

Lessons from Definitive Screening Designs

Stu Janis
 3M Lean Six Sigma Operations
 225-4S-18
 St. Paul, MN 55144

Definitive Screening Designs (DSDs), introduced by Jones and Nachtsheim (2011), are an exciting recent development in the field of DOE. Traditionally, screening experiments generally involved highly fractionated 2-level designs where main effects often were completely confounded with 2-factor interactions and 2-factor interactions were confounded with each other. In addition, screening designs were only good for screening; model building required additional data and the associated time and resources involved in gathering it.

DSDs have many advantages including:

1. The minimum number of runs for k factors is only $2k + 1$ if all factors continuous and $2k + 2$ if a categorical factor is involved (but a few extra runs are recommended).
2. Main effects are completely independent of two-factor interactions.
3. Two-factor interactions are not completely confounded with other two-factor interactions, although they may be correlated.
4. All quadratic effects estimable, orthogonal to main effects, and not completely confounded (though correlated) with interaction effects.
5. With six through (at least) 12 factors, all possible full quadratic models involving three or fewer factors are estimable with very high statistical efficiency.

At their simplest, DSDs consist of pairs of observations, where one factor is held at its center level and the remaining factors are at various combinations of -1 and +1 levels. The two points within a pair have opposite values of -1 and +1 for the factors not held at the 0 level, giving the designs foldover properties (see Figure 1). DSDs that include qualitative factors with only two levels were introduced in Jones and Nachtsheim (2013).

m = 6	m = 7	m = 8
1 0+-----	1 0+---+-	1 0-+++++
2 0-++++	2 0----+-	2 0+-----
3 +0-+-	3 -0+---	3 -0-++++-
4 -0+--	4 +0+---	4 +0+-----
5 --0+--	5 +-0++++	5 --0+---+
6 ++0-+-	6 -+0----	6 ++0-+-+
7 -++0+-	7 +--0+--	7 +-+0+---
8 +--0-+	8 -++0-+-	8 --+0-+-+
9 +--0-0-	9 ---+0--	9 ---+0-+-
10 -+-+0+	10 +-+0-+-	10 +-+0+--
11 +++++0-	11 -+++0+	11 +---+0++
12 -----0	12 +---0-	12 -+++0--
13 000000	13 +++++0-	13 -+-+0+
	14 -----0	14 +---+0-
	15 0000000	15 ++++++0
		16 -----+0
		17 00000000

Figure 1 – Examples of Definitive Screening Designs

This paper will discuss my experience with the design and analysis of two DSDs in an industrial environment.

DSD 1 – Improving Manufacturing Efficiency

The first DSD was conducted in a manufacturing setting in 2015. The experiment consisted of six factors and three responses. One factor was categorical and the experiment used the minimum number of runs of $2k+2$, or 14^1 . The goal was to understand what factors affect process outputs and seek opportunities to improve efficiencies.

With 14 runs and 6 factors, of course, it's not possible to fit a full 2^{nd} order model, so Jones and Nachtsheim's recommended approach at the time was stepwise regression. However, there are a number of approaches to stepwise regression and the question was, which approach makes the most sense. The two main elements I considered in stepwise regression were:

- Stopping criterion – p-value vs. AICc
- Rules regarding hierarchy, sometimes known as heredity – I looked at two of JMP's options: "Combine" and "No Rules"

A p-value stopping criterion is common and easy to understand, but Jones and Nachtsheim recommend AICc based on simulations suggesting that it does a better job of finding active effects for DSDs. Since this was my first DSD I thought I'd try both. It's not clear which approach to heredity is preferred, so it seemed like a good opportunity to try both approaches there, too. For the p-value criterion, I used forward stepwise regression and that is what is shown in the tables below. While writing this paper I also looked at mixed-direction stepwise and found that sometimes it generated similar models and other times the models were quite different.

A "Combine" approach to heredity means that if a 2^{nd} order term enters the model, its parents are included as well. JMP uses a combined F-test of the 2^{nd} order term and its parents to decide whether to include it in the model. The "No Rules" approach is to treat each term individually and not require the parents to be part of a model that includes 2^{nd} order terms.

