New DOE Features in JMP 11

Bradley Jones
Principal Research Fellow
JMP Division of SAS

Statistical Discovery. From SAS®

Copyright © 2009, SAS Institute Inc. All rights reserved.

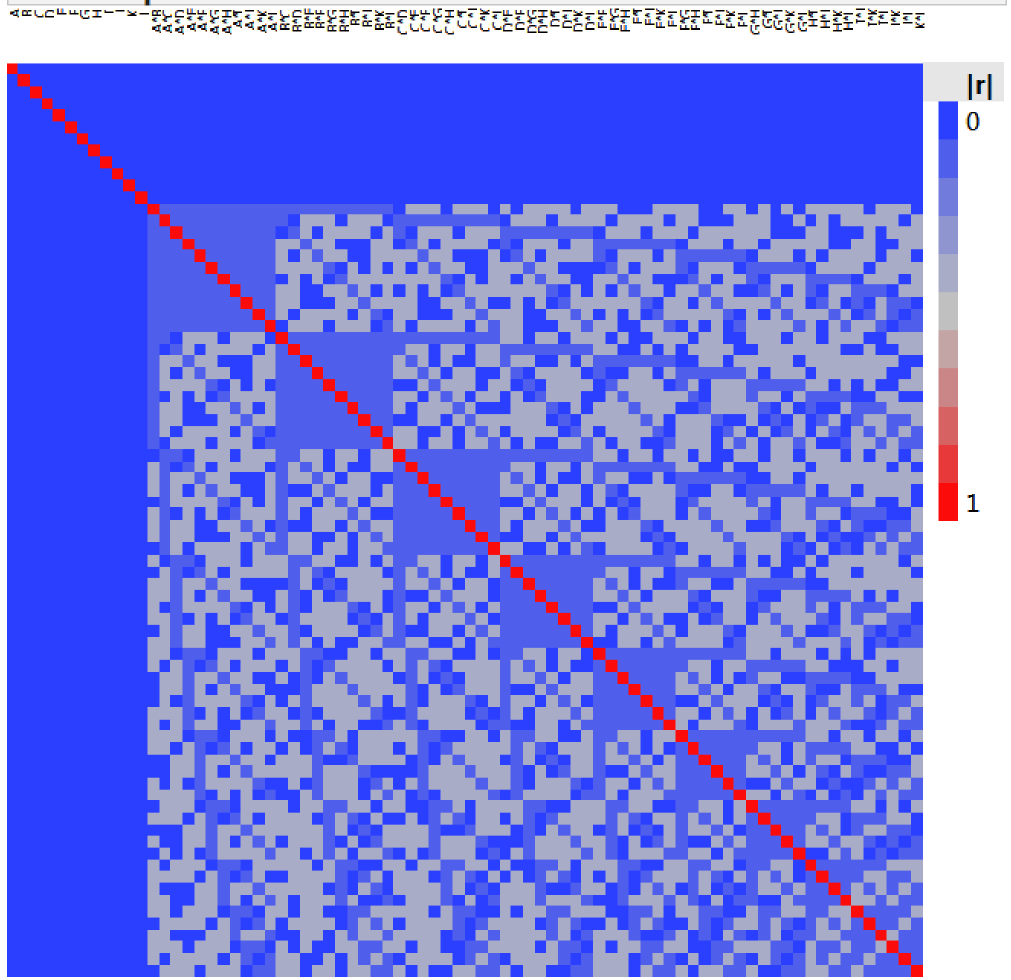
Agenda

1. Definitive Screening Designs
2. Orthogonal or Near Orthogonal Arrays in Screening Designer
3. Fast Flexible Filling (FFF) Designs

Copyright © 2009, SAS Institute Inc. All rights reserved.

Definitive Screening Designs

Color Map On Correlations



Copyright © 2009, SAS Institute Inc. All rights reserved.

Joint work with Chris Nachtsheim

Copyright © 2009, SAS Institute Inc. All rights reserved.

Motivation: Problems with Standard Screening Designs

Resolution III designs confound main effects and two-factor interactions.

Plackett-Burman designs have “complex aliasing “of the main effects by two-factor interactions.

Resolution IV designs confound two-factor interactions with each other, so if one is active, you usually need further runs to resolve the active effects.

Center runs give an overall measure of curvature but you do not know which factor(s) are causing the curvature.

Copyright © 2009, SAS Institute Inc. All rights reserved.

Screening Design – Brad's Wish List

1. Orthogonal main effects.
2. Main effects uncorrelated with two-factor interactions and quadratic effects.
3. Estimable quadratic effects – three-level design.
4. Small number of runs – roughly twice as many runs as factors.
5. Projections to 3 or fewer factors allow fitting RSM model.

Copyright © 2009, SAS Institute Inc. All rights reserved.

Screening Conundrum – Two Models

Suppose we have 6 factors and can afford 12 runs. We are interested in main effects but we are concerned about possible two-factor interactions.

The full model containing both 6 first-order and 15 second-order terms is:

$$\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$$

But n = 12, so we can only fit the intercept and the main effects:

$$\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}^*$$

Standard result: some main effects estimates are biased:

$$\mathrm{E}(\hat{\boldsymbol{\beta}}_1) = \boldsymbol{\beta}_1 + \mathbf{A}\boldsymbol{\beta}_2$$

where the “alias” matrix is: $\mathbf{A} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2$

Alias Matrix of 6 Factor Plackett-Burman design

Alias Matrix

Effect	A*B	A*C	A*D	A*E	A*F	B*C	B*D	B*E	B*F	C*D	C*E	C*F	D*E	D*F	E*F
Intercept	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A	0	0	0	0	0	0.333	-0.33	0.333	0.333	0.333	-0.33	0.333	-0.33	0.333	-0.33
B	0	0.333	-0.33	0.333	0.333	0	0	0	0	-0.33	-0.33	-0.33	-0.33	0.333	0.333
C	0.333	0	0.333	-0.33	0.333	0	-0.33	-0.33	-0.33	0	0	0	-0.33	0.333	-0.33
D	-0.33	0.333	0	-0.33	0.333	-0.33	0	-0.33	0.333	0	-0.33	0.333	0	0	-0.33
E	0.333	-0.33	-0.33	0	-0.33	-0.33	-0.33	0	0.333	-0.33	0	-0.33	0	-0.33	0
F	0.333	0.333	0.333	-0.33	0	-0.33	0.333	0.333	0	0.333	-0.33	0	-0.33	0	0

All main effects are potentially biased by 10 possible two-factor interactions

Copyright © 2009, SAS Institute Inc. All rights reserved.

If only there were another 6 factor design with this alias matrix:

Alias Matrix

Effect	A*B	A*C	A*D	A*E	A*F	B*C	B*D	B*E	B*F	C*D	C*E	C*F	D*E	D*F	E*F
Intercept	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Copyright © 2009, SAS Institute Inc. All rights reserved.

Turns out there is:
Introducing Definitive Screening Designs

Six foldover pairs

Run	A	B	C	D	E	F
1	0	1	-1	-1	-1	-1
2	0	-1	1	1	1	1
3	1	0	-1	1	1	-1
4	-1	0	1	-1	-1	1
5	-1	-1	0	1	-1	-1
6	1	1	0	-1	1	1
7	-1	1	1	0	1	-1
8	1	-1	-1	0	-1	1
9	1	-1	1	-1	0	-1
10	-1	1	-1	1	0	1
11	1	1	1	1	-1	0
12	-1	-1	-1	-1	1	0
13	0	0	0	0	0	0

