Course 02418
Statistical Modelling: Theory and practice

Module 10: Design of Experiments

Murat Kulahci

DTU Informatics
Technical University of Denmark
2800 Lyngby – Denmark
E-mail: mk@imm.dtu.dk

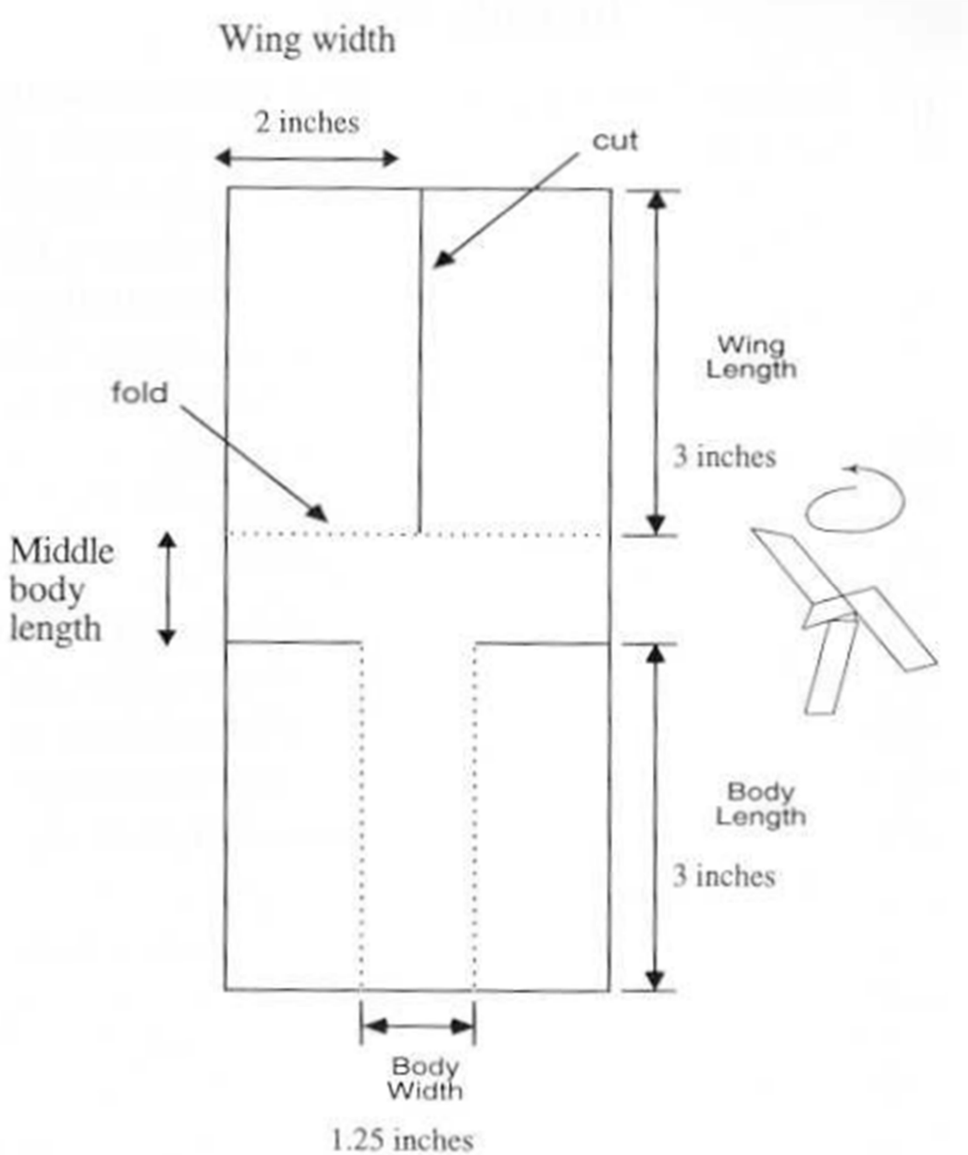
Experimentation

- An **experiment** is a test or a series of tests
- Experiments are used widely in the engineering and physical sciences
  - Process characterization & optimization
  - Evaluation of material properties
  - Product design & development
  - Component & system tolerance determination
- "All experiments are designed experiments, some are poorly designed, some are well-designed"

Strategy of Experimentation

- **“Best-guess” experiments**
  - Used a lot
  - More successful than you might suspect, but there are disadvantages…
- **One-factor-at-a-time (OFAT) experiments**
  - Sometimes associated with the “scientific” or “engineering” method
  - Devastated by interaction, also very inefficient
- **Statistically designed experiments**
  - Based on Fisher’s factorial concept

Paper Helicopter



Basic Principles of DOE

- **Randomization**
  - Running the trials in an experiment in random order
  - Notion of balancing out effects of “lurking” variables
- **Replication**
  - Sample size (improving precision of effect estimation, estimation of error or background noise)
  - Replication versus repeat measurements?
- **Blocking**
  - Dealing with nuisance factors

2^k Factorial Design

- **Special case** of the general factorial design; *k* factors, all at two levels
- The two levels are usually called **low** and **high** (they could be either quantitative or qualitative)
  - Assumes response is approximately linear over the range chosen
- Very widely used in industrial experimentation
- Form a basic “building block” for other very useful experimental designs (DNA)
- Special (short-cut) methods for analysis

Example

- In a chemical process, it is speculated that the reactant concentration and catalyst amount affect the yield
- We would like to investigate whether and how these factor are affecting the yield
- We change the current settings for these factors which were 20% and 1.5lb the reactant concentration and the catalyst amount respectively
- There are 4 possible factorial combinations if both factors are tested only at two levels.
- To estimate the error variance we also decide to replicate the factorial experiments three times

Data

Concentration	Catalyst Amount	Yield
15%	1lb	28
15%	1lb	25
15%	1lb	27
25%	1lb	36
25%	1lb	32
25%	1lb	32
15%	2lb	18
15%	2lb	19
15%	2lb	23
25%	2lb	31
25%	2lb	30
25%	2lb	29

How to analyze this data?

- We can use a simple regression model with the main effects and the two factor interaction

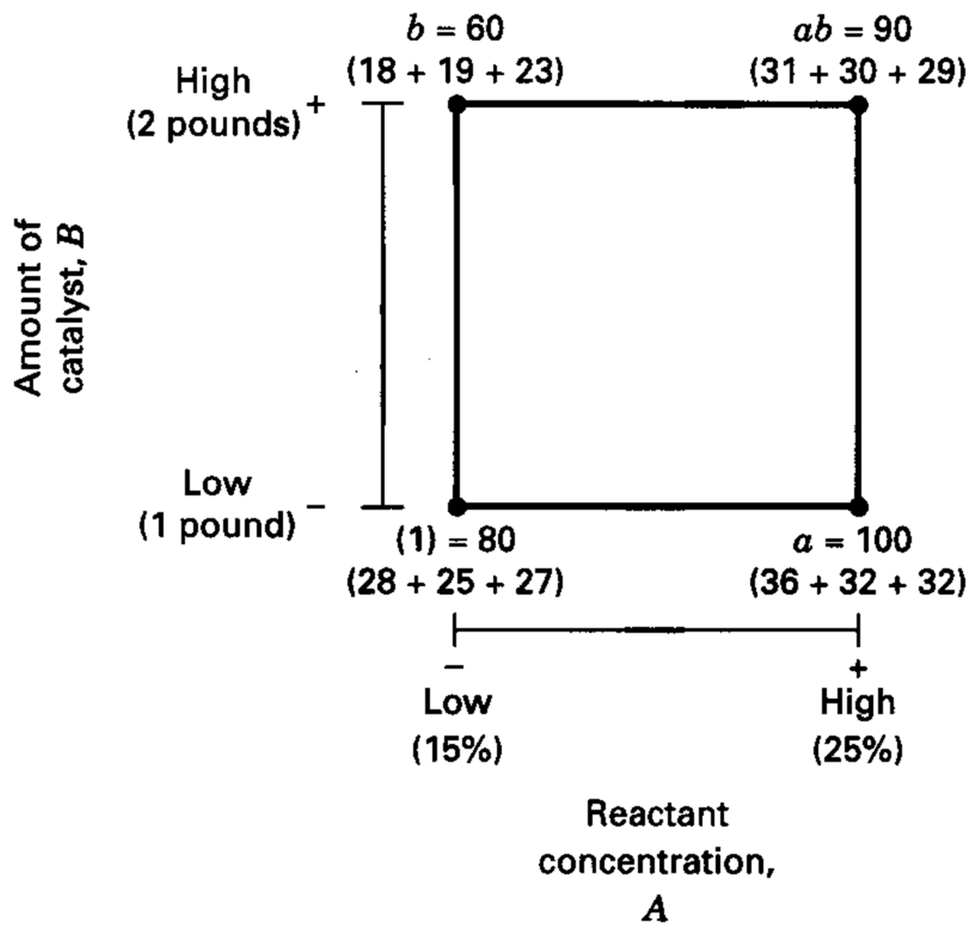
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon$$