The experiment was designed and analyzed with JMP 12. The design appears in Table 1 and a summary of the analyses for Y1 appear in Table 2

Run Order	Block	A	B	C	D	E	F	Y1	Y2	Y3
1	1	B	0	0	0	0	0	588	1.5804	14.47
2	1	B	-1	-1	-1	1	0	591	1.926	13.325
3	1	B	1	-1	0	-1	1	581	1.7196	16.17
4	1	B	0	1	1	1	1	571	1.494	12.44
5	1	B	-1	0	1	-1	-1	575	1.4168	11.14
6	1	B	1	1	-1	0	-1	572	1.446	13.2
7	1	B	1	-1	1	1	-1	570	1.472	11.41
8	1	A	-1	1	0	1	-1	551	1.2572	10.55
9	1	A	0	0	0	0	0	572	1.5488	15.56
10	1	A	-1	1	-1	-1	1	569	1.724	16.445
11	1	A	1	0	-1	1	1	580	1.6268	14.7
12	1	A	0	-1	-1	-1	-1	567	1.52	11.42
13	1	A	-1	-1	1	0	1	584	1.7852	14.29
14	1	A	1	1	1	-1	0	577	1.3708	15.415

Table 1 – Definitive Screening Design 1

¹ Categorical factors require an additional data point, making their minimum number of runs $2k+2$.

Everything went according to plan with the design except for the run order involving Factor A. The experimenters' intent was to use a completely randomized design, but upon beginning the experiment they discovered the material designated for Level A hadn't been prepared properly. In the interest of time they ran all the conditions involving level B while the material for level A was remade.

Stopping Criterion	p to enter and remove	Rules	A	B	C	D	E	F	AC	AF	BC	DF	FF	Cp	R ² Adj	AICc	Terms
p-value	0.05	Combine												20.6	0	108.3	0
p-value	0.05	No Rules												20.6	0	108.3	0
p-value	0.15	Combine	X	X				X	X	X			X	3.6	.796	117.1	6
p-value	0.15	No Rules	X	X				X			X	X	X	2.0	.883	109.3	6
AICc	N/A	Combine						X					X	11.2	.303	108.3	2
AICc	N/A	No Rules	X	X				X			X	X		1.4	.896	104.3	5

Table 2 – Comparison of models for Y1 generated with various approaches to analysis

The first thing to note about Table 2 is that a p-value stopping criterion of 0.05 brings nothing into the model – not a promising start for my first analysis of my first DSD! Relaxing the criterion to 0.15 yielded models with six terms each, though the interaction terms were different for the two heredity rules. Finally, AICc yielded a model with only two terms using combined heredity rules, and using No rules generated a five term model that was similar to the p-value of 0.15 with no rules.

In every model where there was statistical significance, factor F is involved in both a main effect and at least one 2nd order term, so I felt confident that F was an active effect. The rest of the terms weren't as clear, given the mixed messages from the models. One guideline in stepwise regression analysis of DSDs is that the number of terms in the model shouldn't exceed n/2, which in this case is 7. None of the models exceeded 7 terms but some came close. The models with more terms had much higher R²Adj values, but the Cp values of those models were lower than one might expect. On the other hand, the models with fewer terms had Cp values higher than one would expect.

An investigation of model term correlations between the models with p=0.15 is shown in Figure 2. There is some correlation between AC and DF, but very little correlation between AF and BC, meaning that given the restrictions of the Combine rule, AC and DF could be serving similar functions in their respective models.

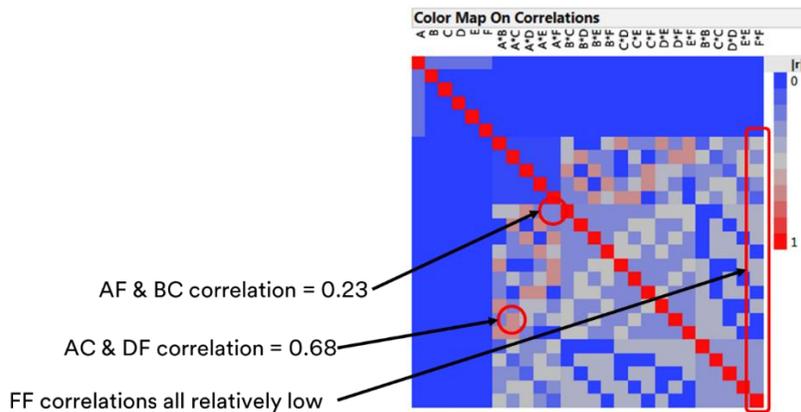


Figure 2 – Color Map of Correlations for 14-run DSD with one categorical factor

Analyses of Y2 and Y3 yielded a variety of models as well, as shown in Table 3 and Table 4. As with Y1, the AICc model with the Combine rule generally makes the simplest models with the fewest number of terms, but at the expense of lower R^2_{Adj} and higher Cp values. Using no rules resulted in models with lots of terms (including a few models where the number of terms exceeded $n/2$), high R^2_{Adj} values, low Cp values, and questions about whether the models were overfitting the data. In all three responses the p-value and AICc approach yielded identical or nearly identical models when no heredity rules were used.