Definitive Screening Design for 6 factors

Middle value in each row

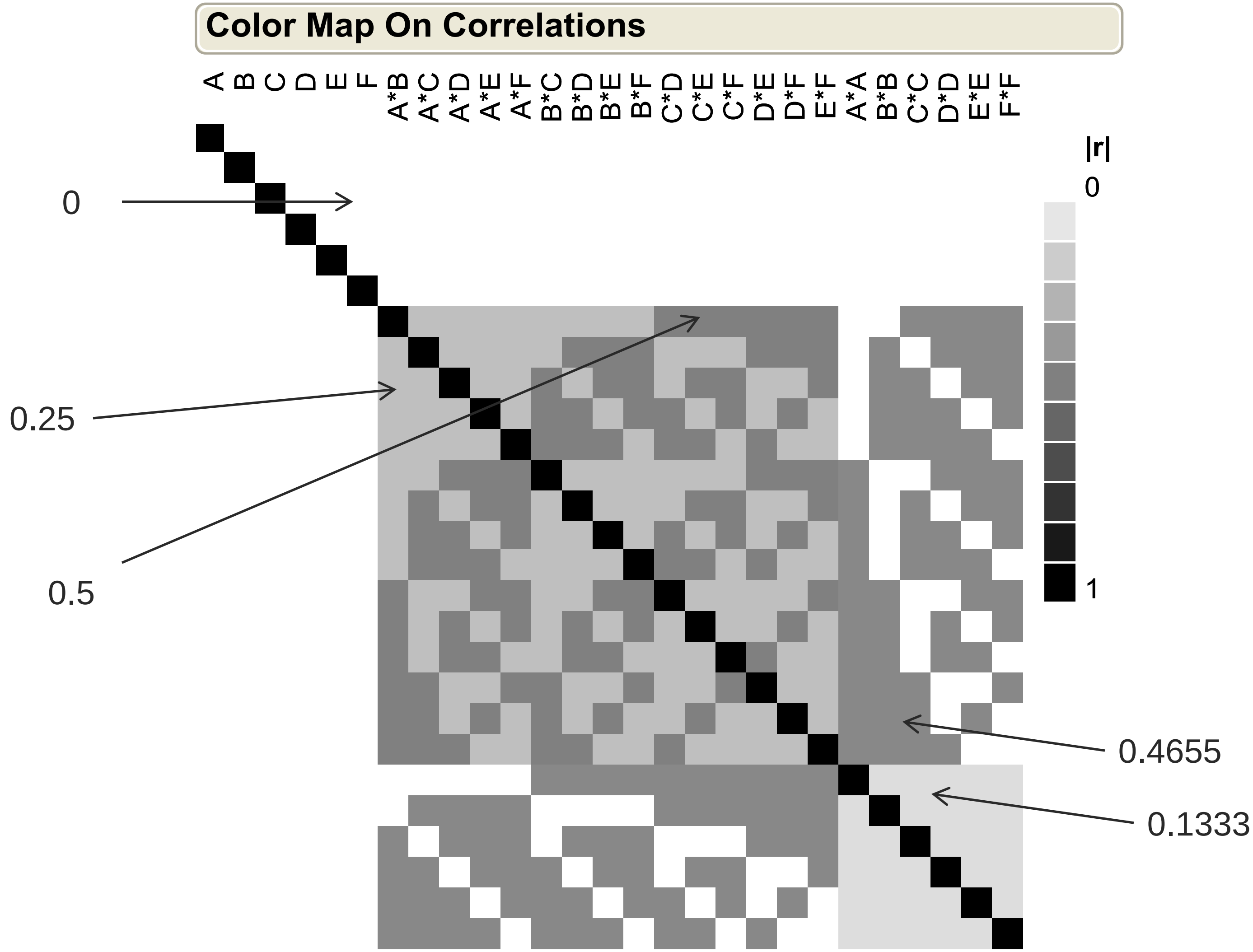
Run	A	B	C	D	E	F
1	0	1	-1	-1	-1	-1
2	0	-1	1	1	1	1
3	1	0	-1	1	1	-1
4	-1	0	1	-1	-1	1
5	-1	-1	0	1	-1	-1
6	1	1	0	-1	1	1
7	-1	1	1	0	1	-1
8	1	-1	-1	0	-1	1
9	1	-1	1	-1	0	-1
10	-1	1	-1	1	0	1
11	1	1	1	1	-1	0
12	-1	-1	-1	-1	1	0
13	0	0	0	0	0	0

Definitive Screening Design for 6 factors

Run	A	B	C	D	E	F
1	0	1	-1	-1	-1	-1
2	0	-1	1	1	1	1
3	1	0	-1	1	1	-1
4	-1	0	1	-1	-1	1
5	-1	-1	0	1	-1	-1
6	1	1	0	-1	1	1
7	-1	1	1	0	1	-1
8	1	-1	-1	0	-1	1
9	1	-1	1	-1	0	-1
10	-1	1	-1	1	0	1
11	1	1	1	1	-1	0
12	-1	-1	-1	-1	1	0
13	0	0	0	0	0	0

Center Point → (row 13)

Useful Column Correlation Pattern



Copyright © 2009, SAS Institute Inc. All rights reserved.

Journal of Quality Technology paper (2011)

A Class of Three-Level Designs for Definitive Screening in the Presence of Second-Order Effects

BRADLEY JONES

SAS Institute, Cary, NC 27513

CHRISTOPHER J. NACHTSHEIM

Carlson School of Management, University of Minnesota, Minneapolis, MN 55455

Copyright © 2009, SAS Institute Inc. All rights reserved.

Table from initial *JQT* paper

Table 1: General design structure for m factors

Foldover Pair	Run (i)	Factor Levels $x_{i,1}$	$x_{i,2}$	$x_{i,3}$	$\cdots$	$x_{i,m}$
1	1	0	± 1	± 1	$\cdots$	± 1
	2	0	∓ 1	∓ 1	$\cdots$	∓ 1
2	3	± 1	0	± 1	$\cdots$	± 1
	4	∓ 1	0	∓ 1	$\cdots$	∓ 1
3	5	± 1	± 1	0	$\cdots$	± 1
	6	∓ 1	∓ 1	0	$\cdots$	∓ 1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
m	$2m-1$	± 1	± 1	± 1	$\cdots$	0
	$2m$	∓ 1	∓ 1	∓ 1	$\cdots$	0
Centerpoint	$m+1$	0	0	0	$\cdots$	0

Copyright © 2009, SAS Institute Inc. All rights reserved.

Follow up paper in *JQT* (2012)

Constructing Definitive Screening Designs Using Conference Matrices

LILI XIAO

Department of Mathematical Sciences, Tsinghua University, Beijing, China 100084

DENNIS K. J. LIN

Department of Statistics, Pennsylvania State University, University Park, PA 16802