- Using the usual equation, we have

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$$

$$Cov\left(\hat{\boldsymbol{\beta}}\right) = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\sigma^2$$

Alternative Approach



Treatment combinations in the 2^2 design.

A = reactant concentration, B = catalyst amount,
y = yield

	Factor		Treatment	Replicate			
	A	B	Combination	I	II	III	Total
(1)	−	−	A low, B low	28	25	27	80
a	+	−	A high, B low	36	32	32	100
b	−	+	A low, B high	18	19	23	60
ab	+	+	A high, B high	31	30	29	90

To represent the experimental runs, we use lower case letters. If for an experimental run a factor is at its high level then we include the corresponding lower case letter in the representation. For example for the run "*a*" only factor *A* will be at the high level and the rest of the factors (in this case only *B*) will be at the low level. To represent the run with both factors are at the low level we use "(1)" representation.

Analysis Procedure for a Factorial Design

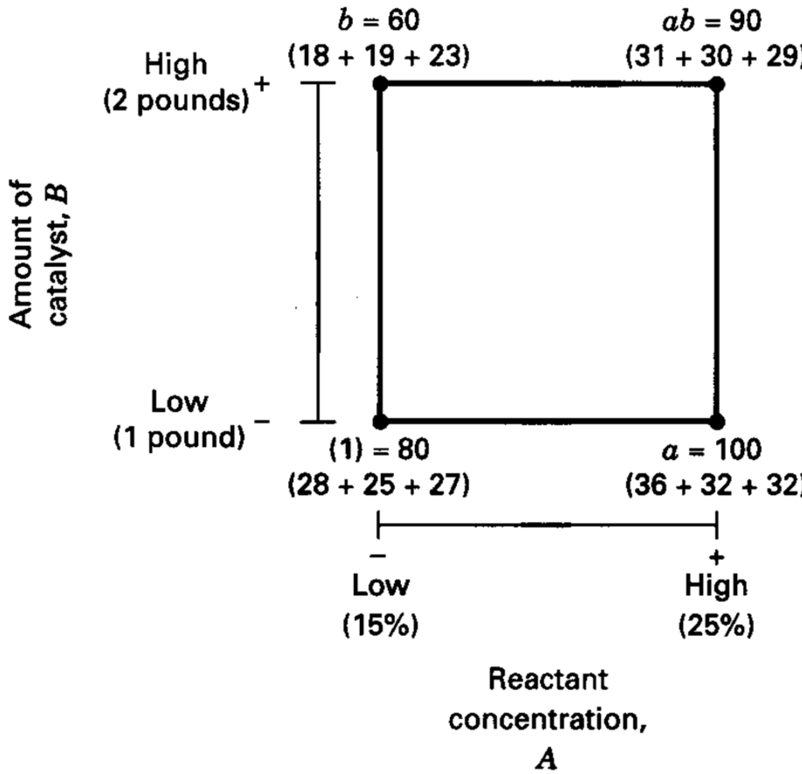
- Estimate factor **effects**
- **Formulate** model
  - With replication, use full model
  - With an unreplicated design, use normal probability plots
- Statistical **testing** (ANOVA)
- **Refine** the model
- Analyze **residuals** (graphical)
- **Interpret** results

Estimation of Factor Effects

$$A = \bar{y}_{A^+} - \bar{y}_{A^-}$$
$$= \frac{ab+a}{2n} - \frac{b+(1)}{2n}$$
$$= \frac{1}{2n}\left[ab + a - b - (1)\right]$$

$$B = \bar{y}_{B^+} - \bar{y}_{B^-}$$
$$= \frac{ab+b}{2n} - \frac{a+(1)}{2n}$$
$$= \frac{1}{2n}\left[ab + b - a - (1)\right]$$

$$AB = \frac{ab+(1)}{2n} - \frac{a+b}{2n}$$
$$= \frac{1}{2n}\left[ab + (1) - a - b\right]$$



Treatment combinations in the 2^2 design.

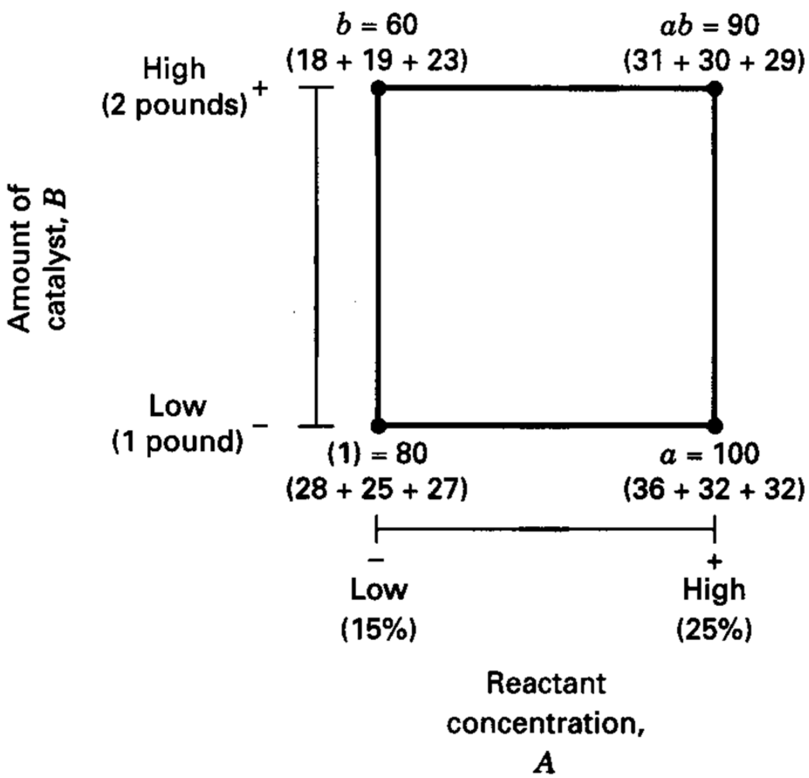
Estimation of Factor Effects

AB = average difference between the effect of B at the high level of A and the effect of B at the low level of A.

$$AB = \frac{ab - b}{2n} - \frac{a - (1)}{2n}$$

$$= \frac{ab + (1)}{2n} - \frac{a + b}{2n}$$

$$= \frac{1}{2n}\left[ab + (1) - a - b\right]$$



Treatment combinations in the 2^2 design.

Estimation of Factor Effects

Factor		Treatment	Replicate			
A	B	Combination	I	II	III	Total
−	−	A low, B low	28	25	27	80
+	−	A high, B low	36	32	32	100
−	+	A low, B high	18	19	23	60
+	+	A high, B high	31	30	29	90

$$A = \frac{1}{2*3}[90+100-60-80] = 8.33$$

$$B = \frac{1}{2*3}[90+60-100-80] = -5.0$$

$$AB = \frac{1}{2*3}[90+80-100-60] = 1.67$$

Practical interpretation?