Stopping Criterion	p to enter and remove	Rules	A	B	C	D	E	F	AD	AF	CE	CF	DF	Cp	R ² Adj	AICc	Terms
p-value	0.05	Combine	X	X	X	X		X	X	X		X		5.9	.984	12.7	8
p-value	0.05	None	X	X	X	X		X			X		X	3.9	.987	-17.9	7
AICc	N/A	Combine			X			X				X		94.5	.578	-7.7	3
AICc	N/A	None	X	X	X	X		X			X		X	3.9	.987	-17.9	7

Table 3 – Comparison of models for Y2

Stopping Criterion	p to enter and remove	Rules	A	B	C	D	E	F	CD	FF	Cp	R ² Adj	AICc	Terms
p-value	0.05	Combine						X			40.5	.485	55.7	1
p-value	0.05	None		X		X	X	X	X	X	2.6	.976	50.6	6
AICc	N/A	Combine						X		X	28.8	.592	55.3	2
AICc	N/A	None		X		X	X	X	X	X	2.6	.976	50.6	6

Table 4 – Comparison of models for Y3

One other note on the models relates to the AICc Combine for Y2. The model included C, F, and C*F, however the C*F p-value was 0.995. JMP 12 only allows forward selection for the AICc criterion, and manually removing C*F dropped AICc from -7.7 to -12.7 (Table 5).

SSE	DFE	RMSE	RSquare	RSquare Adj	Cp	p	AICc	BIC
0.1355312	10	0.1164179	0.6746	0.5770	94.538242	4	-7.69627	-12.001

Lock	Entered	Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Intercept	1.5634	1	0	0.000	1
<input type="checkbox"/>	<input type="checkbox"/>	A[A-B]	0	1	0.004487	0.308	0.59235
<input type="checkbox"/>	<input type="checkbox"/>	B	0	1	0.022468	1.788	0.21393
<input type="checkbox"/>	<input checked="" type="checkbox"/>	C	-0.11308	2	0.127871	4.717	0.03607
<input type="checkbox"/>	<input type="checkbox"/>	D	0	1	0.049562	5.189	0.04874
<input type="checkbox"/>	<input type="checkbox"/>	E	0	1	6.15e-5	0.004	0.95043
<input type="checkbox"/>	<input checked="" type="checkbox"/>	F	0.12376	2	0.153166	5.651	0.0228
<input type="checkbox"/>	<input type="checkbox"/>	A[A-B]*B	0	3	0.031828	0.716	0.573
<input type="checkbox"/>	<input type="checkbox"/>	B*B	0	2	0.027685	1.027	0.40092
<input type="checkbox"/>	<input type="checkbox"/>	A[A-B]*C	0	2	0.005467	0.168	0.84815
<input type="checkbox"/>	<input type="checkbox"/>	B*C	0	2	0.039337	1.636	0.25377
<input type="checkbox"/>	<input type="checkbox"/>	C*C	0	1	0.002285	0.154	0.70356
<input type="checkbox"/>	<input type="checkbox"/>	A[A-B]*D	0	3	0.092192	4.963	0.03729
<input type="checkbox"/>	<input type="checkbox"/>	B*D	0	3	0.079877	3.349	0.08511
<input type="checkbox"/>	<input type="checkbox"/>	C*D	0	2	0.051384	2.443	0.1486
<input type="checkbox"/>	<input type="checkbox"/>	D*D	0	2	0.058742	3.060	0.10305
<input type="checkbox"/>	<input type="checkbox"/>	A[A-B]*E	0	3	0.029095	0.638	0.61422
<input type="checkbox"/>	<input type="checkbox"/>	B*E	0	3	0.022733	0.470	0.71245
<input type="checkbox"/>	<input type="checkbox"/>	C*E	0	2	0.033595	1.318	0.32
<input type="checkbox"/>	<input type="checkbox"/>	D*E	0	3	0.052501	1.475	0.30164
<input type="checkbox"/>	<input type="checkbox"/>	E*E	0	2	0.004946	0.152	0.86183
<input type="checkbox"/>	<input type="checkbox"/>	A[A-B]*F	0	2	0.017329	0.586	0.57856
<input type="checkbox"/>	<input type="checkbox"/>	B*F	0	2	0.044476	1.951	0.20371
<input type="checkbox"/>	<input checked="" type="checkbox"/>	C*F	0.00025	1	5e-7	0.000	0.99527
<input type="checkbox"/>	<input type="checkbox"/>	D*F	0	2	0.049567	2.306	0.16185
<input type="checkbox"/>	<input type="checkbox"/>	E*F	0	2	0.001846	0.055	0.94661
<input type="checkbox"/>	<input type="checkbox"/>	F*F	0	1	0.010403	0.748	0.4095