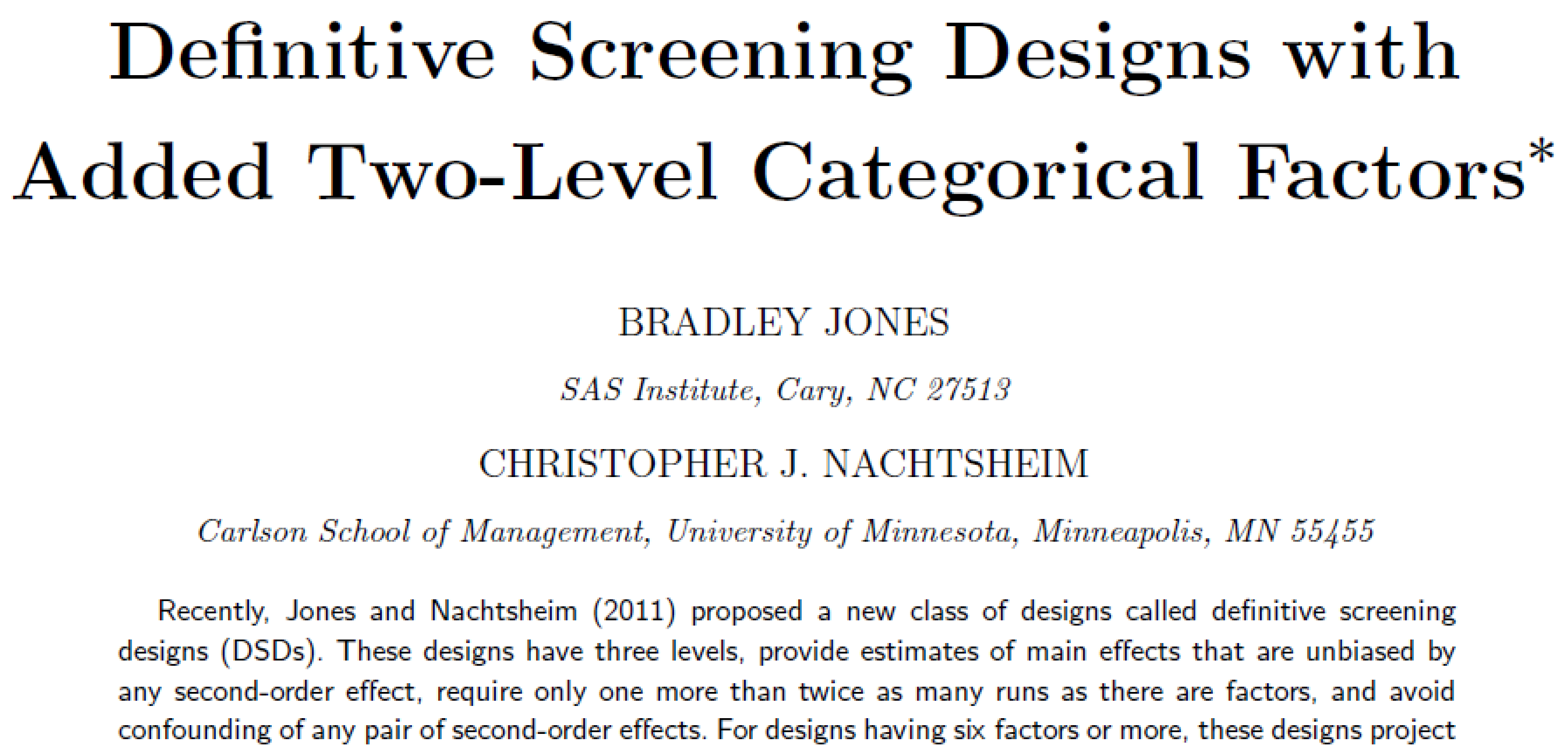
FENGSHAN BAI

Department of Mathematical Sciences, Tsinghua University, Beijing, China 100084

Jones and Nachtsheim (2011) propose a new class of designs for definitive screening. These designs

Copyright © 2009, SAS Institute Inc. All rights reserved.

Follow up paper in *JQT* (2013)


Definitive Screening Designs with Added Two-Level Categorical Factors*

BRADLEY JONES

SAS Institute, Cary, NC 27513

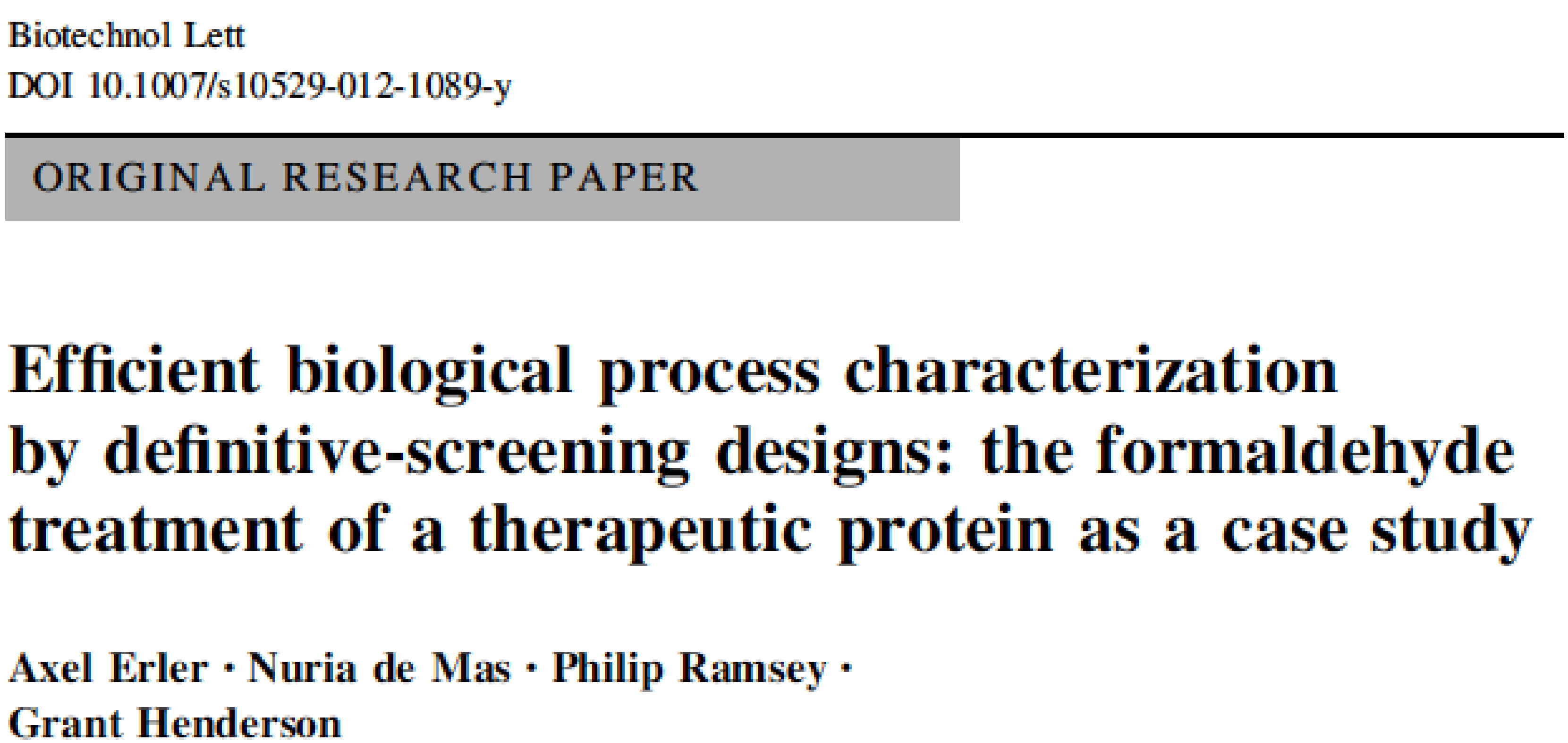
CHRISTOPHER J. NACHTSHEIM

Carlson School of Management, University of Minnesota, Minneapolis, MN 55455

Recently, Jones and Nachtsheim (2011) proposed a new class of designs called definitive screening designs (DSDs). These designs have three levels, provide estimates of main effects that are unbiased by any second-order effect, require only one more than twice as many runs as there are factors, and avoid confounding of any pair of second-order effects. For designs having six factors or more, these designs project

Copyright © 2009, SAS Institute Inc. All rights reserved.

Impact: First published DSD case study


Biotechnol Lett
DOI 10.1007/s10529-012-1089-y

ORIGINAL RESEARCH PAPER

Efficient biological process characterization by definitive-screening designs: the formaldehyde treatment of a therapeutic protein as a case study

Axel Erler · Nuria de Mas · Philip Ramsey · Grant Henderson

From the conclusions of first paper:

"***Definitive-screening designs were used to efficiently select a model describing the formulation of a protein under clinical development. The ability of the single definitive screening design to identify and model all the active effects obviated the need for further experimentation, reducing the total number of experimental runs required to 17 from the greater than or equal to 70 runs that would have been necessary using the traditional screening/RSM approach.***"

Impact: A break-through solution for sequestering greenhouse gasses

2012 Statistics in Chemistry Award Winner

Bradley Jones, SAS Institute, JMP Division

Scott Allen, Novomer, Inc.

To Bradley Jones and Scott Allen for outstanding collaborative work in developing a new catalyst for CO_2-based polymers that sequester CO_2.