Magnitude

Direction

Sum of Squares: Use of Contrasts

$$SS_{contrast} = \frac{Contrast^2}{4n}$$

Factor A	Factor B	Treatment Combination	Replicate I	Replicate II	Replicate III	Total
−	−	*A* low, *B* low	28	25	27	80
+	−	*A* high, *B* low	36	32	32	100
−	+	*A* low, *B* high	18	19	23	60
+	+	*A* high, *B* high	31	30	29	90

$$SS_A = \tfrac{1}{4*3}[90+100-60-80]^2 = 208.33$$

$$SS_B = \tfrac{1}{4*3}[90+60-100-80]^2 = 75$$

$$SS_{AB} = \tfrac{1}{4*3}[90+80-100-60]^2 = 8.33$$

Statistical Testing - ANOVA

Response: Conversion

ANOVA for Selected Factorial Model

Analysis of variance table [Partial sum of squares]

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F
Model	291.67	3	97.22	24.82	0.0002
A	*208.33*	*1*	*208.33*	*53.19*	*< 0.0001*
B	*75.00*	*1*	*75.00*	*19.15*	*0.0024*
AB	*8.33*	*1*	*8.33*	*2.13*	*0.1828*
Pure Error	31.33	8	3.92		
Cor Total	323.00	11			

Std. Dev.	1.98	R-Squared	0.9030
Mean	27.50	Adj R-Squared	0.8666
		Pred R-Squared	0.7817
PRESS	70.50		

The *F*-test for the "model" source is testing the significance of the overall model; that is, is either *A*, *B*, or *AB* or some combination of these effects important?

Statistical Testing - ANOVA

Factor	Coefficient Estimate	DF	Standard Error	95% CI Low	95% CI High
Intercept	27.50	1	0.57	26.18	28.82
A-Concert	4.17	1	0.57	2.85	5.48
B-Catalyst	-2.50	1	0.57	-3.82	-1.18
AB	0.83	1	0.57	-0.48	2.15

General formulas for the standard errors of the model coefficients and the confidence intervals are available. They will be given later.

Refine Model

Response: Conversion

ANOVA for Selected Factorial Model

Analysis of variance table [Partial sum of squares]

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F
Model	283.33	2	141.67	32.14	< 0.0001
A	*208.33*	*1*	*208.33*	*47.27*	*< 0.0001*
B	*75.00*	*1*	*75.00*	*17.02*	*0.0026*
Residual	39.67	9	4.41		
Lack of Fit	*8.33*	*1*	*8.33*	*2.13*	*0.1828*
Pure Error	*31.33*	*8*	*3.92*		
Cor Total	323.00	11			

Std. Dev.	2.10	R-Squared	0.8772
Mean	27.50	Adj R-Squared	0.8499
		Pred R-Squared	0.7817
PRESS	70.52		

There is now a residual sum of squares, partitioned into a “lack of fit” component (the AB interaction) and a “pure error” component

Regression Model for the Process

Factor	Coefficient Estimate	DF	Standard Error	95% CI Low	95% CI High
Intercept	27.50	1	0.61	26.13	28.87
A-Concent	4.17	1	0.61	2.80	5.54
B-Catalyst	-2.50	1	0.61	-3.87	-1.13

Final Equation in Terms of Coded Factors:

Conversior =
27.5
4.166667 * A
-2.5 * B

Final Equation in Terms of Actual Factors:

Conversior =
18.33333
0.833333 * Concentration
-5 * Catalyst

Conversion between Coded & Natural Variables

Factor A : Concentration

$$X_A = \frac{\text{Conc-}(\text{Conc}_{\text{Low}} + \text{Conc}_{\text{High}})/2}{(\text{Conc}_{\text{High}} - \text{Conc}_{\text{Low}})/2}$$

e.g.

Low = 15 High = 25

$$X_A = \frac{\text{Conc-}(15+25)/2}{(25-15)/2} = \frac{\text{Conc-}20}{5}$$

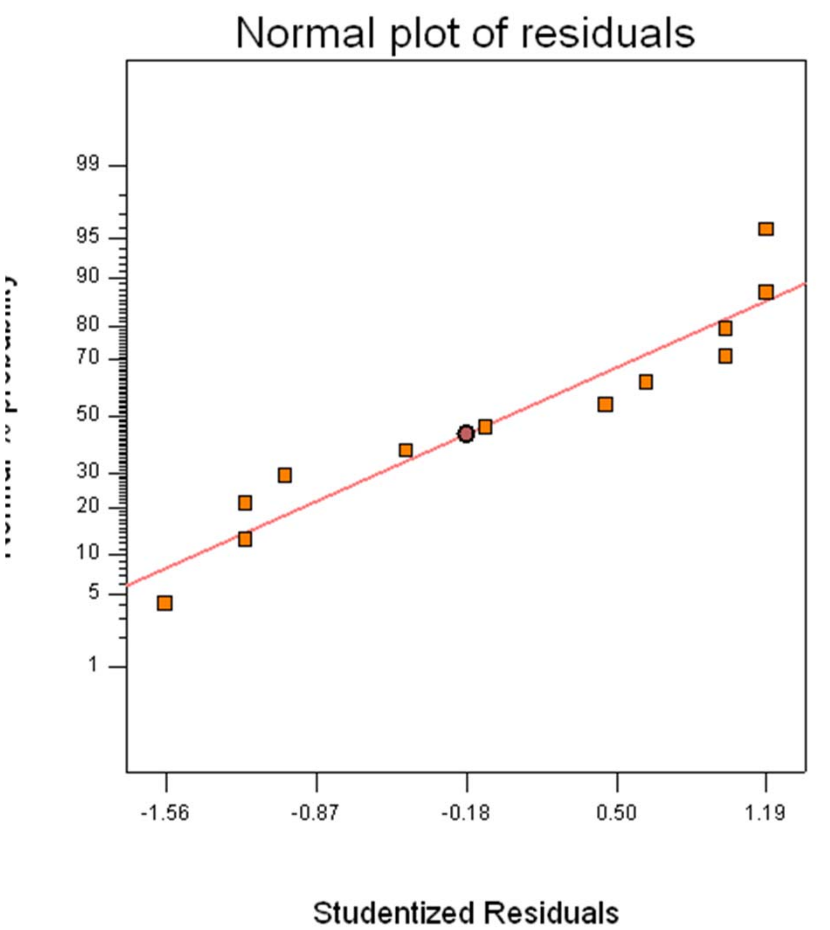
What happens if:

