	-3.778
7	115	15	1000	-1.95	0.797	0.227	1.594	-1.351	-1.951	-1.935	-2.025	-1.947
8	115	15	1250	-3.32	0.932	0.07	2.693	-2.962	-3.082	-3.031	-3.236	-2.82
9	115	15	1250	-2.79	0.342	1.073	0.418	-2.962	-3.082	-3.031	-3.236	-2.82
10	115	30	1000	-1.23	0.694	0.366	1.183	-1.351	-1.229	-1.255	-1.237	-1.287
11	115	30	1000	-1.31	0.3	1.203	0.357	-1.351	-1.229	-1.255	-1.237	-1.287
12	115	30	1250	-2.35	0.395	0.928	0.503	-2.962	-2.36	-2.35	-2.448	-2.354

jmp Statistical Discovery.™ From SAS.

sas THE POWER TO KNOW.

34

ROAD TO SVEM

$$\text{Mean}\left(\begin{array}{l}\textit{Shrinkage Prediction Formula 1},\\ \textit{Shrinkage Prediction Formula 2},\\ \textit{Shrinkage Prediction Formula 3},\\ \textit{Shrinkage Prediction Formula 4},\\ \textit{Shrinkage Prediction Formula 5}\end{array}\right)$$

jmp Statistical Discovery.™ From SAS.

sas THE POWER TO KNOW.

35

ROAD TO SVEM

	Compaction Pressure	Sintering Time	Sintering Temp	Shrinkage	U	Training Weight	Validation Weight	Shrinkage Prediction Formula 1	Shrinkage Prediction Formula 2	Shrinkage Prediction Formula 3	Shrinkage Prediction Formula 4	Shrinkage Prediction Formula 5
1	60	15	1000	-1.85	0.096	2.345	0.101	-2.01	-1.902	-1.885	-1.921	-1.873
2	60	15	1000	-2.01	0.671	0.399	1.112	-2.01	-1.902	-1.885	-1.921	-1.873
3	60	15	1250	-3.75	0.161	1.827	0.175	-3.62	-3.82	-3.885	-3.797	-3.749
4	60	30	1000	-1.74	0.987	0.013	4.322	-2.01	-1.826	-1.746	-1.856	-1.707
5	60	30	1250	-3.8	0.592	0.524	0.898	-3.62	-3.745	-3.747	-3.731	-3.778
6	60	30	1250	-3.74	0.736	0.307	1.331	-3.62	-3.745	-3.747	-3.731	-3.778
7	115	15	1000	-1.95	0.797	0.227	1.594	-1.351	-1.951	-1.935	-2.025	-1.947
8	115	15	1250	-3.32	0.932	0.07	2.693	-2.962	-3.082	-3.031	-3.236	-2.82
9	115	15	1250	-2.79	0.342	1.073	0.418	-2.962	-3.082	-3.031	-3.236	-2.82
10	115	30	1000	-1.23	0.694	0.366	1.183	-1.351	-1.229	-1.255	-1.237	-1.287
11	115	30	1000	-1.31	0.3	1.203	0.357	-1.351	-1.229	-1.255	-1.237	-1.287
12	115	30	1250	-2.35	0.395	0.928	0.503	-2.962	-2.36	-2.35	-2.448	-2.354

jmp Statistical Discovery.™ From SAS.

sas THE POWER TO KNOW.

36

ROAD TO SVEM

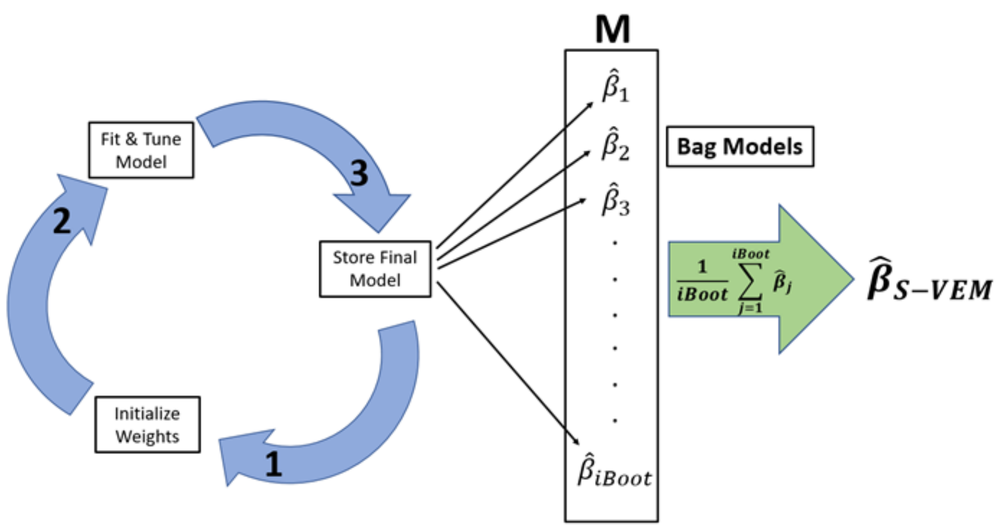
	Compaction Pressure	Sintering Time	Sintering Temp	Shrinkage	U	Training Weight	Validation Weight	Shrinkage Prediction Formula 1	Shrinkage Prediction Formula 2	Shrinkage Prediction Formula 3	Shrinkage Prediction Formula 4	Shrinkage Prediction Formula 5	SVEM Prediction
1	60	15	1000	-1.85	0.096	2.345	0.101	-2.01	-1.902	-1.885	-1.921	-1.873	-1.9183
2	60	15	1000	-2.01	0.671	0.399	1.112	-2.01	-1.902	-1.885	-1.921	-1.873	-1.9183
3	60	15	1250	-3.75	0.161	1.827	0.175	-3.62	-3.82	-3.885	-3.797	-3.749	-3.7744
4	60	30	1000	-1.74	0.987	0.013	4.322	-2.01	-1.826	-1.746	-1.856	-1.707	-1.8293
5	60	30	1250	-3.8	0.592	0.524	0.898	-3.62	-3.745	-3.747	-3.731	-3.778	-3.7242
6	60	30	1250	-3.74	0.736	0.307	1.331	-3.62	-3.745	-3.747	-3.731	-3.778	-3.7242
7	115	15	1000	-1.95	0.797	0.227	1.594	-1.351	-1.951	-1.935	-2.025	-1.947	-1.842
8	115	15	1250	-3.32	0.932	0.07	2.693	-2.962	-3.082	-3.031	-3.236	-2.82	-3.026
9	115	15	1250	-2.79	0.342	1.073	0.418	-2.962	-3.082	-3.031	-3.236	-2.82	-3.026
10	115	30	1000	-1.23	0.694	0.366	1.183	-1.351	-1.229	-1.255	-1.237	-1.287	-1.272
11	115	30	1000	-1.31	0.3	1.203	0.357	-1.351	-1.229	-1.255	-1.237	-1.287	-1.272
12	115	30	1250	-2.35	0.395	0.928	0.503	-2.962	-2.36	-2.35	-2.448	-2.354	-2.4948

jmp Statistical Discovery.™ From SAS.

sas THE POWER TO KNOW.