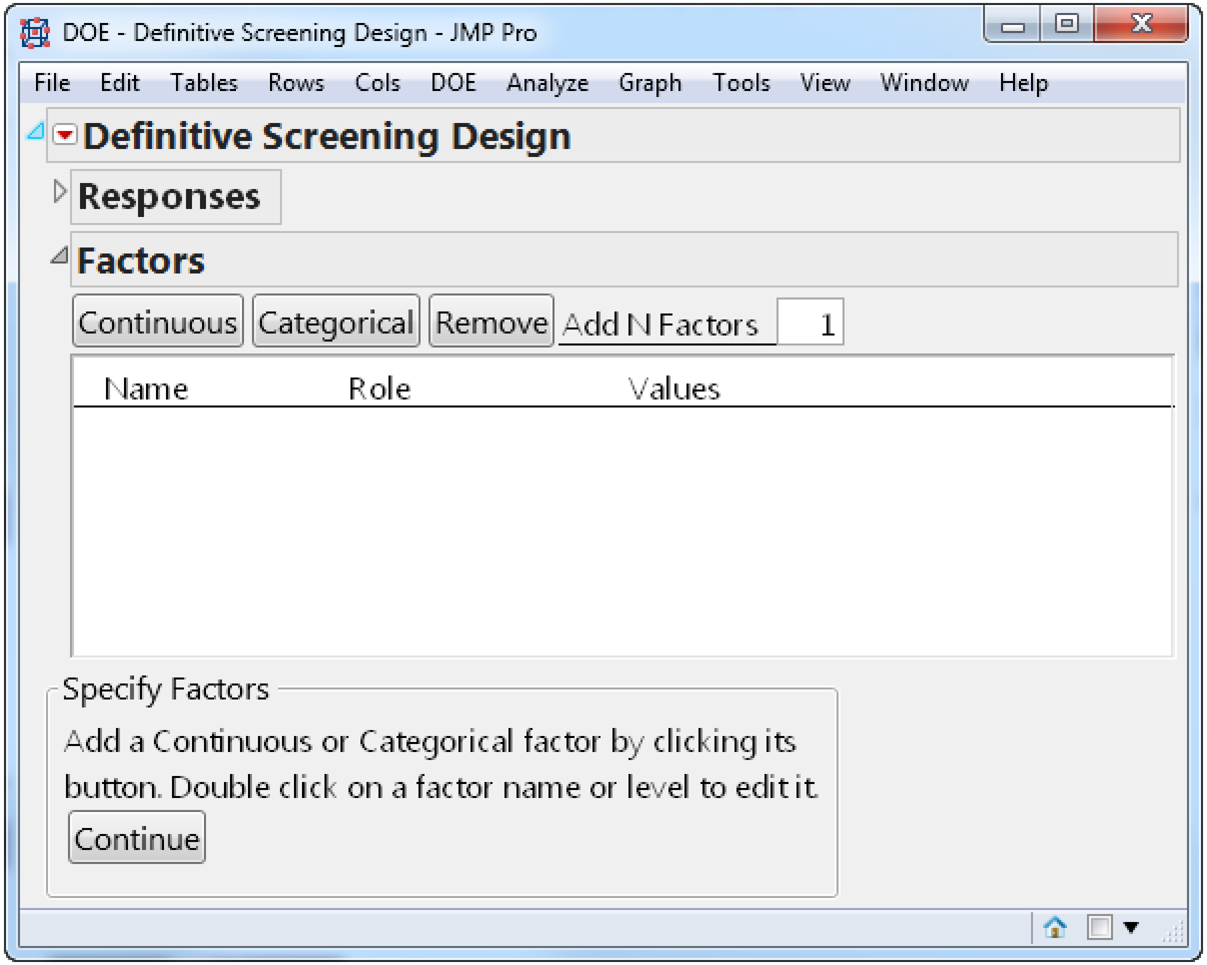
Awards for Definitive Screening Designs

"A New Class of Three-Level Screening Designs for Definitive Screening in the Presence of Second-Order Effects", *Journal of Quality Technology*, Jan., 2011

1. Brumbaugh Award, 2011
2. Lloyd S. Nelson Award, 2012
3. Statistics in Chemistry Award, 2012

Copyright © 2009, SAS Institute Inc. All rights reserved.

JMP Demo



Copyright © 2009, SAS Institute Inc. All rights reserved.

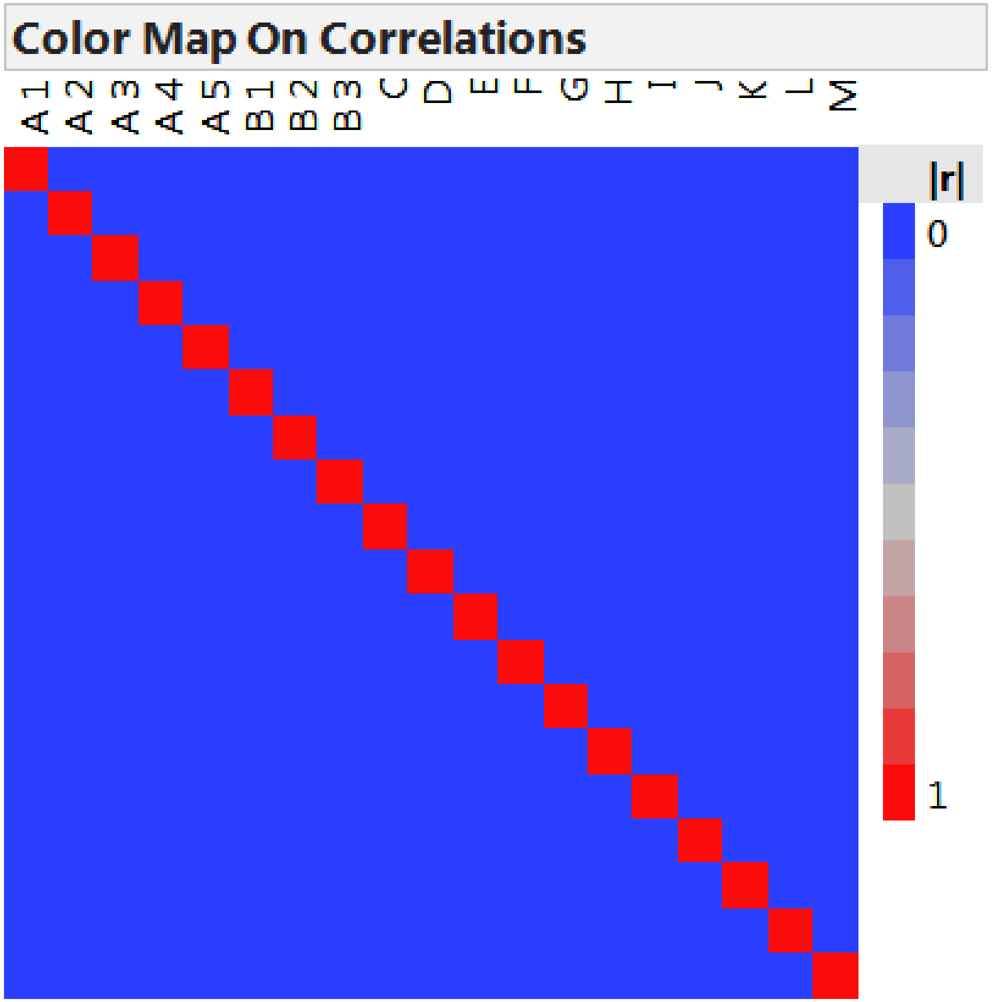
Bottom Line

Recommendation:

Stop using 2^{k-p} designs for five or more continuous factors!!

Copyright © 2009, SAS Institute Inc. All rights reserved.

Orthogonal or Near Orthogonal Arrays



Copyright © 2009, SAS Institute Inc. All rights reserved.

Work with Ryan Lekivetz

Copyright © 2009, SAS Institute Inc. All rights reserved.

Motivating Example

1. Suppose there are 6 machines and 4 suppliers of raw material.
2. The machines have 11 controllable settings.
3. You want to know whether the machines and suppliers make a difference
4. You also want to know what the vital few controls are.

Copyright © 2009, SAS Institute Inc. All rights reserved.