SSE	DFE	RMSE	RSquare	RSquare Adj	Cp	p	AICc	BIC
0.1355317	11	0.1110003	0.6746	0.6155	92.538613	3	-12.7518	-14.64

Lock	Entered	Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Intercept	1.5634	1	0	0.000	1
<input type="checkbox"/>	<input type="checkbox"/>	A[A-B]	0	1	0.004487	0.342	0.57143
<input type="checkbox"/>	<input type="checkbox"/>	B	0	1	0.022468	1.987	0.18897
<input type="checkbox"/>	<input checked="" type="checkbox"/>	C	-0.11308	1	0.127871	10.378	0.00814
<input type="checkbox"/>	<input type="checkbox"/>	D	0	1	0.049562	5.765	0.03725
<input type="checkbox"/>	<input type="checkbox"/>	E	0	1	6.15e-5	0.005	0.94761
<input type="checkbox"/>	<input checked="" type="checkbox"/>	F	0.12376	1	0.153165	12.431	0.00475
<input type="checkbox"/>	<input type="checkbox"/>	A[A-B]*B	0	3	0.031813	0.818	0.51937
<input type="checkbox"/>	<input type="checkbox"/>	B*B	0	2	0.026733	1.106	0.37206
<input type="checkbox"/>	<input type="checkbox"/>	A[A-B]*C	0	2	0.005467	0.189	0.83086
<input type="checkbox"/>	<input type="checkbox"/>	B*C	0	2	0.038238	1.769	0.225
<input type="checkbox"/>	<input type="checkbox"/>	C*C	0	1	0.002285	0.171	0.68753
<input type="checkbox"/>	<input type="checkbox"/>	A[A-B]*D	0	3	0.077661	3.579	0.06629
<input type="checkbox"/>	<input type="checkbox"/>	B*D	0	3	0.07797	3.612	0.06498
<input type="checkbox"/>	<input type="checkbox"/>	C*D	0	2	0.051255	2.373	0.11789
<input type="checkbox"/>	<input type="checkbox"/>	D*D	0	2	0.057187	2.785	0.08489
<input type="checkbox"/>	<input type="checkbox"/>	A[A-B]*E	0	3	0.017813	0.404	0.75457
<input type="checkbox"/>	<input type="checkbox"/>	B*E	0	3	0.022674	0.536	0.67072
<input type="checkbox"/>	<input type="checkbox"/>	C*E	0	2	0.031562	1.366	0.30332
<input type="checkbox"/>	<input type="checkbox"/>	D*E	0	3	0.051749	1.647	0.25432
<input type="checkbox"/>	<input type="checkbox"/>	E*E	0	2	0.004054	0.139	0.87228
<input type="checkbox"/>	<input type="checkbox"/>	A[A-B]*F	0	2	0.017329	0.660	0.54031
<input type="checkbox"/>	<input type="checkbox"/>	B*F	0	2	0.043153	2.102	0.17819
<input type="checkbox"/>	<input type="checkbox"/>	C*F	0	1	5e-7	0.000	0.99527
<input type="checkbox"/>	<input type="checkbox"/>	D*F	0	2	0.049566	2.595	0.12891
<input type="checkbox"/>	<input type="checkbox"/>	E*F	0	2	0.00172	0.058	0.94413
<input type="checkbox"/>	<input type="checkbox"/>	F*F	0	1	0.010403	0.831	0.38334

Table 5 – Comparison of models for Y2 with and without C*F (p = 0.995)

With so many models ranging from very simple to fairly complex, it was time to bring in the subject matter experts to address the important question, Which models make sense? They generally agreed that the simpler models aligned more with their understanding of how the process works.

A couple of follow-up options would have been to augment the design to a) untangle some of the interactions that are highly correlated and/or b) collect data at points in the design space with the greatest disagreement between the models. In the end, the team opted to do neither because they were able to implement a reduction in Factor F, leading to a cycle time improvement and opening the door to further optimization of similar lines. Those improvements would delay or even eliminate the need to purchase and install a new manufacturing line.

Lessons from DSD 1

Several useful lessons came from my first DSD. They include:

- Different model selection techniques can yield widely varying models. When using JMP 12 analysis, keep in mind that the number of model terms should be no larger than $n/2$, and verify the models with subject matter experts. In addition, the AICc criterion generally results in simpler models than a p-value criterion, and combined rules tends to yield simpler models than the No Rules option.
- Add some extra runs, if possible, to give the design more power.
- Pay attention to the data sheet. For unknown reasons, the experimenter listed the results for one response in reverse run order. I failed to notice it in my initial analysis, and from a purely statistical perspective, it generated a fine model with some statistically significant effects and good-looking residual plots. Of course, it was meaningless, but it shows the dangers of blindly analyzing a small data set with a relatively large number of factors and no context.