37

SVEM

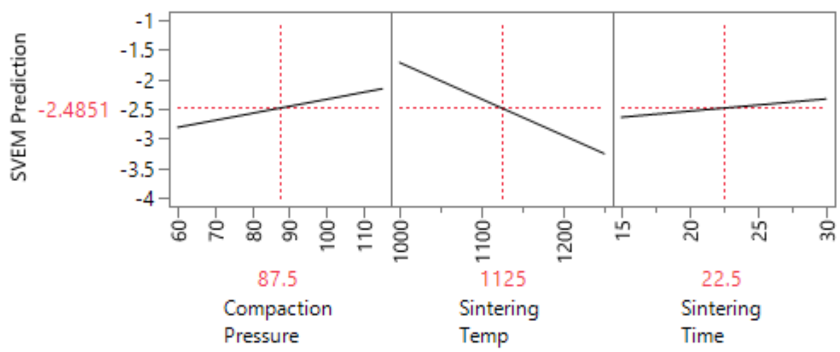


jmp Statistical Discovery.™ From SAS.

sas THE POWER TO KNOW.

38

ROAD TO SVEM



jmp Statistical Discovery.™ From SAS.

39

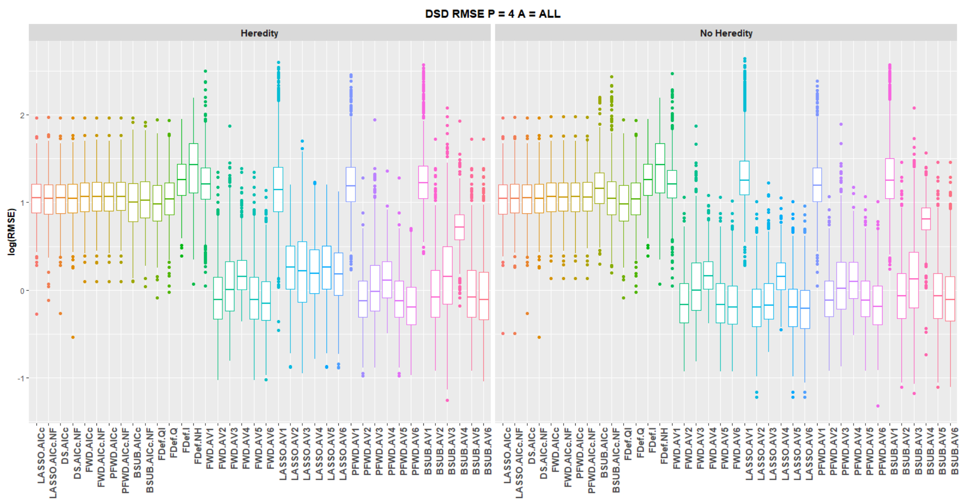
SIMULATIONS

- Definitive Screening and Box Behnken Designs in 4 and 8 DoE factors
- Tried many "classical" and autovalidation based modeling approaches, all based on quadratic RSM.
- 1000 simulation reps per situation
- Each sim rep had its own set of "true" model coefficients
  - Nonzero model coefficients were double exponentially distributed
  - "True" nonzero coefficients represented 50%-100% of all possible coefficients
- Models evaluated on an independent set RMSE (n=10k, spacefilling design) vs. true model

jmp Statistical Discovery.™ From SAS.

40

SIMULATIONS

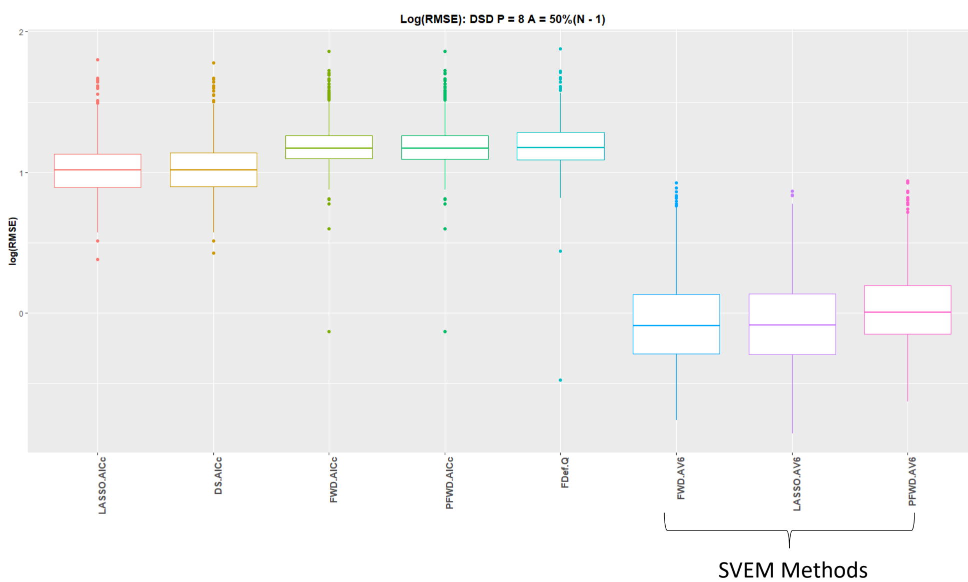


jmp Statistical Discovery.™ From SAS.

sas THE POWER TO KNOW.

41

SIMULATIONS

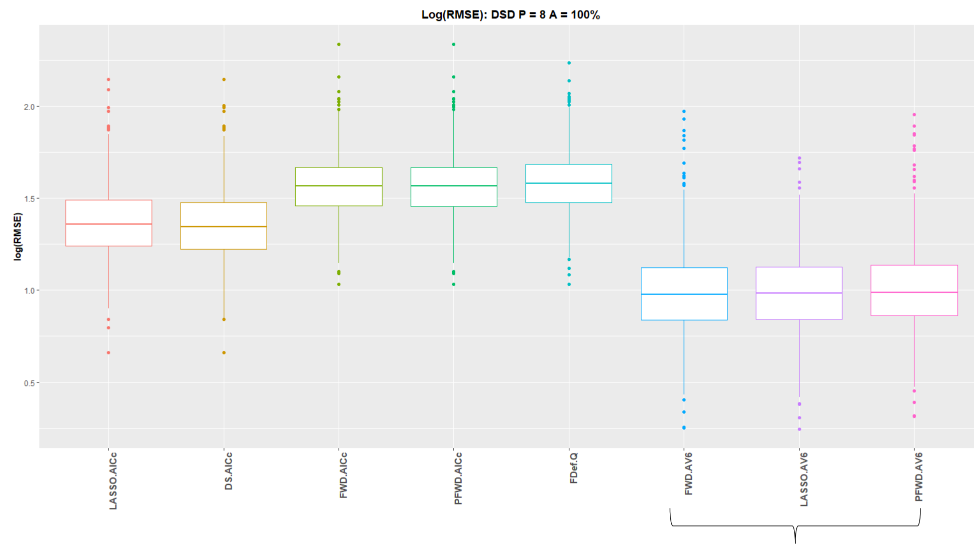


jmp Statistical Discovery.™ From SAS.

sas THE POWER TO KNOW.

42

SIMULATIONS

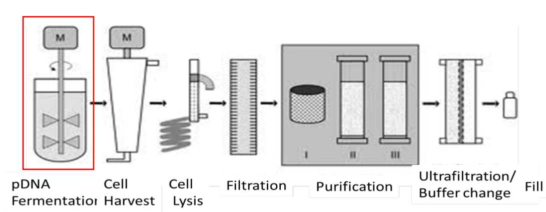


jmp Statistical Discovery.™ From SAS.

sas THE POWER TO KNOW.

43

THE PLASMID DNA (pDNA) EXPERMENT



Fermentation
Genome *E. coli*

→ **Cell Harvest**
Continuous centrifugation

→ **Cell Lysis**
Continuous chemical lysis

→ **Clarification & Concentration**
Flocculation / RNA Precipitation

→ **Purification**
HPLC based

→ **Final Product**
pDNA

jmp Statistical Discovery.™ From SAS.

sas THE POWER TO KNOW.