JMP Demo

Screening Design

Responses

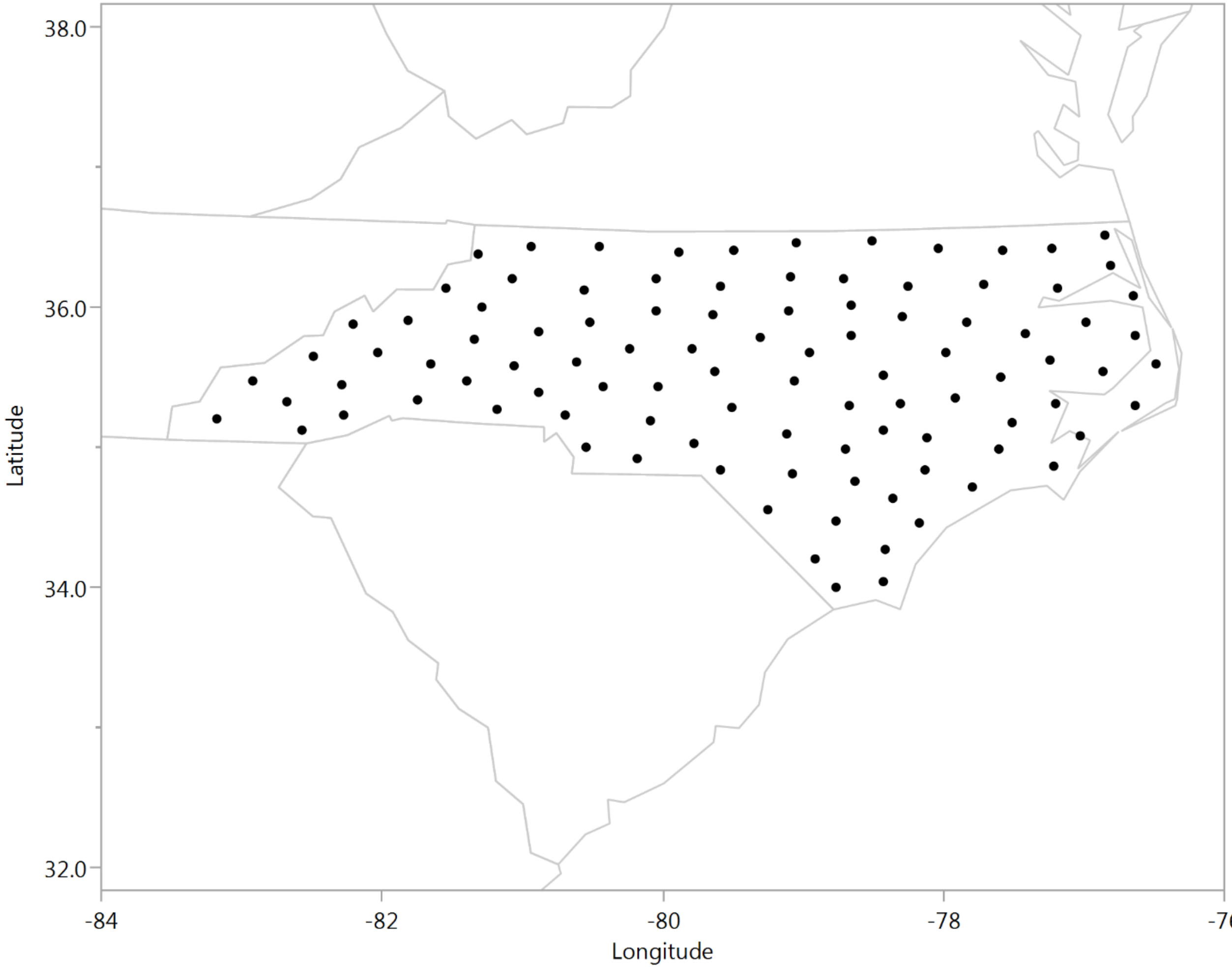
Factors

Continuous | Discrete Numeric | Categorical | Remove | Add N Factors 1

Name	Role	Values					
Machine	Categorical	1	2	3	4	5	6
Supplier	Categorical	1	2	3	4		
C	Continuous	-1	1				
D	Continuous	-1	1				
E	Continuous	-1	1				
F	Continuous	-1	1				
G	Continuous	-1	1				
H	Continuous	-1	1				
I	Continuous	-1	1				
J	Continuous	-1	1				
K	Continuous	-1	1				
L	Continuous	-1	1				
M	Continuous	-1	1				

Copyright © 2009, SAS Institute Inc. All rights reserved.

Fast Flexible Filling (FFF) Designs



Copyright © 2009, SAS Institute Inc. All rights reserved

Latin Hypercube Designs

"It ain't what you don't know that gets you into trouble. It's what you know for sure that just ain't so." – *Mark Twain*