DSD 2 – Test Method Ruggedness

The second DSD was conducted in a product development lab and was used to evaluate non-product-related factors that could affect test results. The product in question is attached to a coated substrate, and adhesion is evaluated by performing a peel test. When two product samples have different adhesion results, we'd prefer that it be because the samples are actually different as opposed to the difference being caused by variation in the operators' substrate preparation techniques.

The experimenters considered a number of opportunities for variation in how the operators prepare the substrates. They also wanted to evaluate the effects on two different products. In particular they looked at:

- A. Time between steps – Time between preparation steps 1 and 2
- B. Time before peel – Time between preparation step 2 and the peel test
- C. Coating amount – Amount of coating used during substrate preparation
- D. Pressure – how much pressure was used when applying the coating
- E. Application rate – how quickly the coating was applied to the substrate
- F. Final preparation step – Yes or No
- G. Product type – A or B

In the interval between running DSD 1 and DSD 2, Jones and Nachtsheim became stronger advocates for using pairs of “fake factors” to generate additional power (Jones 2015). In JMP 12, this simply involves adding extra factors to the design so that additional points are generated, and then not including those fake factors in the analysis. In JMP 13, the dialog box contains a field for “extra runs,” which accomplishes the same thing.

Fake factors have some obvious advantages. The question in my mind was how many to add: two or four. Each fake factor adds two extra runs, so starting from a base of 20 runs, two fake factors would give a total of 24 runs, and four fake factors would require 28 runs. Would the additional four runs from fake factors 3 and 4 would add enough benefit to offset the cost of the extra runs? The correlation plots in Figure 3 show that the highest correlations among 2-factor interactions drop pretty dramatically with four fake factors, and since experimental runs were relatively inexpensive, we proceeded with the 28-run design.

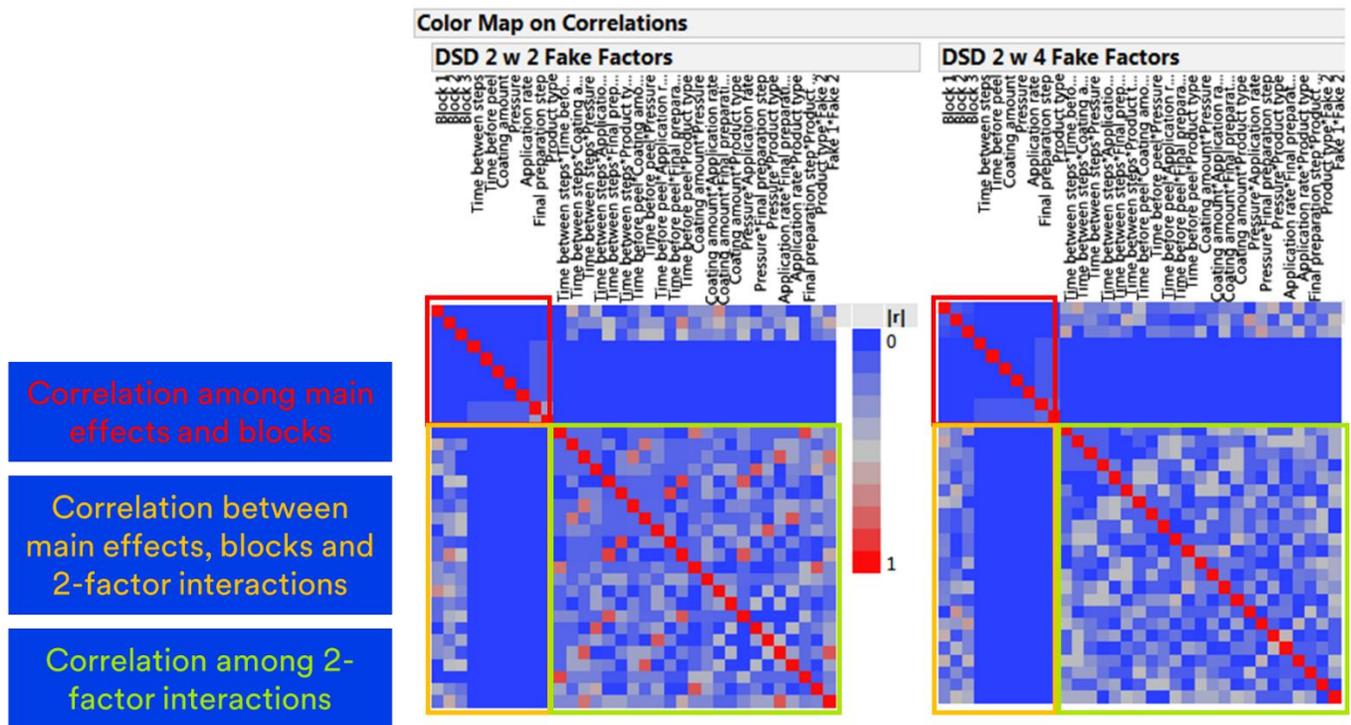


Figure 3 – Color map of correlations comparing 7-factor DSDs with four and eight extra runs