44

THE PLASMID DNA (pDNA) EXPERMENT

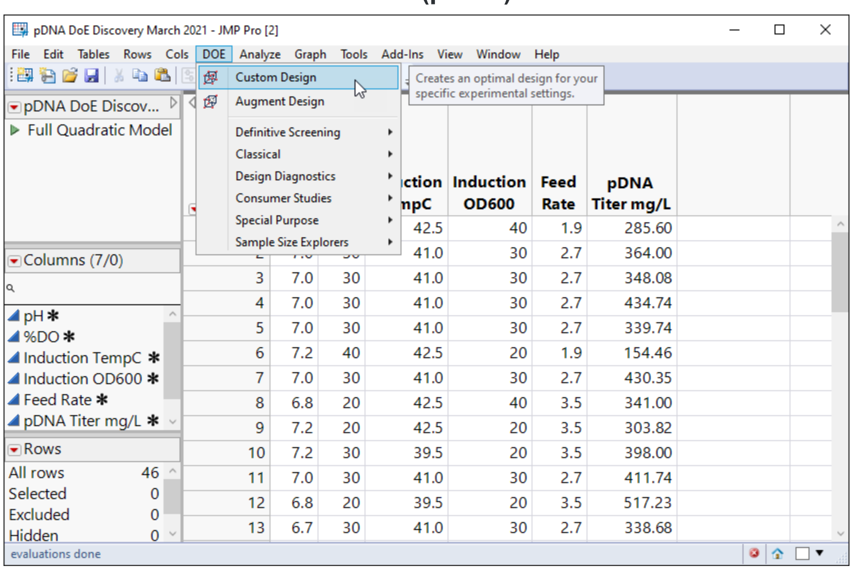
	pH	%DO	Induction TempC	Induction OD600	Feed Rate	pDNA Titer mg/L
1	6.8	30	42.5	40	1.9	285.60
2	7.0	30	41.0	30	2.7	364.00
3	7.0	30	41.0	30	2.7	348.08
4	7.0	30	41.0	30	2.7	434.74
5	7.0	30	41.0	30	2.7	339.74
6	7.2	40	42.5	20	1.9	154.46
7	7.0	30	41.0	30	2.7	430.35
8	6.8	20	42.5	40	3.5	341.00
9	7.2	20	42.5	20	3.5	303.82
10	7.2	30	39.5	20	3.5	398.00
11	7.0	30	41.0	30	2.7	411.74
12	6.8	20	39.5	20	3.5	517.23
13	6.7	30	41.0	30	2.7	338.68
14	7.2	20	41.0	40	1.9	229.00
15	7.0	30	41.0	17	2.7	282.29
16	6.8	40	41.0	20	3.5	377.00

jmp Statistical Discovery.™ From SAS.

sas THE POWER TO KNOW.

45

THE PLASMID DNA (pDNA) EXPERMENT

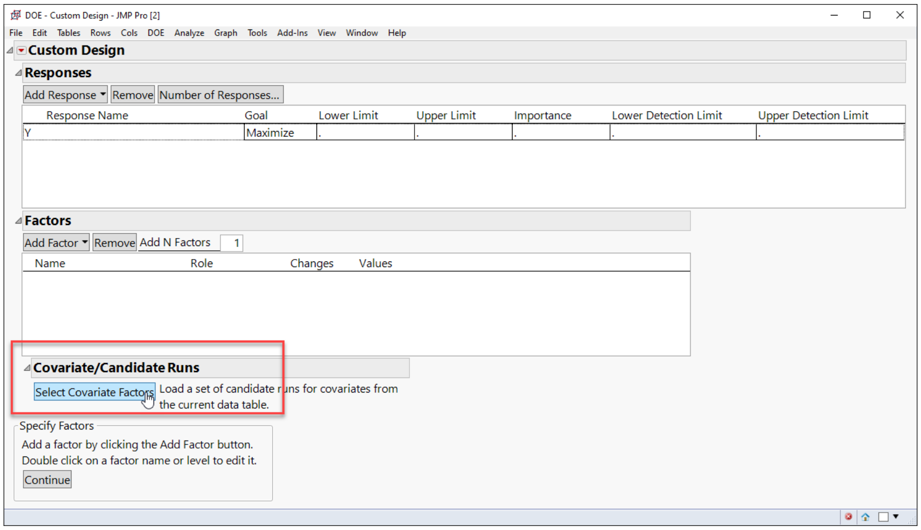


jmp Statistical Discovery.™ From SAS.

sas THE POWER TO KNOW.

46

THE PLASMID DNA (pDNA) EXPERMENT

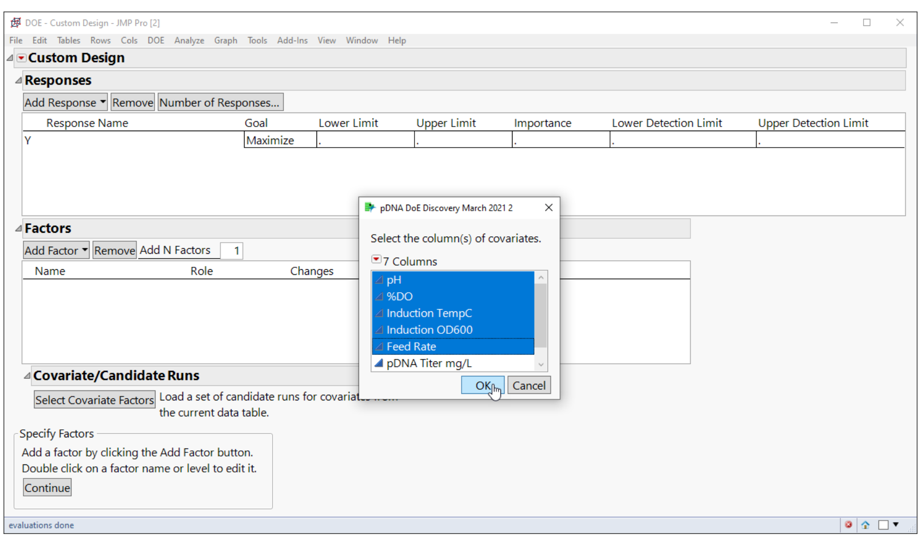


jmp Statistical Discovery.™ From SAS.

sas THE POWER TO KNOW.

47

THE PLASMID DNA (pDNA) EXPERMENT

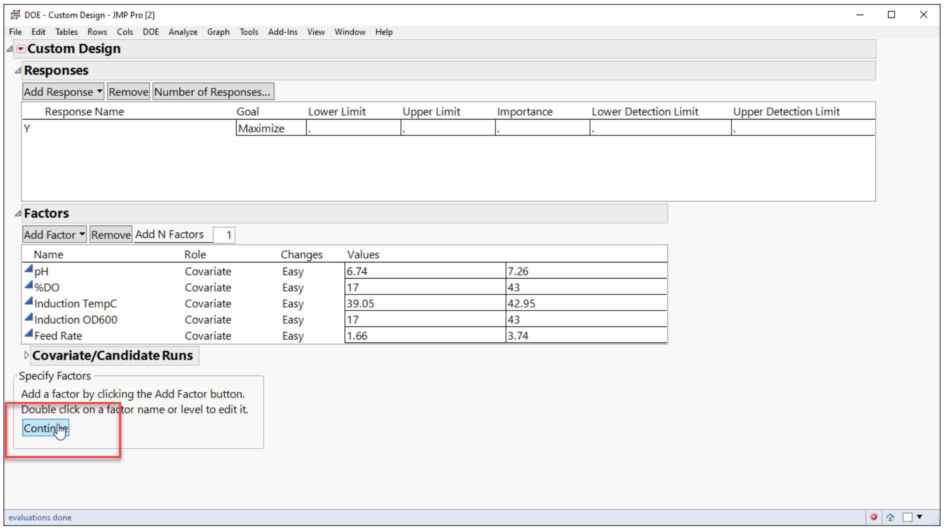


jmp Statistical Discovery.™ From SAS.

sas THE POWER TO KNOW.