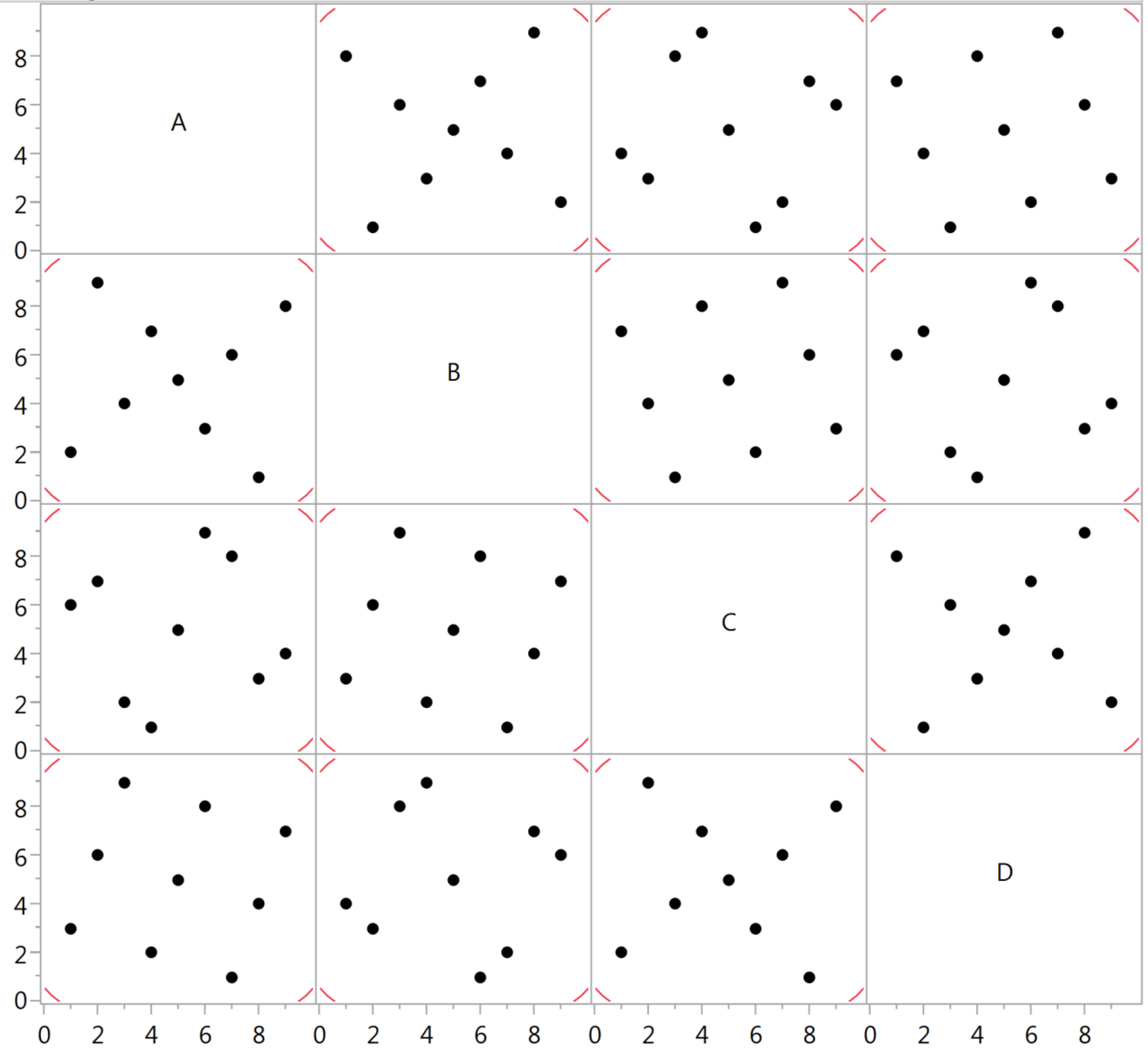
Copyright © 2009, SAS Institute Inc. Al

Multivariate

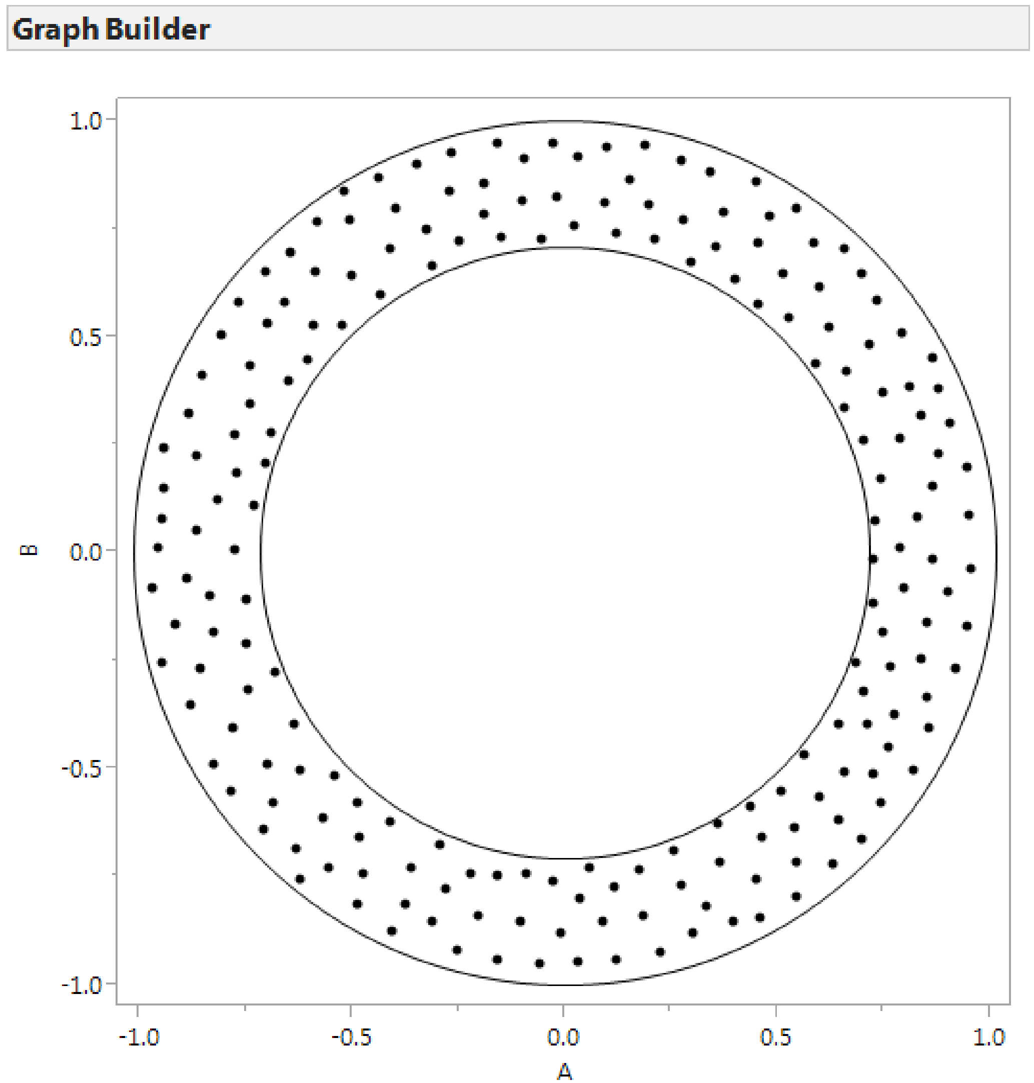
Correlations

	A	B	C	D
A	1.0000	0.0000	0.0000	0.0000
B	0.0000	1.0000	0.0000	0.0000
C	0.0000	0.0000	1.0000	0.0000
D	0.0000	0.0000	0.0000	1.0000