Jones and Nachtsheim also developed a new 2-stage approach to DSD analysis in the interval between my two DSDs that involves decomposing the response into one component to estimate main effects, and another to estimate 2nd order terms (Jones 2015). Stage 1 of the analysis fits a first order model to the main effect component; Stage 2 uses the 2nd order component to look for 2nd order terms only

involving the factors found significant in the main effects analysis, also known as strong heredity. JMP 13 incorporates this approach, and the analysis is shown in Table 7.

The screenshot shows the JMP Pro interface for a Definitive Screening Design 2. The table contains 28 rows and 9 columns. The columns are: Block, Time between steps, Time before peel, Coating amount, Pressure, Application rate, Final preparation step, and Product Type. The rows are numbered 1 through 28, with the last row (28) being the only one selected. The 'Block' column has values A, B, C, and D. The 'Time between steps' column has values 0, 1, and -1. The 'Time before peel' column has values -1, 0, and 1. The 'Coating amount' column has values -1, 0, and 1. The 'Pressure' column has values -1, 0, and 1. The 'Application rate' column has values -1, 0, and 1. The 'Final preparation step' column has values Y and N. The 'Product Type' column has values A and B.

	Block	Time between steps	Time before peel	Coating amount	Pressure	Application rate	Final preparation step	Product Type
1	A	1	-1	1	-1	0	Y	B
2	A	0	1	1	1	1	Y	A
3	A	-1	1	-1	1	0	N	A
4	A	-1	-1	-1	1	1	N	A
5	A	1	1	1	-1	-1	Y	B
6	A	0	-1	-1	-1	-1	N	B
7	A	0	0	0	0	0	N	B
8	A	0	0	0	0	0	Y	A
9	B	0	0	0	0	0	Y	A
10	B	-1	1	1	-1	1	N	B
11	B	-1	0	-1	1	-1	Y	B
12	B	1	1	1	1	-1	Y	A
13	B	1	0	1	-1	1	N	A
14	B	-1	-1	-1	-1	1	N	B
15	B	1	-1	-1	1	-1	Y	A
16	B	0	0	0	0	0	N	B
17	C	1	-1	-1	-1	1	Y	A
18	C	-1	1	1	1	-1	N	B
19	C	1	-1	1	1	1	N	A
20	C	-1	1	-1	-1	-1	Y	B
21	C	-1	1	0	-1	1	Y	A
22	C	1	-1	0	1	-1	N	B
23	D	1	1	-1	1	1	Y	B
24	D	1	1	-1	-1	-1	N	A
25	D	-1	-1	1	0	-1	Y	A
26	D	-1	-1	1	1	1	Y	B
27	D	-1	-1	1	-1	-1	N	A
28	D	1	1	-1	0	1	N	B

Table 6 – Definitive Screening Design 2

Stage 1 from this analysis shows Coating amount and Product type as having significant main effects – small for Coating amount and very large for Product type. Stage 2 brings in a relatively small interaction between the two factors and an even smaller quadratic term for Coating amount ($p = 0.45$). These can be seen graphically in Figure 4.

The significance of Product type wasn't surprising since the products had different constructions, and it also revealed a potential flaw in the planning of the experiment. Since the product constructions are different, it isn't out of the question to imagine a 2-factor interaction being important for one product and not for the other – in essence, a 3-factor interaction. A better approach may have been to treat this as two 6-factor experiments, one on each product type, with a separate analysis for each.² Ultimately,

² This approach was taken in a follow-up experiment with a different type of coating and it worked well.

though, this analysis shows that the test method is very insensitive to the factors varied in the experiment.