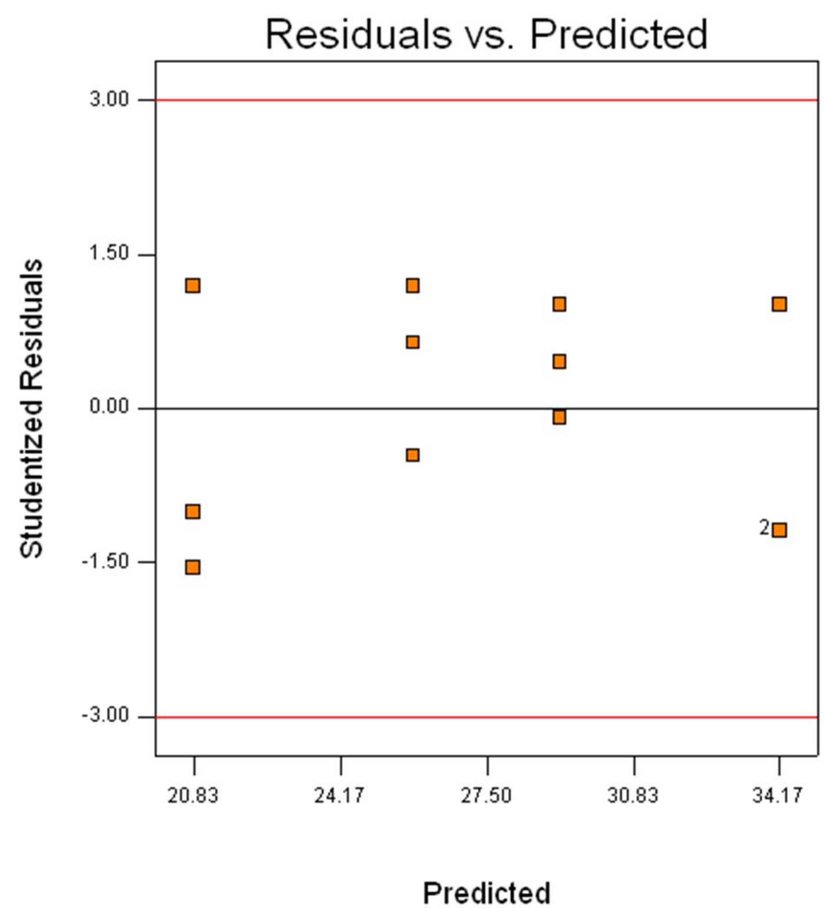
Conc = 15

Conc = 25

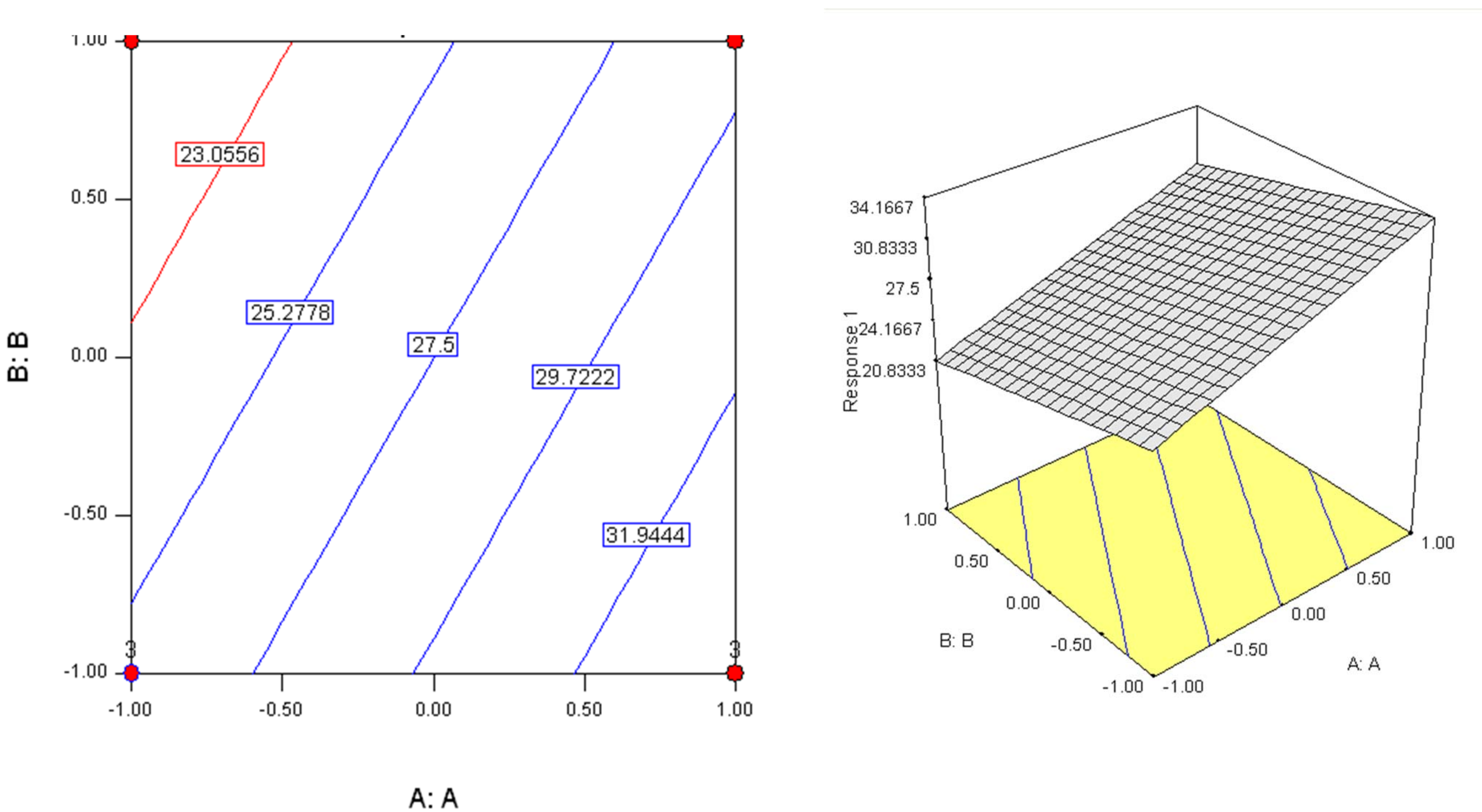
Conc = 20

Residuals and Diagnostic Checking

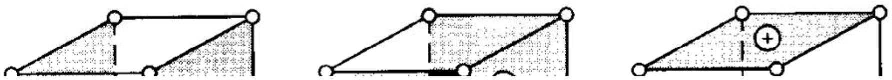


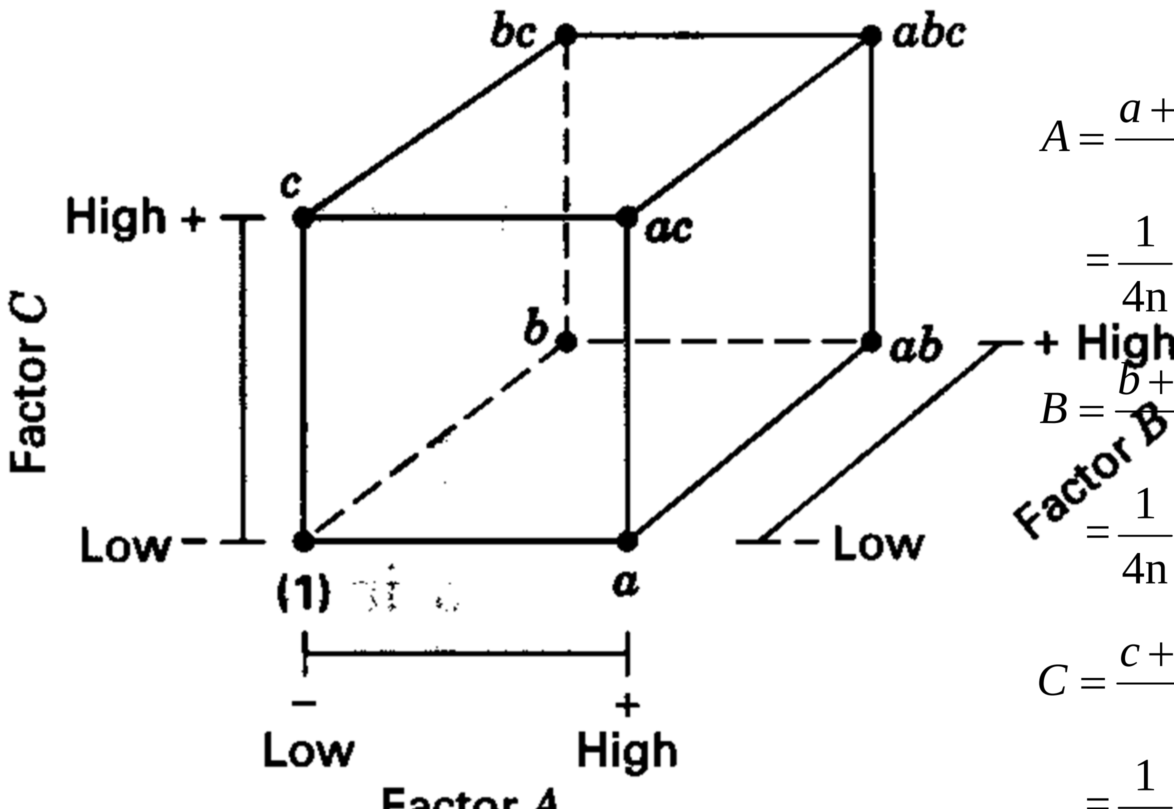


The Response Surface

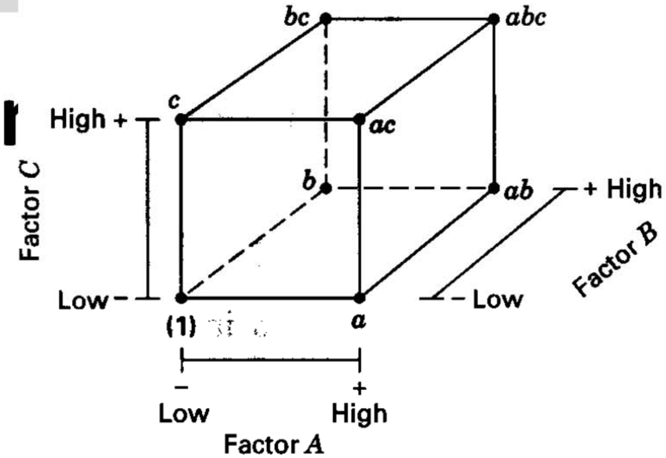


The 2^3 Factorial Design





Geometric presentation of contrasts corresponding to the main effects and interactions in the 2^3 design.



$$A = \frac{a+ab+ac+abc}{4n} - \frac{(1)+b+c+bc}{4n}$$

$$= \frac{1}{4\text{n}}[a+ab+ac+abc-(1)-b-c-bc]$$

$$B = \frac{b+ab+bc+abc}{4n} - \frac{(1)+a+c+ac}{4n}$$

$$= \frac{1}{4\text{n}}[b+ab+bc+abc-(1)-a-c-ac]$$

$$C = \frac{c+ac+bc+abc}{4n} - \frac{(1)+a+b+ab}{4n}$$

$$= \frac{1}{4\text{n}}[c+ac+bc+abc-(1)-a-b-ab]$$

$$A = \frac{1}{4n}[a + ab + ac + abc - (1) - b - c - bc]$$

Treatment Combo.	Factorial Effect							
	I	A	B	AB	C	AC	BC	ABC
(1)	+	-						
a	+	+						
b	+	-						
ab	+	+						
c	+	-						
ac	+	+						
bc	+	-						
abc	+	+						

$$B = \frac{1}{4n}[b + ab + bc + abc - (1) - a - c - ac]$$

Treatment	Factorial Effect							
Combo.	I	A	B	AB	C	AC	BC	ABC
(1)	+	-	-					
a	+	+	-					
b	+	-	+					
ab	+	+	+					
c	+	-	-					
ac	+	+	-					
bc	+	-	+					
abc	+	+	+					

$$C = \frac{1}{4n}[c + ac + bc + abc - (1) - a - b - ab]$$

Treatment	Factorial Effect							
Combo.	I	A	B	AB	C	AC	BC	ABC
(1)	+	-	-		-			
a	+	+	-		-			
b	+	-	+		-			
ab	+	+	+		-			
c	+	-	-		+			
ac	+	+	-		+			
bc	+	-	+		+			
abc	+	+	+		+			

$$AB = \frac{1}{4n}[abc + ab + c + (1) - bc - b - ac - a]$$

Treatment	Factorial Effect							
Combo.	I	A	B	AB	C	AC	BC	ABC
(1)	+	-	-	+	-			
a	+	+	-	-	-			
b	+	-	+	-	-			
ab	+	+	+	+	-			
c	+	-	-	+	+			
ac	+	+	-	-	+			
bc	+	-	+	-	+			
abc	+	+	+	+	+			