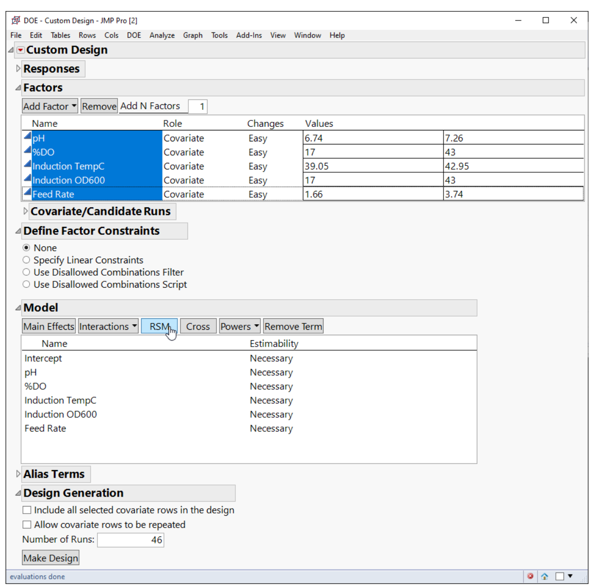
48

THE PLASMID DNA (pDNA) EXPERMENT



49

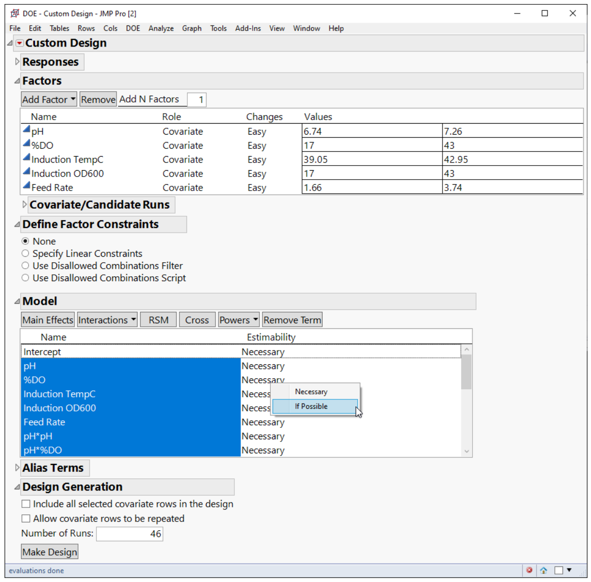
THE PLASMID DNA (pDNA) EXPERMENT



jmp Statistical Discovery.™ From SAS.
Copyright © 2018, SAS Institute, Inc. All rights reserved.

50

THE PLASMID DNA (pDNA) EXPERMENT

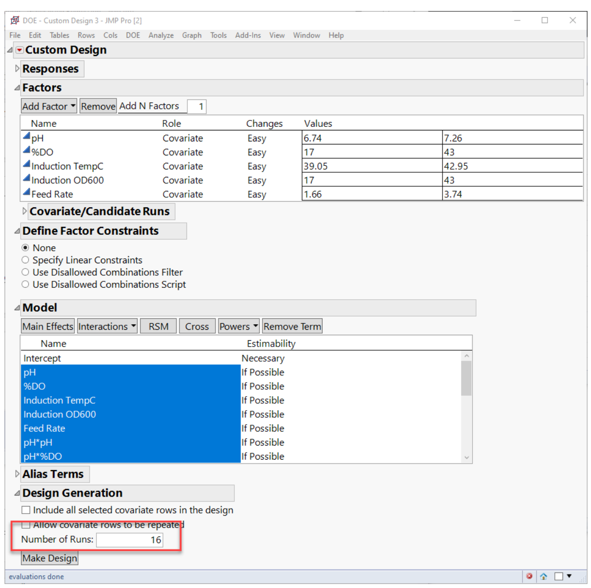


jmp Statistical Discovery.™ From SAS.

sas THE POWER TO KNOW.

51

THE PLASMID DNA (pDNA) EXPERMENT

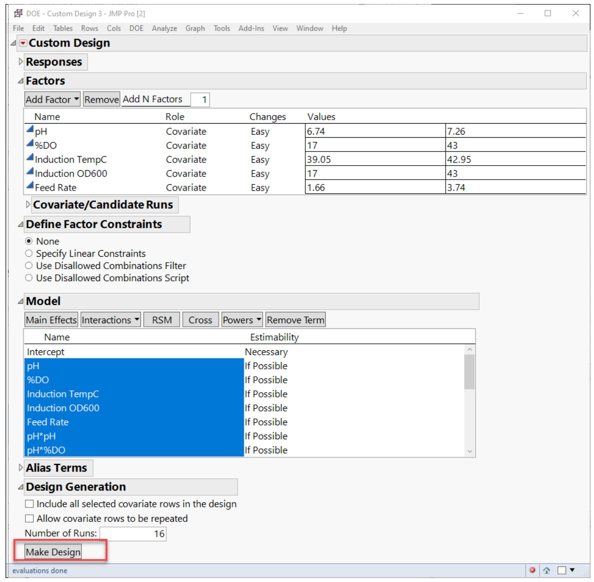


jmp Statistical Discovery.™ From SAS.

sas THE POWER TO KNOW.