Scatterplot Matrix



Fast Flexible Filling (FFF) – Amuse-bouche

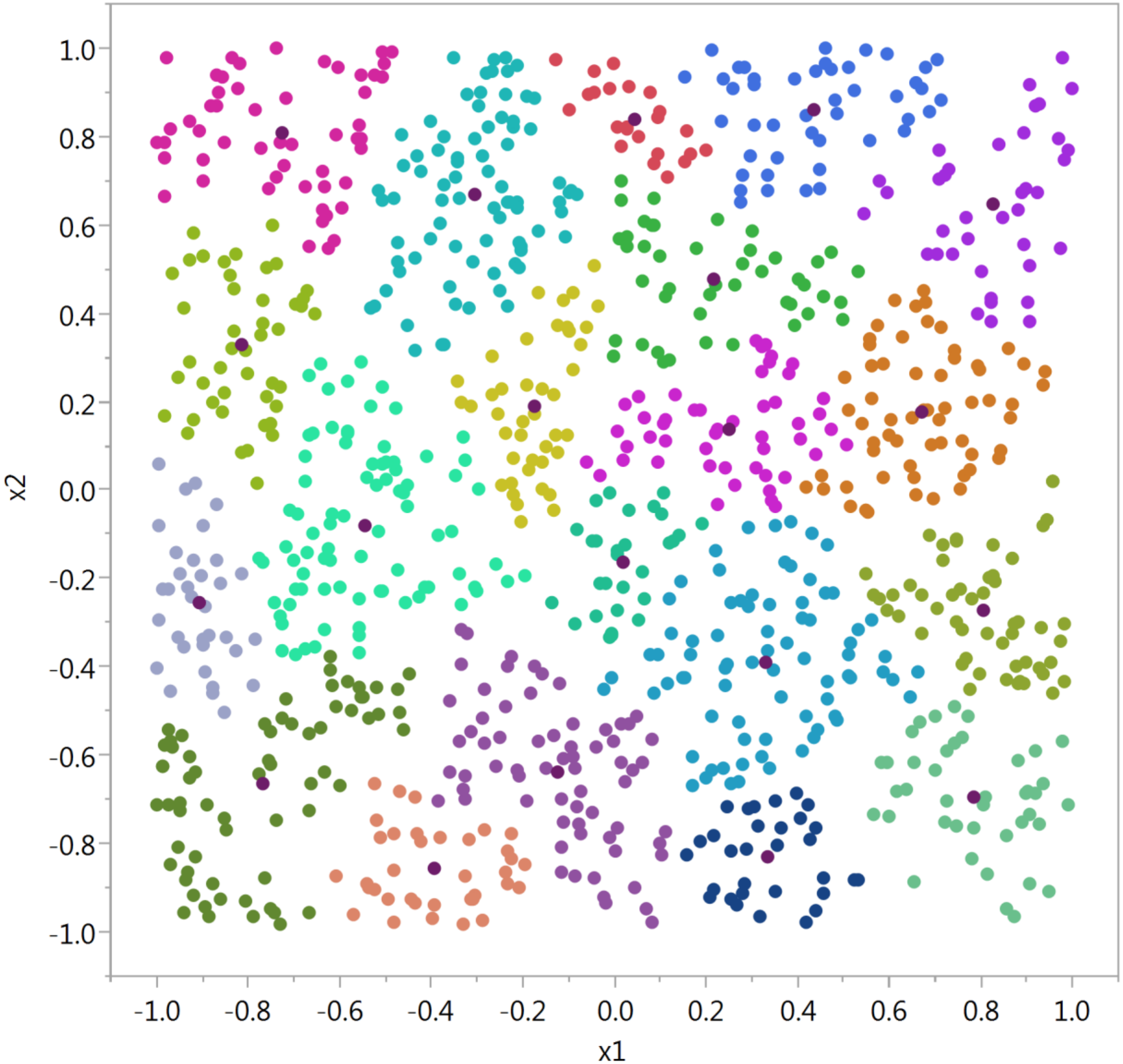


FFF Design Construction

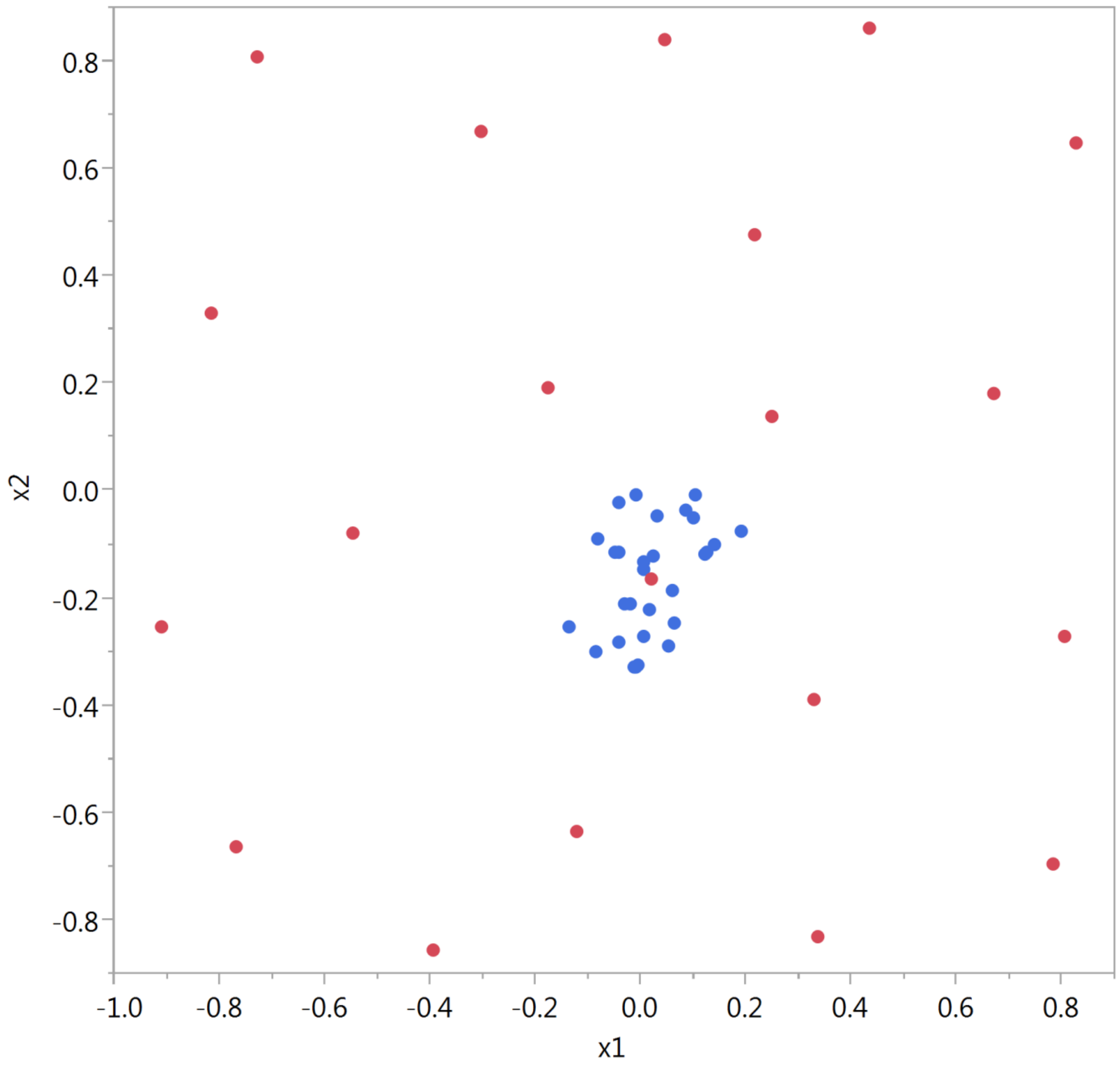
1. Generate a large random set of feasible points.
    a) If *n* is the desired number of runs
    b) Use maximum of 10,000 or 50*n* feasible points
2. Cluster the points using Fast Ward adaptation
    a) Create *n* clusters
    b) One for each design point
3. Compute the centroids of each cluster

Copyright © 2009, SAS Institute Inc. All rights reserved.

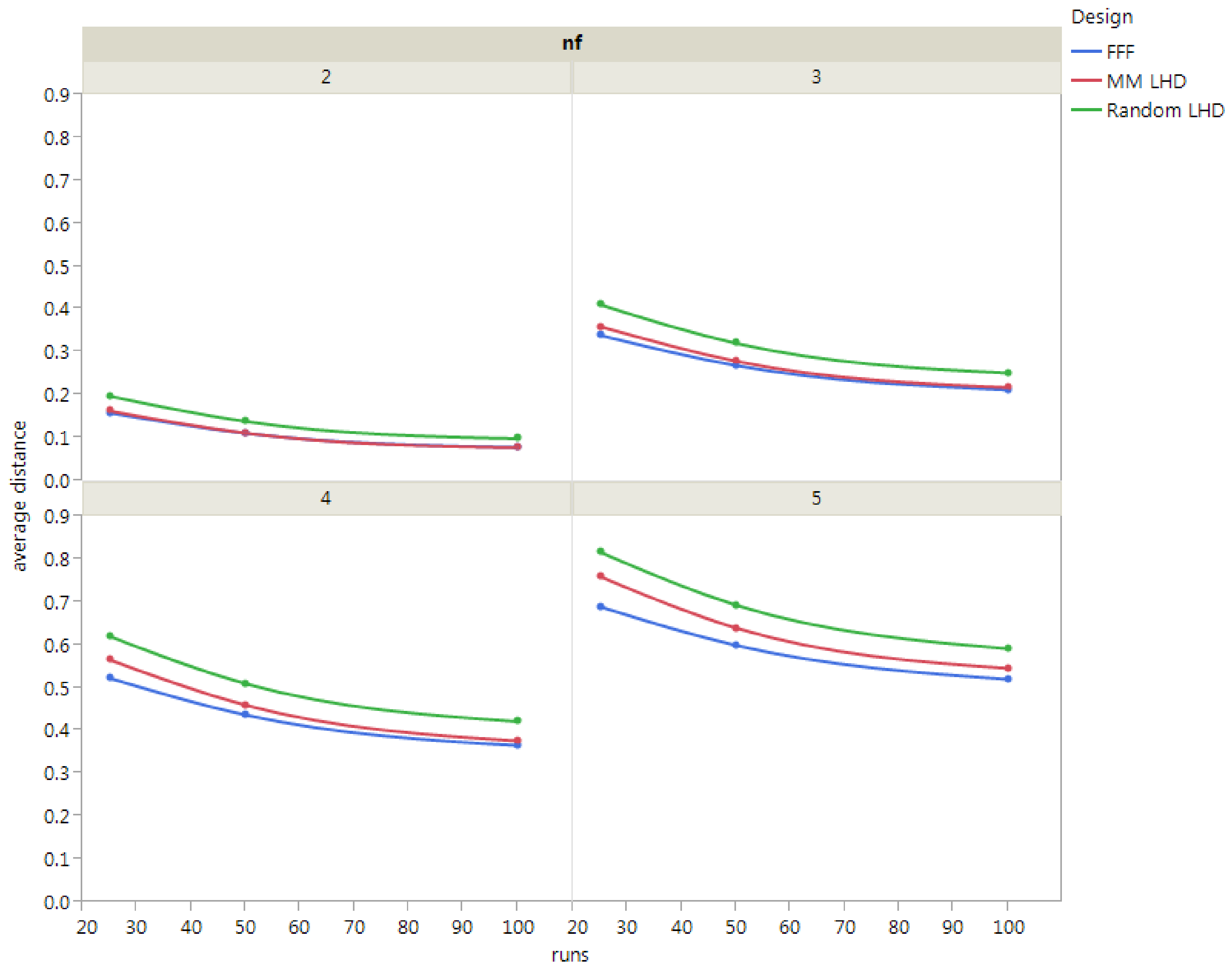
1000 points clustered into 20 groups



20 group centroids with one cluster



Average distance to the nearest design point from an arbitrary point by design type, number of factors and number of runs