$$ABC = \frac{1}{4n}[-(1) + a + b - ab + c - ab - bc + abc]$$

Treatment	Factorial Effect							
Combo.	I	A	B	AB	C	AC	BC	ABC
(1)	+	-	-	+	-	+	+	-
a	+	+	-	-	-	-	+	+
b	+	-	+	-	-	+	-	+
ab	+	+	+	+	-	-	-	-
c	+	-	-	+	+	-	-	+
ac	+	+	-	-	+	+	-	-
bc	+	-	+	-	+	-	+	-
abc	+	+	+	+	+	+	+	+

Example: Plasma Etching Experiment

Run	Coded Factors A	B	C	Etch Rate Replicate 1	Replicate 2	Total
1	−1	−1	−1	550	604	(1) = 1154
2	1	−1	−1	669	650	a = 1319
3	−1	1	−1	633	601	b = 1234
4	1	1	−1	642	635	ab = 1277
5	−1	−1	1	1037	1052	c = 2089
6	1	−1	1	749	868	ac = 1617
7	−1	1	1	1075	1063	bc = 2178
8	1	1	1	729	860	abc = 1589

Factor Levels	Low (−1)	High (+1)
A (Gap, cm)	0.80	1.20
B (C_2F_6 flow, SCCM)	125	200
C (Power, W)	275	325

A = Gap between electrodes, B = Gas Flow, C = RF Power,
y = Etch Rate

Treatment Combo.	Factorial Effect							
	I	A	B	AB	C	AC	BC	ABC
(1) =1154	+	-	-	+	-	+	+	-
a = 1319	+	+	-	-	-	-	+	+
b =1234	+	-	+	-	-	+	-	+
ab = 1277	+	+	+	+	-	-	-	-
c = 2089	+	-	-	+	+	-	-	+
ac = 1617	+	+	-	-	+	+	-	-
bc = 2178	+	-	+	-	+	-	+	-
abc = 1589	+	+	+	+	+	+	+	+
Contrast		-813	59	2449	-199	-1229	-17	45

The General 2^k Factorial Design

- There will be k main effects, and

$$\binom{k}{2} \text{ two-factor interactions}$$

$$\binom{k}{3} \text{ three-factor interactions}$$

$$\vdots$$

$$1\ k-\text{factor interaction}$$

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

General Factorial Equations

$$Effect = \frac{2 * Contrast}{n * 2^k}$$

$$SS_{contrast} = \frac{Contrast^2}{n * 2^k}$$

Unreplicated 2^k Factorial Designs

- These are 2^k factorial designs with **one observation** at each corner of the "cube"
- An unreplicated 2^k factorial design is also sometimes called a "**single replicate**" of the 2^k
- These designs are very widely used
- Modeling "noise"?