Stage 1 - Main Effect Estimates				
Term	Estimate	Std Error	t Ratio	Prob> t
Coating amount	0.4497	0.1976	2.2757	0.0420*
Product Type	6.8132	0.1752	38.896	<.0001*
Statistic	Value			
RMSE	0.9239			
DF	12			

Stage 2 - Even Order Effect Estimates				
Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	27.406	0.6439	42.559	<.0001*
Block[1]	-0.54	0.4895	-1.103	0.3021
Block[2]	-0.241	0.4798	-0.501	0.6296
Block[3]	-0.093	0.5318	-0.174	0.8662
Coating amount*Product Type	-0.719	0.3353	-2.144	0.0644
Coating amount*Coating amount	-0.443	0.7262	-0.61	0.5591
Statistic	Value			
RMSE	1.5118			
DF	8			

Combined Model Parameter Estimates				
Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	27.406	0.5087	53.874	<.0001*
Block[1]	-0.54	0.3867	-1.396	0.1780
Block[2]	-0.241	0.379	-0.635	0.5329
Block[3]	-0.093	0.4201	-0.22	0.8279
Coating amount	0.4497	0.2555	1.7604	0.0936
Product Type	6.8132	0.2264	30.089	<.0001*
Coating amount*Product Type	-0.719	0.2649	-2.714	0.0134*
Coating amount*Coating amount	-0.443	0.5737	-0.772	0.4494
Statistic	Value			
RMSE	1.1943			
DF	20			

Table 7 – Model generated by JMP 13 two-stage approach

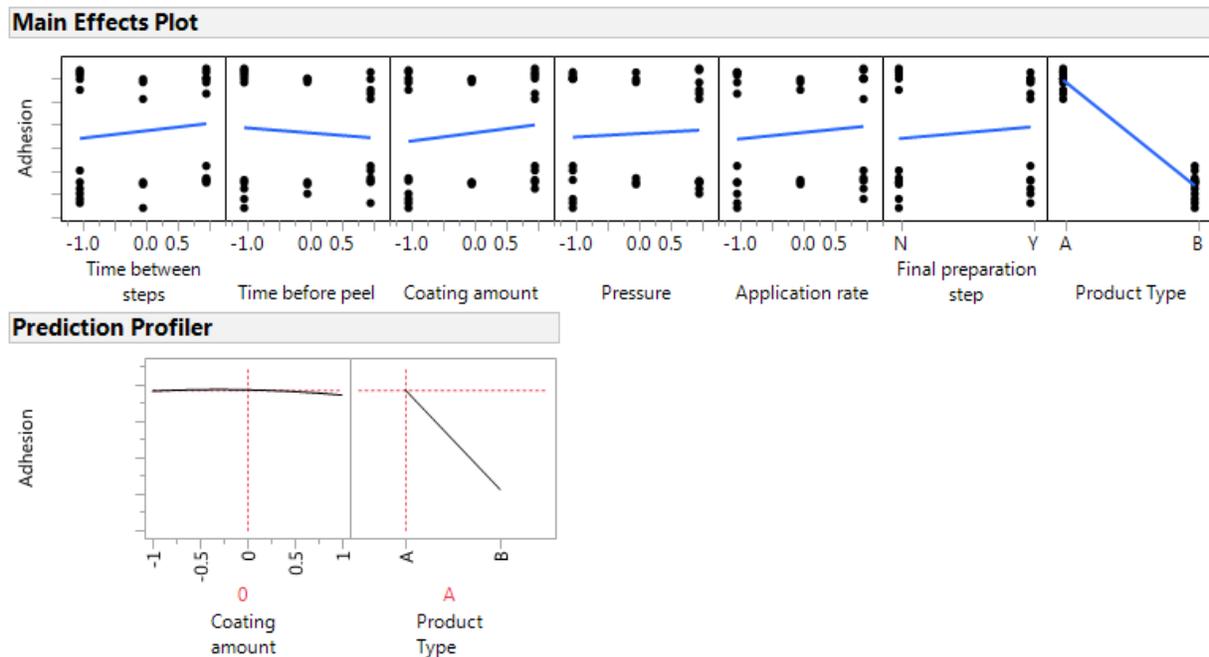


Figure 4 – Factor plots from two-stage model

One issue with the 2-stage analysis approach is that it's quite conservative. Experience shows that it is not unusual to find important 2-factor interactions where only one main effect is statistically significant (i.e., weak heredity), and a strong heredity approach will not detect them. Thus, I tried using the JMP 12

analysis approach of stepwise regression with AICc stopping criteria and the Combine rule. The analysis appears in Table 8 with interaction plots in Figure 5.

This analysis added two additional interaction terms, both involving Product type, thus providing further evidence that the products perform differently and may be affected differently by the other factors in the experiment. When discussing the results with subject matter experts, it wasn't clear which model made the most physical sense³. Some of the results appeared counter-intuitive, but were consistent with other experimental work. Fortunately, though, the magnitude of the effects and interactions (other than Product type) was very small from a practical perspective, and the test method ultimately appears rugged to the factors involved in substrate preparation.