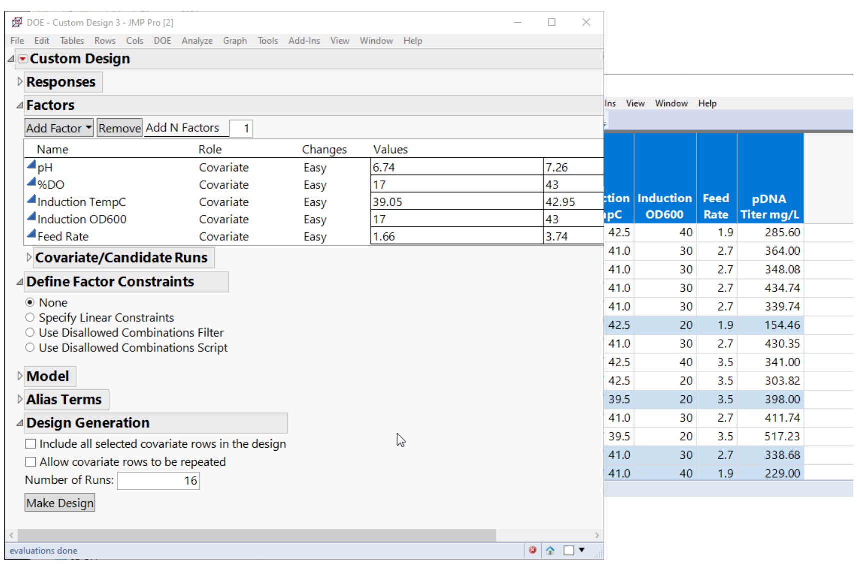
52

THE PLASMID DNA (pDNA) EXPERMENT



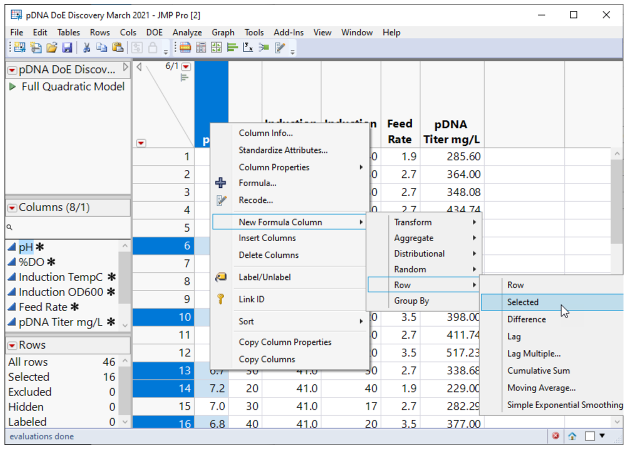
53

THE PLASMID DNA (pDNA) EXPERMENT



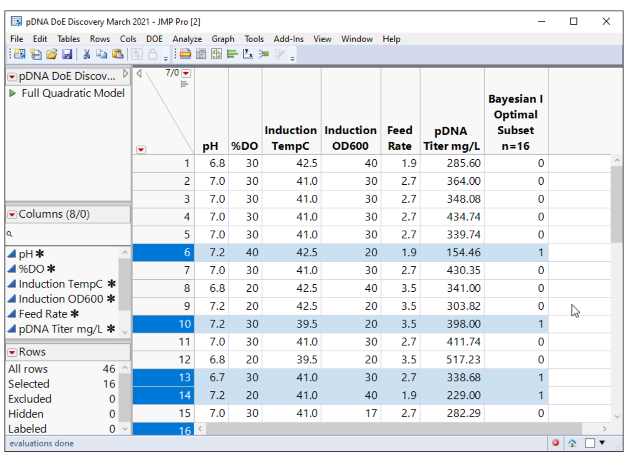
54

THE PLASMID DNA (pDNA) EXPERMENT



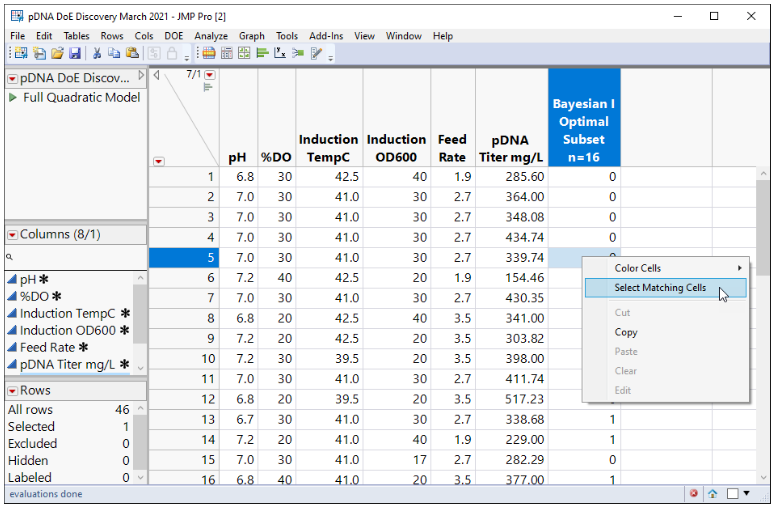
55

THE PLASMID DNA (pDNA) EXPERMENT



56

THE PLASMID DNA (pDNA) EXPERMENT



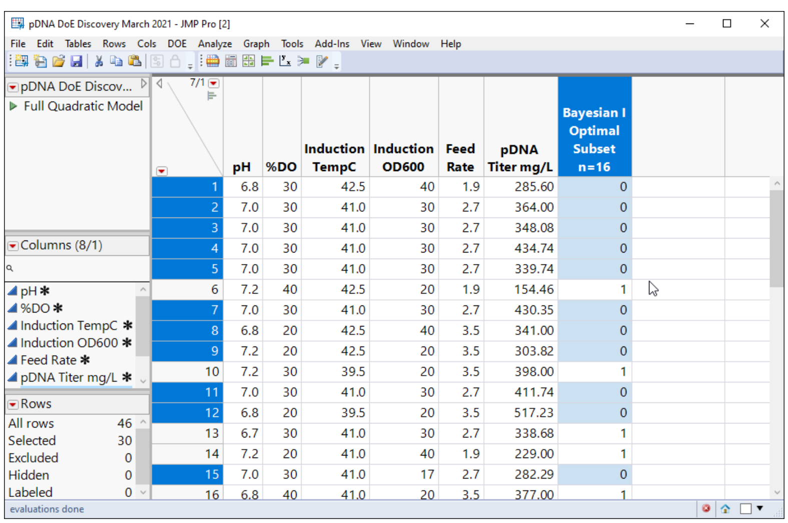
	pH	%DO	Induction TempC	Induction OD600	Feed Rate	pDNA Titer mg/L	Bayesian I Optimal Subset n=16
1	6.8	30	42.5	40	1.9	285.60	0
2	7.0	30	41.0	30	2.7	364.00	0
3	7.0	30	41.0	30	2.7	348.08	0
4	7.0	30	41.0	30	2.7	434.74	0
5	7.0	30	41.0	30	2.7	339.74	0
6	7.2	40	42.5	20	1.9	154.46	
7	7.0	30	41.0	30	2.7	430.35	
8	6.8	20	42.5	40	3.5	341.00	
9	7.2	20	42.5	20	3.5	303.82	
10	7.2	30	39.5	20	3.5	398.00	
11	7.0	30	41.0	30	2.7	411.74	
12	6.8	20	39.5	20	3.5	517.23	
13	6.7	30	41.0	30	2.7	338.68	1
14	7.2	20	41.0	40	1.9	229.00	1
15	7.0	30	41.0	17	2.7	282.29	0
16	6.8	40	41.0	20	3.5	377.00	1

jmp Statistical Discovery.™ From SAS.

sas THE POWER TO KNOW.

57

THE PLASMID DNA (pDNA) EXPERMENT



	pH	%DO	Induction TempC	Induction OD600	Feed Rate	pDNA Titer mg/L	Bayesian I Optimal Subset n=16
1	6.8	30	42.5	40	1.9	285.60	0
2	7.0	30	41.0	30	2.7	364.00	0
3	7.0	30	41.0	30	2.7	348.08	0
4	7.0	30	41.0	30	2.7	434.74	0
5	7.0	30	41.0	30	2.7	339.74	0
6	7.2	40	42.5	20	1.9	154.46	1
7	7.0	30	41.0	30	2.7	430.35	0
8	6.8	20	42.5	40	3.5	341.00	0
9	7.2	20	42.5	20	3.5	303.82	0
10	7.2	30	39.5	20	3.5	398.00	1
11	7.0	30	41.0	30	2.7	411.74	0
12	6.8	20	39.5	20	3.5	517.23	0
13	6.7	30	41.0	30	2.7	338.68	1
14	7.2	20	41.0	40	1.9	229.00	1
15	7.0	30	41.0	17	2.7	282.29	0
16	6.8	40	41.0	20	3.5	377.00	1

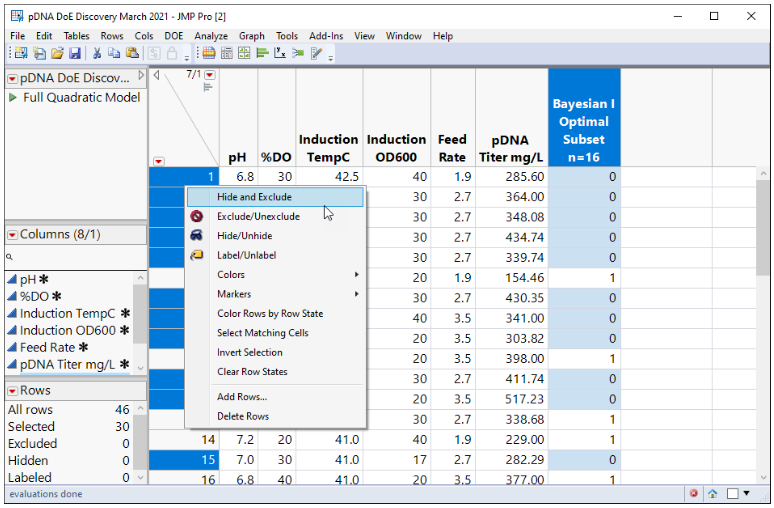
jmp Statistical Discovery.™ From SAS.