Unreplicated 2^k Factorial Designs

- Lack of replication causes potential **problems** in statistical testing
  - Replication admits an estimate of "pure error" (a better phrase is an **internal estimate** of error)
  - With no replication, fitting the full model results in zero degrees of freedom for error
- Potential **solutions** to this problem
  - Pooling high-order interactions to estimate error
  - **Normal probability plotting** of effects (Daniels, 1959)
  - Other methods…

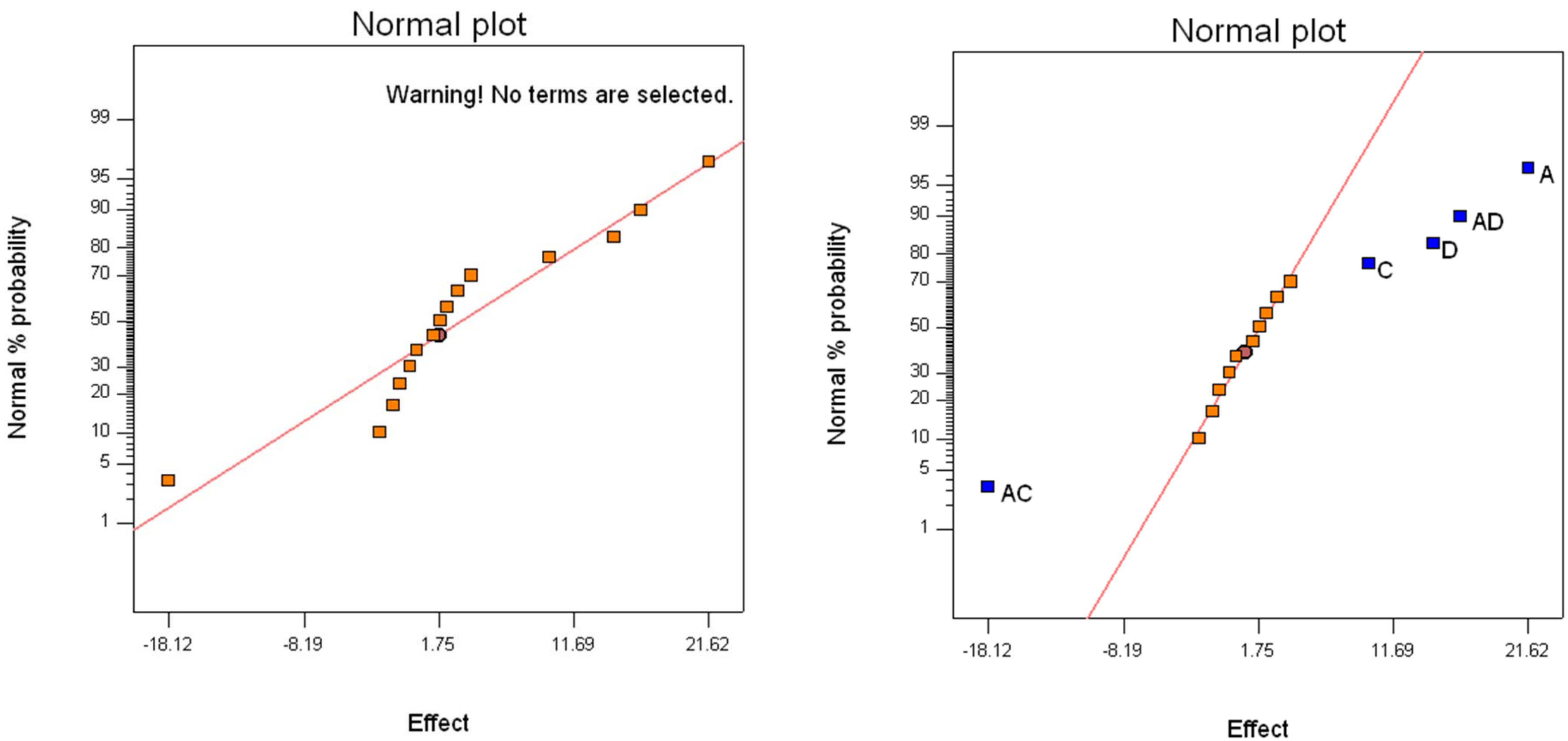
Example

Run Number	Factor A	Factor B	Factor C	Factor D	Run Label	Filtration Rate (gal/h)
1	−	−	−	−	(1)	45
2	+	−	−	−	*a*	71
3	−	+	−	−	*b*	48
4	+	+	−	−	*ab*	65
5	−	−	+	−	*c*	68
6	+	−	+	−	*ac*	60
7	−	+	+	−	*bc*	80
8	+	+	+	−	*abc*	65
9	−	−	−	+	*d*	43
10	+	−	−	+	*ad*	100
11	−	+	−	+	*bd*	45
12	+	+	−	+	*abd*	104
13	−	−	+	+	*cd*	75
14	+	−	+	+	*acd*	86
15	−	+	+	+	*bcd*	70
16	+	+	+	+	*abcd*	96

Estimates of the Effects

Term	Effect	SumSqr	% Contribtn
A	21.6	1870.6	32.6
B	3.1	39.1	0.7
C	9.9	390.1	6.8
D	14.6	855.6	14.9
AB	0.1	0.1	0.0
AC	-18.1	1314.1	22.9
AD	16.6	1105.6	19.3
BC	2.4	22.6	0.4
BD	-0.4	0.6	0.0
CD	-1.1	5.1	0.1
ABC	1.9	14.1	0.2
ABD	4.1	68.1	1.2
ACD	-1.6	10.6	0.2
BCD	-2.6	27.6	0.5
ABCD	1.4	7.6	0.1

Normal Probability Plot of Effects



Effects that lie along the line are negligible… large effects are far from the line

ANOVA Summary for the Model

Response: Filtration Rate
ANOVA for Selected Factorial Model
Analysis of variance table [Partial sum of squares]

Source	Sum of Squares	DF	Mean Square	F Value	Prob >F
Model	5535.81	5	1107.16	56.74	< 0.0001
A	*1870.56*	*1*	*1870.56*	*95.86*	*< 0.0001*
C	*390.06*	*1*	*390.06*	*19.99*	*0.0012*
D	*855.56*	*1*	*855.56*	*43.85*	*< 0.0001*
AC	*1314.06*	*1*	*1314.06*	*67.34*	*< 0.0001*
AD	*1105.56*	*1*	*1105.56*	*56.66*	*< 0.0001*
Residual	195.12	10	19.51		
Cor Total	5730.94	15			

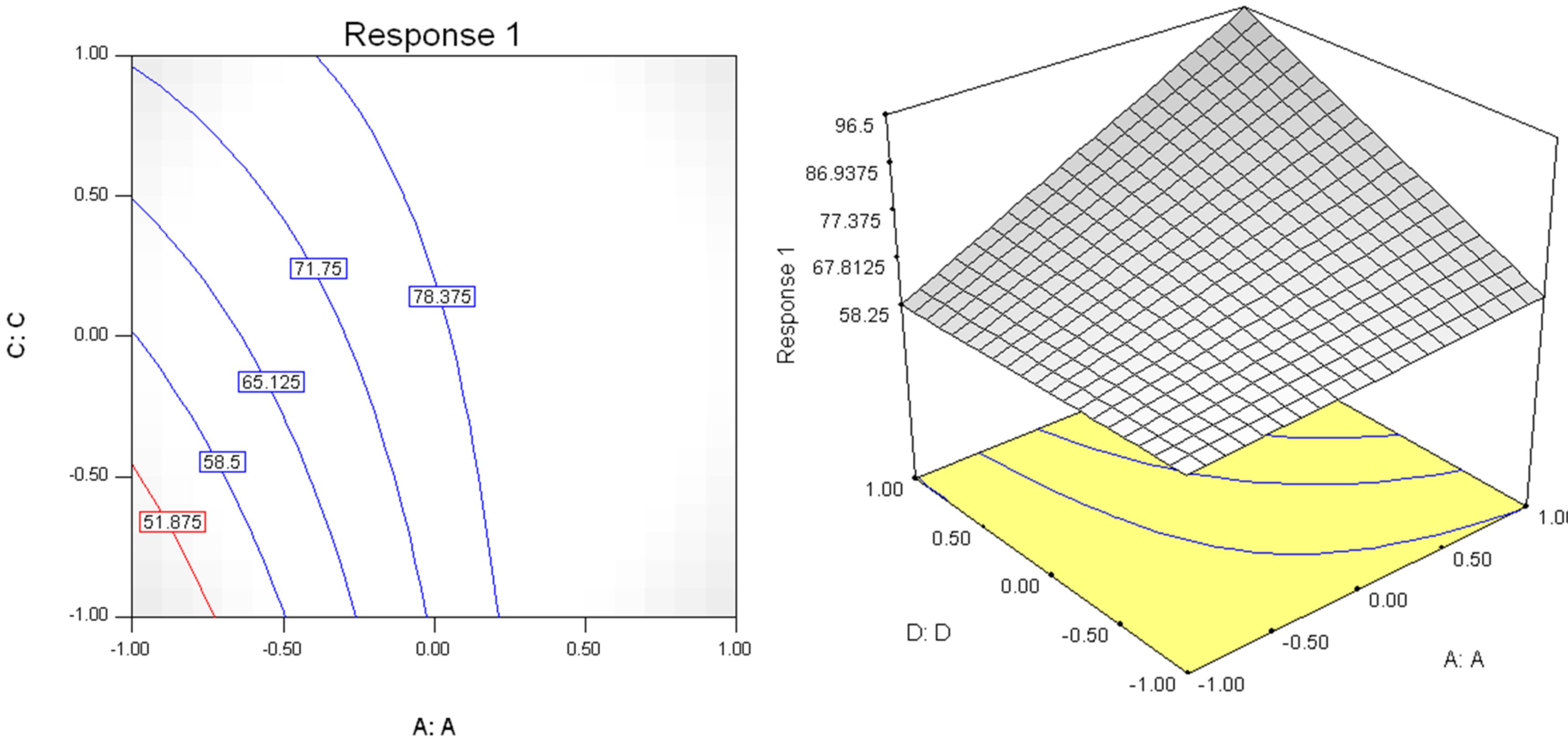
Std. Dev.	4.42	R-Squared	0.9660
Mean	70.06	Adj R-Squared	0.9489
		Pred R-Squared	0.9128
PRESS	499.52		