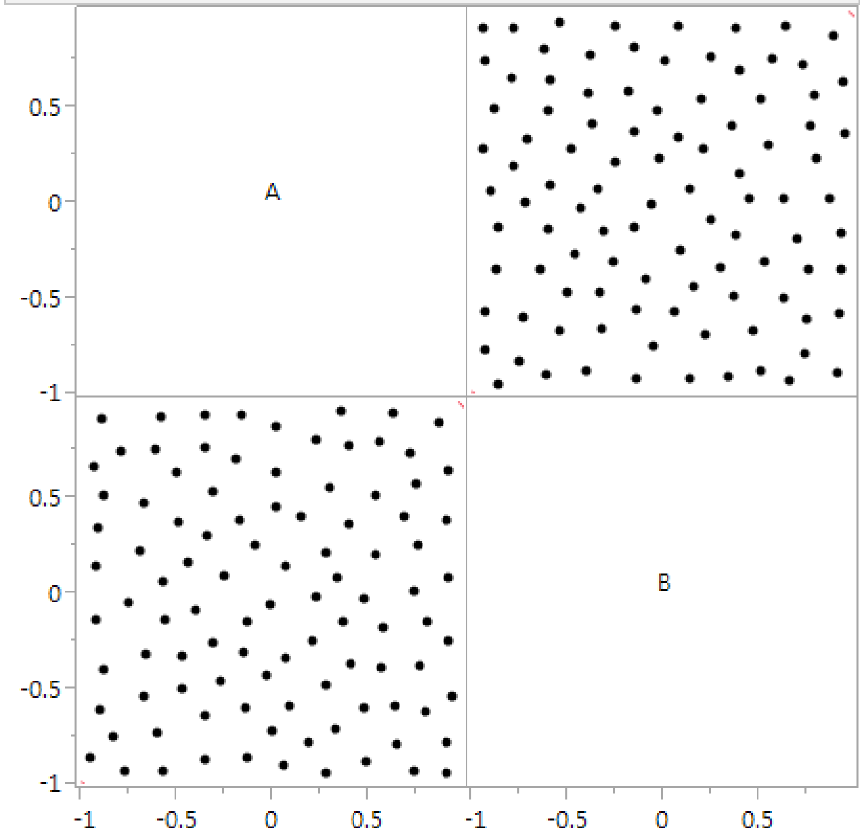


Two Factors – 100 Runs

Correlations

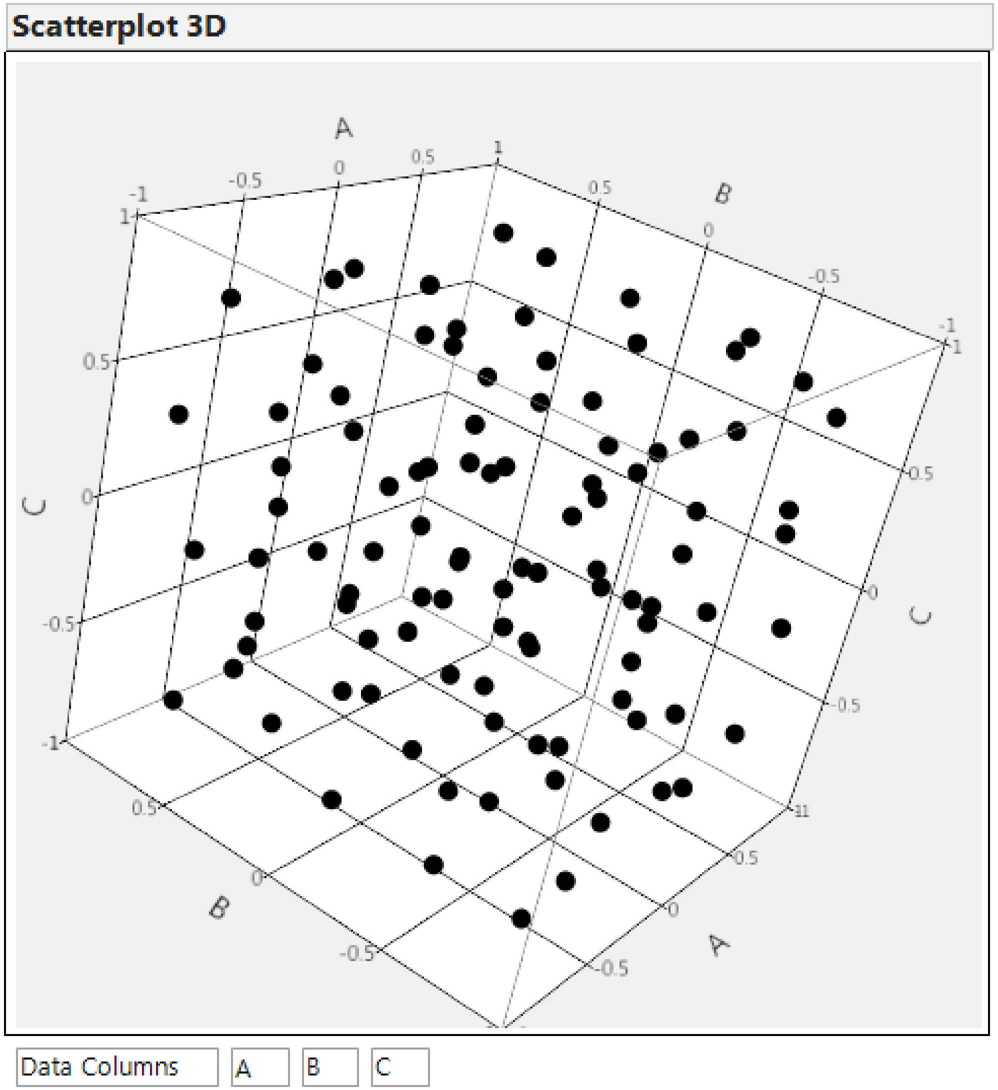
	A	B
A	1.0000	-0.0286
B	-0.0286	1.0000

Scatterplot Matrix



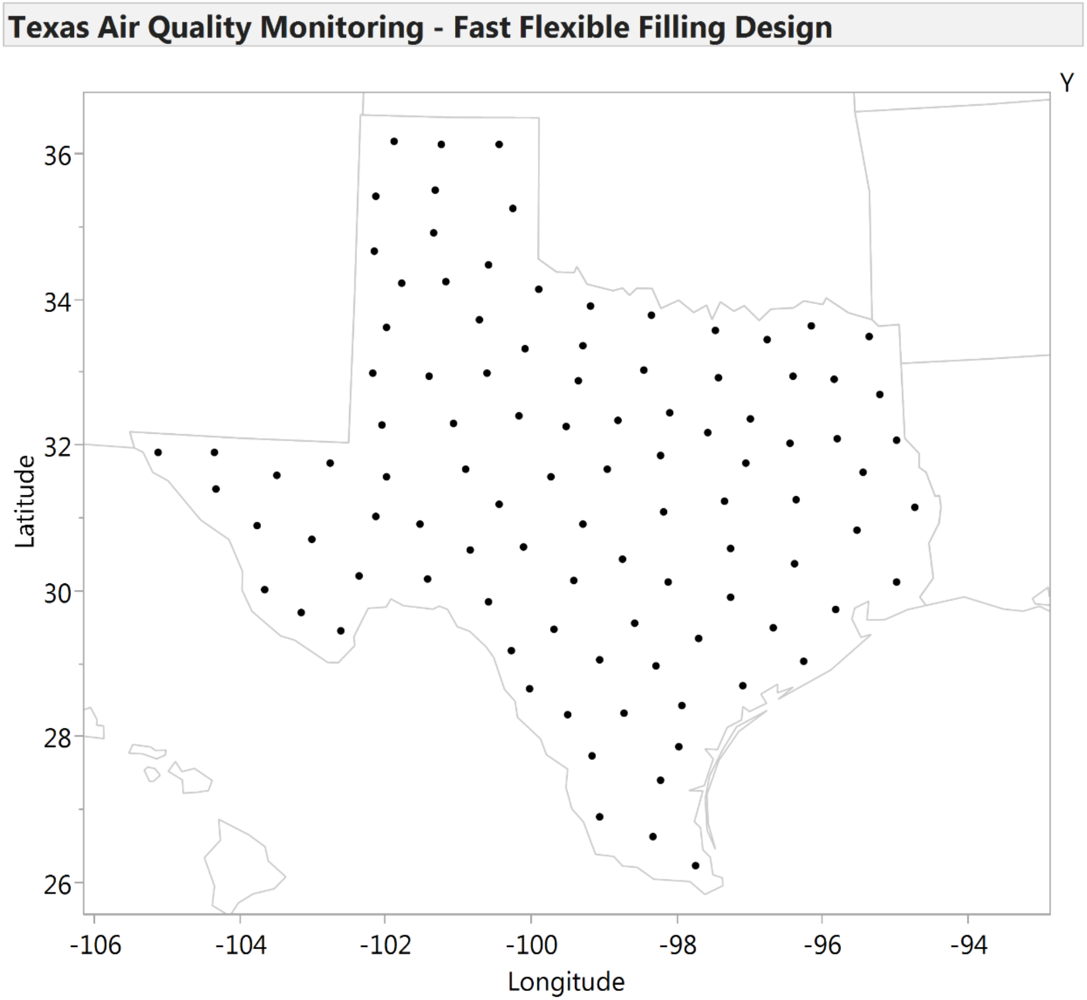
Copyright © 2009, SAS Institute Inc. All rights reserved.

Three Factors – 100 Runs



Copyright © 2009, SAS Institute Inc. All rights reserved.

JMP Demonstration



Copyright © 2009, SAS Institute Inc. All rights reserved.

jmp®

Statistical Discovery.™ From SAS.

SEEING IS BELIEVING

Copyright © 2009 SAS Institute Inc. All rights reserved.