Copyright © 2018, SAS Institute, Inc. All rights reserved.

sas THE POWER TO KNOW.

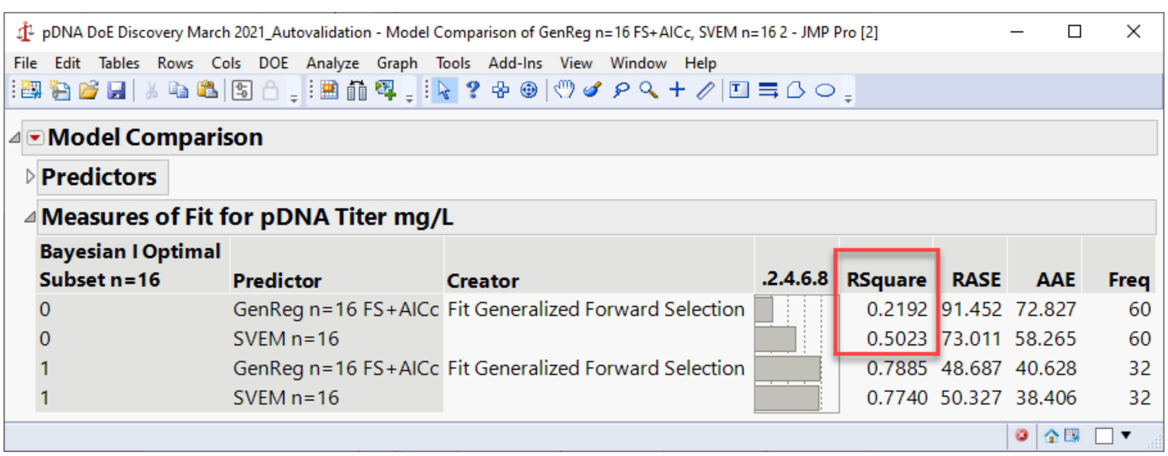
58

THE PLASMID DNA (pDNA) EXPERMENT



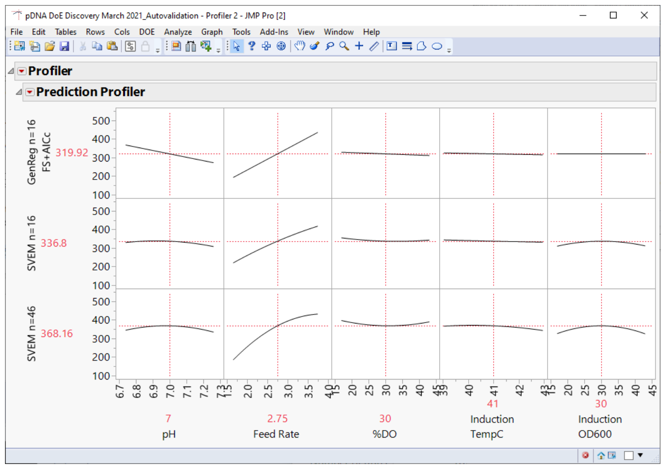
59

THE PLASMID DNA (pDNA) EXPERMENT



60

THE PLASMID DNA (pDNA) EXPERMENT



61

OBSERVATIONS AND FURTHER QUESTIONS

- Diminishes the role of p-value based thinking.
- Establish best practice for uncertainty analysis of predictions
- What are the best families of base models?
- What kinds of designs do we use?
- What is the role of screening experiments?
- What are power and sample size considerations?
- What about blocking and split-plot experiments?

62

Philip Ramsey

Case-Study & SVEM Product Demonstration

63

Philip Ramsey

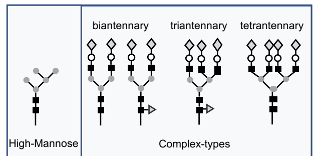
Characterizing a Glycoprofiling Analytical Method for Therapeutic Proteins using SVEM and Bayesian I-optimal Designs

64

Glycoproteins are the largest group of biologically-derived drugs.

ICH Q6B guideline requires extensive physiochemical characterization of biopharmaceuticals including inherent structural heterogeneity due to glycosylation (post-translationally modified) and lot-to-lot consistency is required.

- Carbohydrate content.
- Carbohydrate chain structure.
- Oligosaccharide pattern (antennary profile).
- Glycosylation site.

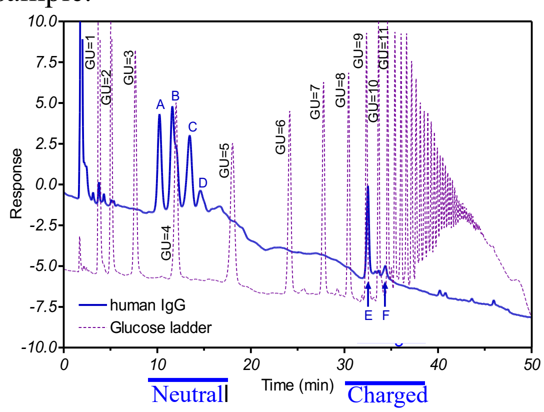


Common types of *N*-glycans. UPS General Chapter <1084>.

Currently, there is a lack of a universally accepted analysis technique for glycosylation characterization.

The goal is to develop a robust, cost effective characterization method.

Two independent experiments was used to characterize & optimize an HPAE-PAD[1] method. The approach uses a glucose ladder (GU) as a reference to identify glycoform peaks from an actual human antibody sample.



Glycoform glycan	G-Unit
A	3.59
B	3.89
C	4.23
D	4.42
E	9.17
F	10.8

[1]High Performance Anion Exchange Chromatography with Pulsed Amperometic Detection

The first experiment consisted of a 16 run, 3-level design.

The second experiment consisted of 28 run, 3-level design.

Since the two experiments are independent, one set can be used as a training set and the second set as a validation set.

For purposes of today's discussion, both designs were combined into a single 44 run design.

The candidate selector in Custom Design (JMP 16) was used to create Bayesian I-optimal designs from the 44 combined runs:

- **Design #1 n = 10 runs**, Full Quadratic Model specified
- **Design #2 n = 13 runs**, Full Quadratic Model specified
- **Design #3 n = 16 runs**, Partial Cubic Model specified

The runs not selected are used as a validation set for predictive models.

Traditionally in statistics the full quadratic model (FQM) is used to build models to optimize physical processes. The FQM has the following mathematical form.

$$Y = \beta_0 + \sum_{i=1}^{K} \beta_i X_i + \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \beta_{ij} X_i X_j + \sum_{i=1}^{K} \beta_{ii} X_i^2$$

Although popular for optimization the FQM is often not sufficient to model the complex kinetics over the experimental region.