The Regression Model

Final Equation in Terms of Coded Factors:

Filtration Rate =
+70.06250
+10.81250 * Temperature
+4.93750 * Concentration
+7.31250 * Stirring Rate
-9.06250 * Temperature * Concentration
+8.31250 * Temperature * Stirring Rate

Model Interpretation – Response Surface Plots



With concentration at either the low or high level, high temperature and high stirring rate results in high filtration rates

Addition of Center Points to a 2^k Designs

- Based on the idea of replicating **some** of the runs in a factorial design
- Runs at the center provide an estimate of error and allow the experimenter to distinguish between two possible models:

$$\text{First-order model (interaction) } y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \sum_{j>i}^{k} \beta_{ij} x_i x_j + \varepsilon$$

$$\text{Second-order model } y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \sum_{j>i}^{k} \beta_{ij} x_i x_j + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \varepsilon$$

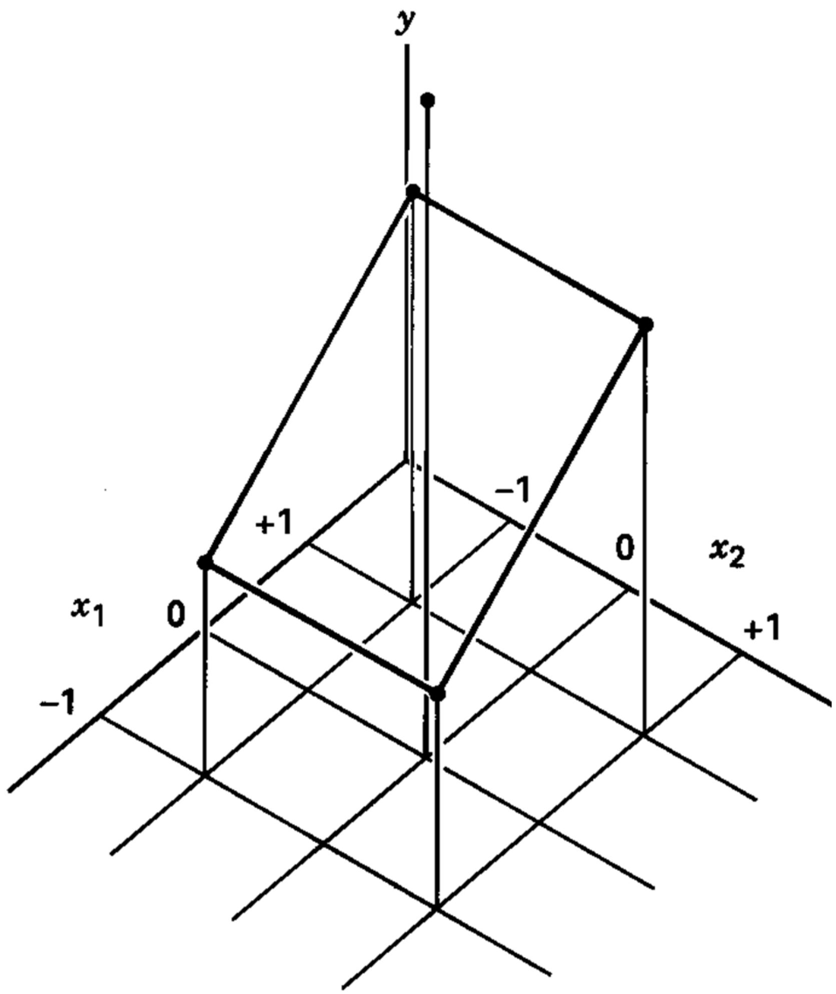


Figure 6-34 A 2^2 design with center points.

$$\bar{y}_F = \bar{y}_C \Rightarrow \text{ no "curvature"}$$

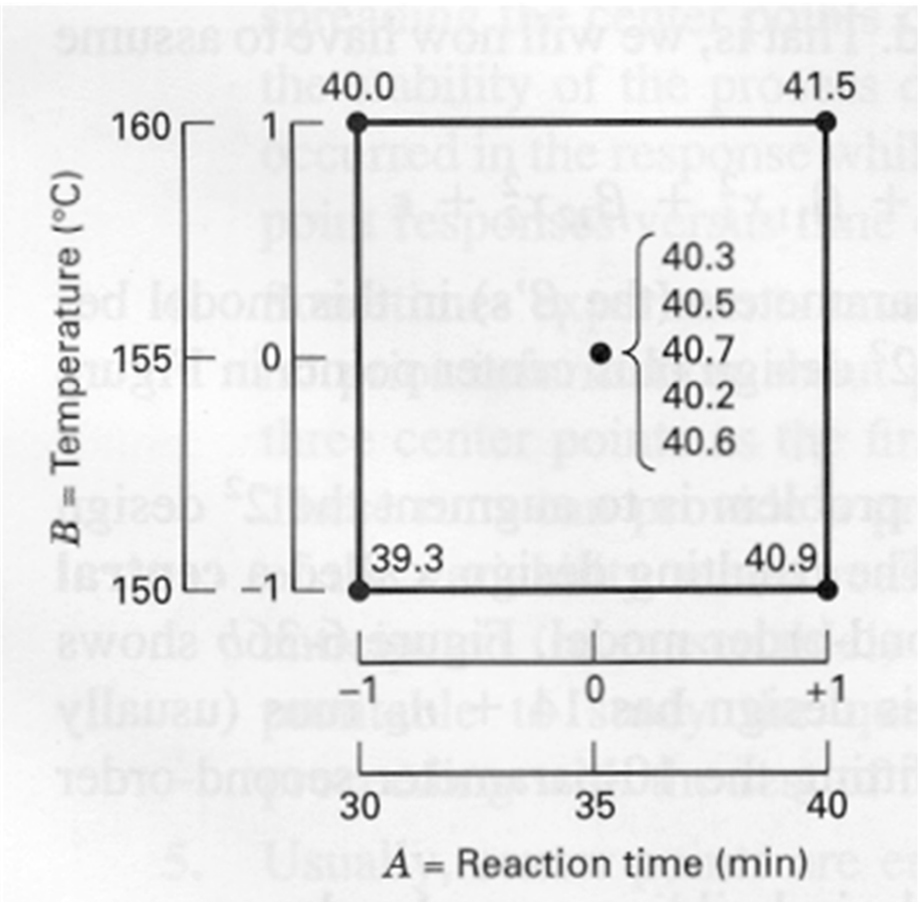
The hypotheses are:

$$H_0 : \sum_{i=1}^{k} \beta_{ii} = 0$$

$$H_1 : \sum_{i=1}^{k} \beta_{ii} \neq 0$$

$$SS_{\text{Pure Quad}} = \frac{n_F n_C (\bar{y}_F - \bar{y}_C)^2}{n_F + n_C}$$

Example



$$n_C = 5$$

Usually between 3 and 6 center points will work well

Design-Expert provides the analysis, including the *F*-test for pure quadratic curvature