Parameter Estimates				
Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	26.970099	0.170704	157.99	<.0001*
Time between steps	0.3398254	0.191326	1.78	0.0909
Time before peel	-0.026189	0.191326	-0.14	0.8925
Coating amount	0.4520981	0.191326	2.36	0.0284*
Product Type[A]	6.7869205	0.170704	39.76	<.0001*
Time between steps*Product Type[A]	-0.419554	0.191326	-2.19	0.0403*
Time before peel*Product Type[A]	-0.7659	0.191326	-4.00	0.0007*
Coating amount*Product Type[A]	-0.612736	0.191326	-3.20	0.0045*

Table 8 – Model from stepwise regression, AICc, combined rules

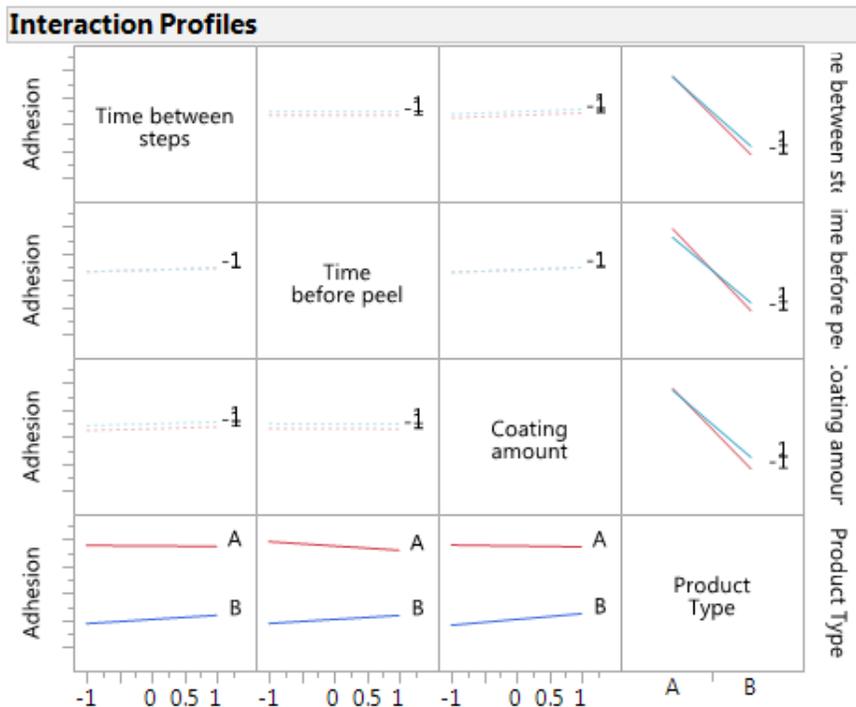


Figure 5 – Interaction plot from stepwise model

³ It's not shown here, but I also tried the "No rules" approach to stepwise regression with AICc. It added even more terms to the model.

Lessons from DSD 2

As with the first DSD, this experiment provided several useful lessons:

- Extra runs (i.e., fake factors) provide more power and better estimation of 2nd order terms. However, it's worth evaluating the benefits and costs of different numbers of extra runs. In this experiment eight extra runs provided substantial benefits compared to four extras. However, in a 6-factor experiment with all continuous factors, there appears to be no improvement in correlation among 2nd order terms when going from four extra runs to eight.
- The question of whether Product type is a factor or a response has come up in other experiments as well. In a screening design it makes more sense to treat the different types as separate responses.
- Good planning prior to executing the experiment helps immensely. Some of the factors weren't well-defined prior to the experiment, and the technicians who ran it spent a day developing good definitions of the -1, 0, and +1 levels for those factors. It allowed them to be consistent and made the results more meaningful.

Summary

Definitive Screening Designs are an exciting new development in the DOE world, with great potential for studying more factors with dramatically fewer runs. Even with “fake factors” the designs are considerably smaller than resolution V classical designs, and provide better estimates of 2nd order terms than resolution III or IV designs.

Depending on one's approach to the analysis, it is possible to generate many different models from the same data, ranging from conservative (2-stage approach) to generous (stepwise with AICc, combined rules) to perhaps overly generous (stepwise with AICc, no rules; p-value based stepwise). Ultimately it is best to review potential models with subject matter experts to help decide which is the most meaningful.

References

Jones, B. and Nachtsheim, C. J. (2011). “A Class of Three-Level Designs for Definitive Screening in the Presence of Second-Order Effects”. *Journal of Quality Technology* 43, pp. 1–14.

Jones, B. and Nachtsheim, C. J. (2013). “Definitive Screening Designs with Added Two-Level Categorical Factors”. *Journal of Quality Technology* 45, pp. 121–129.

Jones, B (2015). “A New Analytical Approach for Definitive Screening Designs”. JMP Discovery Summit, <https://community.jmp.com/docs/DOC-7738>