Cornell and Montgomery (1998) suggested augmenting the FQM with partial cubic (PC) terms that allow more flexibility to fit the complex kinetics in physical systems. The form of the PC model is:

$$Y = \beta_0 + \sum_{i=1}^{K} \beta_i X_i + \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \beta_{ij} X_i X_j + \sum_{i=1}^{K} \beta_{ii} X_i^2 + \sum_{i=1}^{K} \sum_{j \neq i}^{K} \beta_{iij} X_i^2 X_j + \sum_{i=1}^{K} \sum_{j \neq i}^{K} \beta_{ijj} X_i X_j^2 + \sum_{i=1}^{K} \sum_{j \neq i}^{K} \beta_{iijj} X_i^2 X_j^2$$

Five factors were selected to manipulate in the experiments.

Factor (level)	-1	0	1
Initial %NaOAc (% A)	0	10	20
Initial %NaOH (% B)	30	40	50
Gradient_01-12 (mM NaOAc /min)	0.415	1.25	2.085
Gradient_12-24 (mM NaOAc /min)	1.25	2.085	2.915
Gradient_24-42 (mM NaOAc /min)	4.72	5.555	6.39

Two responses were chosen to optimize in the experiment; a total of 28 responses exist.

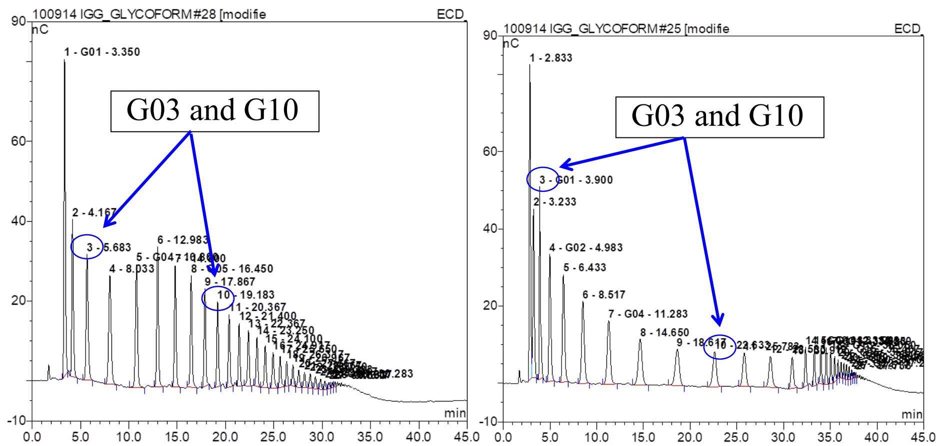
Retention Time for glycan 3 (RT_G03) was most important as it anchors the position of the glucose ladder used to identify specific glycan peaks.

The second response is peak resolution for charged glycan G10 (Resol_G10).

Glycan G10 elutes late with a number of other charged glycans.

Response	Description	Optimization
RT_G03	Retention Time	Target ~ 8.5 min
RT_G10	Retention Time	None

Below are pictures of typical chromatograms from two different runs. Notice the shapes change markedly with experimental conditions.



71

71

The original experimental designs and the experimental work was performed by Dr. Eliza Yeung of Cytovance Biologics, OKC.

Dr. Yeung is the Director of Process Characterization for Cytovance.

We thank Dr. Yeung and Cytovance for their willingness to share the experimental results.

Dr. Yeung is an outstanding scientist and a JMP user.

Eliza Yeung, Ph.D.
Director of Process Characterization
Cytovance Biologics, Inc
800 Research Parkway, Ste 200
Oklahoma City, OK, USA
eyeung@cytovance.com

72

72

A limitation of the PC model for traditional statistical modeling methods is the number of potential PC model terms.

For K factors the full PC model has $2K^2+1$ terms including the intercept; e.g., for K = 5 there exist 51 potential model terms.

Using **SVEM** it is actually possible to fit full PC models with relatively small number of design points.

We illustrate **SVEM** with the case study.

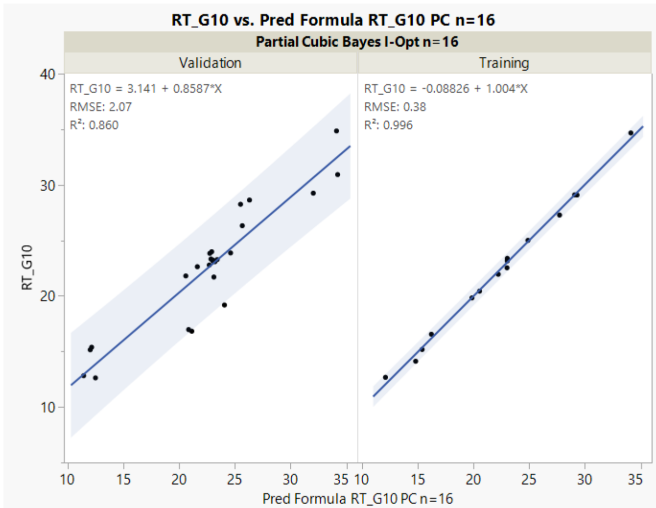
All of the SVEM modeling was performed using the **SVEM addin available from Predictum Inc**.

If interested in the addin send email to Wayne Levin at Predictum: levin@predictum.com.

Below is summary table of the SVEM modeling results. All SVEM runs used Best Subsets with Max Model Size = 5 and Nboot = 1,000 FWB runs for each model.

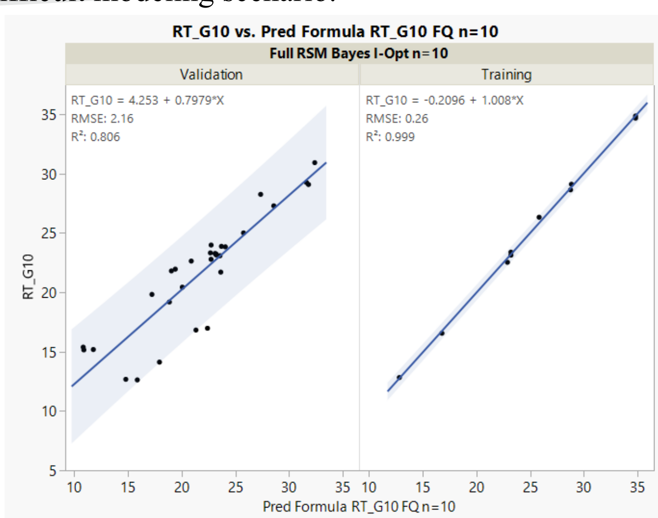
Bayesian I-optimal Design	Response	Full Model (No. Predictors)	RASE Training	RASE Validation	R^2 Validation
N = 16	RT_G03	PC (40)	0.190	**0.387**	**0.98**
N = 16	RT_G10	PC (40)	0.353	**2.148**	**0.84**
N = 13	RT_G03	PC (40)	0.065	**0.480**	**0.97**
N = 13	RT_G10	PC (40)	0.172	**2.671**	**0.75**
N = 10	RT_G03	FQ (20)	0.225	**0.730**	**0.94**
N = 10	RT_G10	FQ (20)	0.240	**2.354**	**0.75**

Below are typical actual by predicted plots for RT_G10 and N = 16, the hardest modeling scenario.



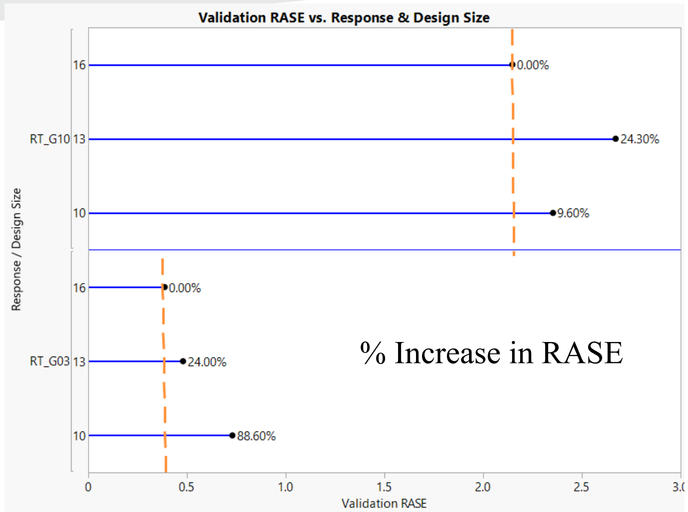
75

Below are typical actual by predicted plots for RT_G10 and N = 10, the more difficult modeling scenario.



76

Validation RASE values vs Design Size for RT_G03 and RT_G10.



SVEM provides the ability to fit supersaturated models (p > n) that are often needed for prediction in complex physical systems.

With SVEM we combine methods from machine learning with Design of Experiments.

Modeling with SVEM allows one to use highly efficient experimental designs speeding up the pace of innovation and reducing overall cost of experimentation.

Bayeian I-optimal designs, available in Custom Design, have the potential to substantially decrease the sizes of experimental designs when use in combination with SVEM.

With a 10 run Bayesian I-optimal design we were able to use SVEM to fit a 20 predictor model that successfully predicted responses on a true validation dataset.

SVEM Add-in

DSD pDNA DoE.jmp

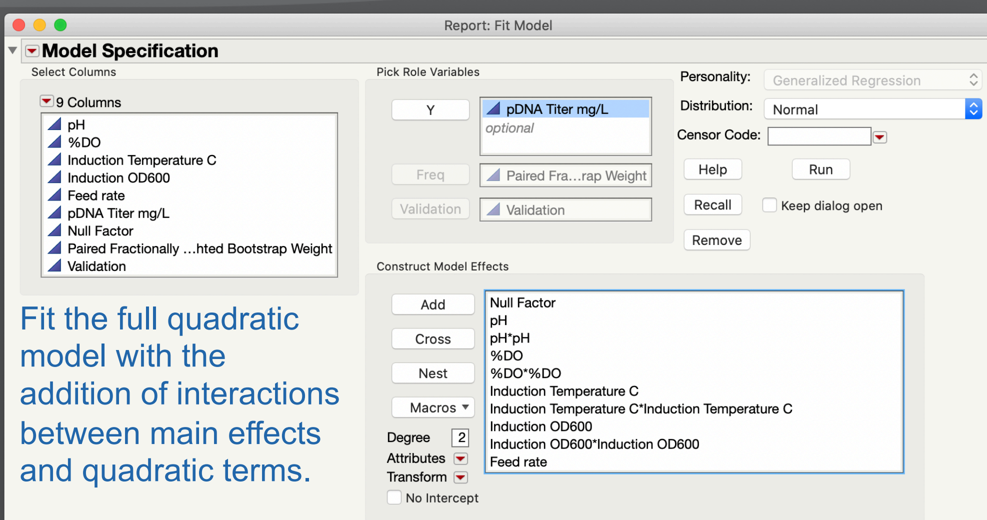
	pH	%DO	Induction Temperature C	Induction OD600	Feed rate	pDNA Titer mg/L
1	7.0	40	42.5	20	1.9	156.20
2	7.0	20	39.5	40	3.5	318.45
3	7.2	30	39.5	20	3.5	398.00
4	6.8	30	42.5	40	1.9	285.60
5	7.2	20	41.0	40	1.9	229.00
6	6.8	40	41.0	20	3.5	377.00
7	7.2	20	42.5	30	3.5	290.00
8	6.8	40	39.5	30	1.9	123.00
9	7.2	40	42.5	40	2.7	299.00
10	6.8	20	39.5	20	2.7	428.00
11	7.0	30	41.0	30	2.7	327.80
12	7.0	30	41.0	30	2.7	339.74
13	7.0	30	41.0	30	2.7	387.35
14	7.0	30	41.0	30	2.7	393.97
15	7.0	30	41.0	30	2.7	348.08

Predictum

- Capability Analysis
- Auto Word Output
- Developer Panel
- Self-Validating Ensemble Modeling
- Demo Central

79

Fit Full Quadratic Model +



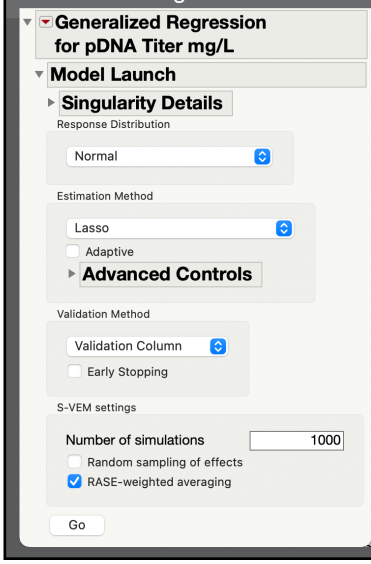
Fit the full quadratic model with the addition of interactions between main effects and quadratic terms.

80

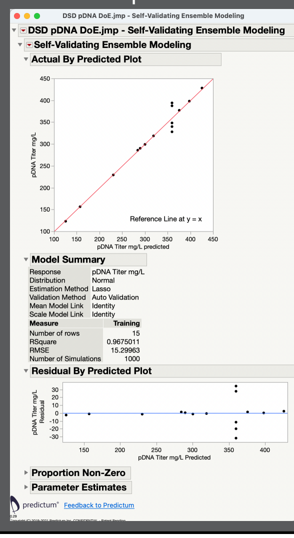
Saving the Prediction Formula

predictum®

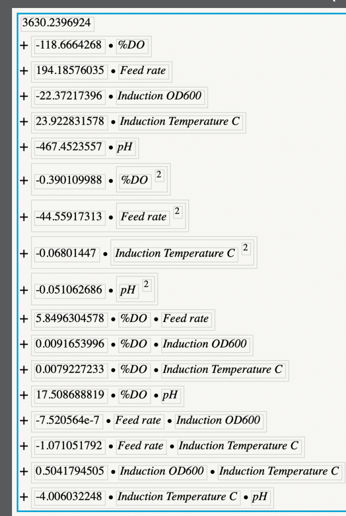
S-VEM Settings at bottom



S-VEM Output



Saved Prediction Formula



81

ANNOUNCEMENT

ON-DEMAND COURSE
MIXTURE DESIGN EXPERIMENTS

82

ANNOUNCEMENT

SVEM ADD-IN

83

Next Steps

predictum®

- 2 ways to try this yourself
  1. Try the Product
     - JMP Pro v14.3 or 15.x
     - JMP Pro trials are available via this link
  2. Contact Predictum
     - Working with your historical DOEs, we can run a proof-of-concept, comparative analyses for you, on a limited basis
     - Review the analysis with you

Contact Wayne:
levin@predictum.com

84

84

predictum®

Self-Validated, Ensemble Models (S-VEM)

Remarkable new method generating more insightful and accurate models from small sets of data

Request SVEM Evaluation: levin@predictum.com

Click here to request JMP Pro Evaluation

85

predictum®

CONTACT US
416.398.8900
info@predictum.com

USA HEADQUARTERS
2375 E Camelback Rd, Suite 600
Phoenix, AZ 85016

CANADA HEADQUARTERS
150 Eglinton Avenue East, Suite 805
Toronto, Ontario M4P 1E8

86