ANOVA

Response: yield

ANOVA for Selected Factorial Model

Analysis of variance table [Partial sum of squares]

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F
Model	2.83	3	0.94	21.92	0.0060
A	*2.40*	*1*	*2.40*	*55.87*	*0.0017*
B	*0.42*	*1*	*0.42*	*9.83*	*0.0350*
AB	*2.500E-003*	*1*	*2.500E-003*	*0.058*	*0.8213*
Curvature	2.722E-003	1	2.722E-003	0.063	0.8137
Pure Error	0.17	4	0.043		
Cor Total	3.00	8			

Std. Dev.	0.21	R-Squared	0.9427
Mean	40.44	Adj R-Squared	0.8996
		Pred R-Squared	N/A
PRESS	N/A		

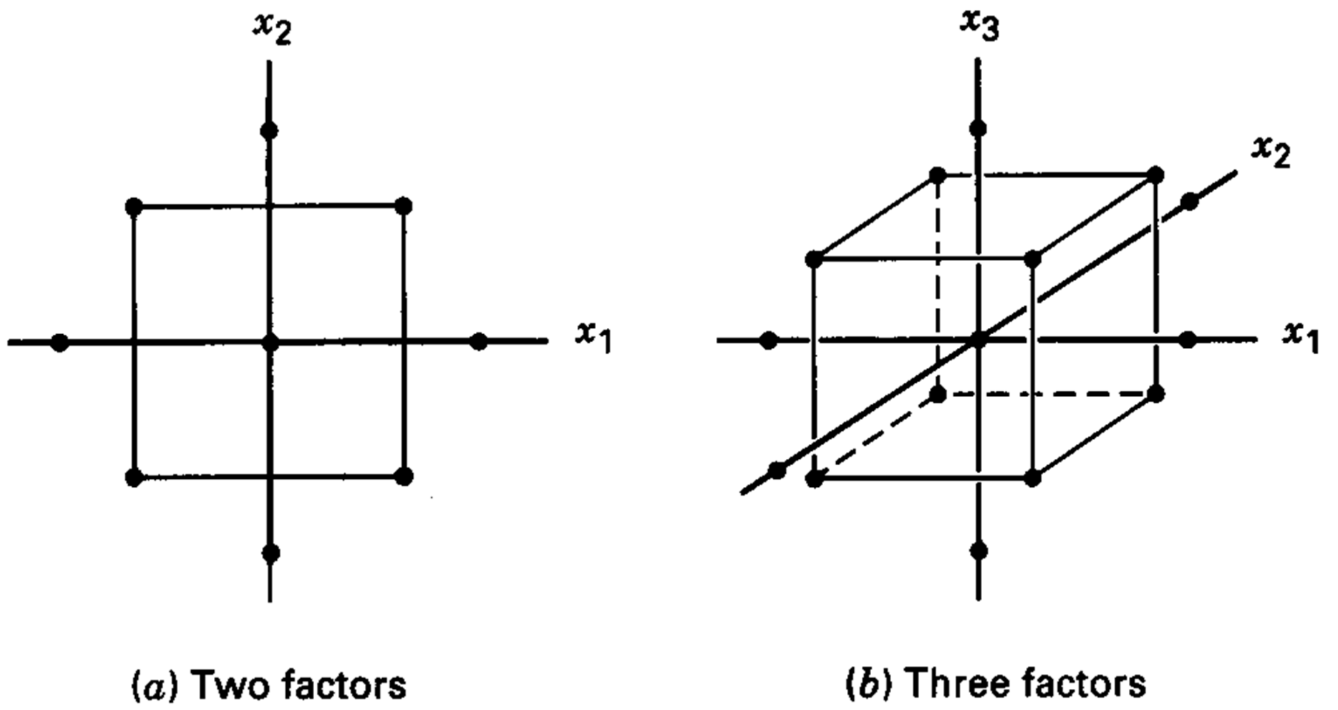
What if curvature was significant?

- Can we determine parameters for the second order model?

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2$$

- No – only have 5 independent runs and we need to estimate 6 parameters.
- Have to do something else.

If curvature is significant, augment the design with axial runs to create a central composite design. The CCD is a very effective design for fitting a second-order response surface model



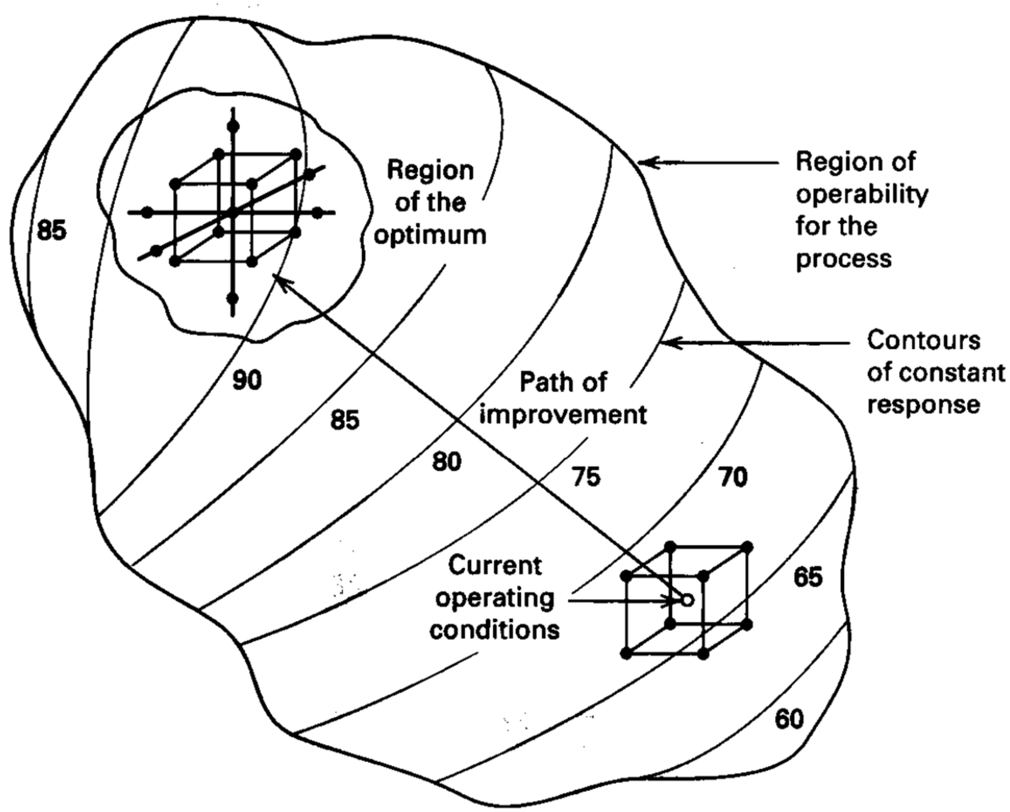
Central composite designs.

Overview of Response Surface Methods

- Primary focus of previous chapters is **factor screening**
  - Two-level factorials, fractional factorials are widely used
- Objective of **RSM** is optimization
- RSM dates from the 1950s; early applications in chemical industry

RSM is a Sequential Procedure

- Factor screening
- Finding the **region** of the optimum
- **Modeling & Optimization** of the response



The sequential nature of RSM.

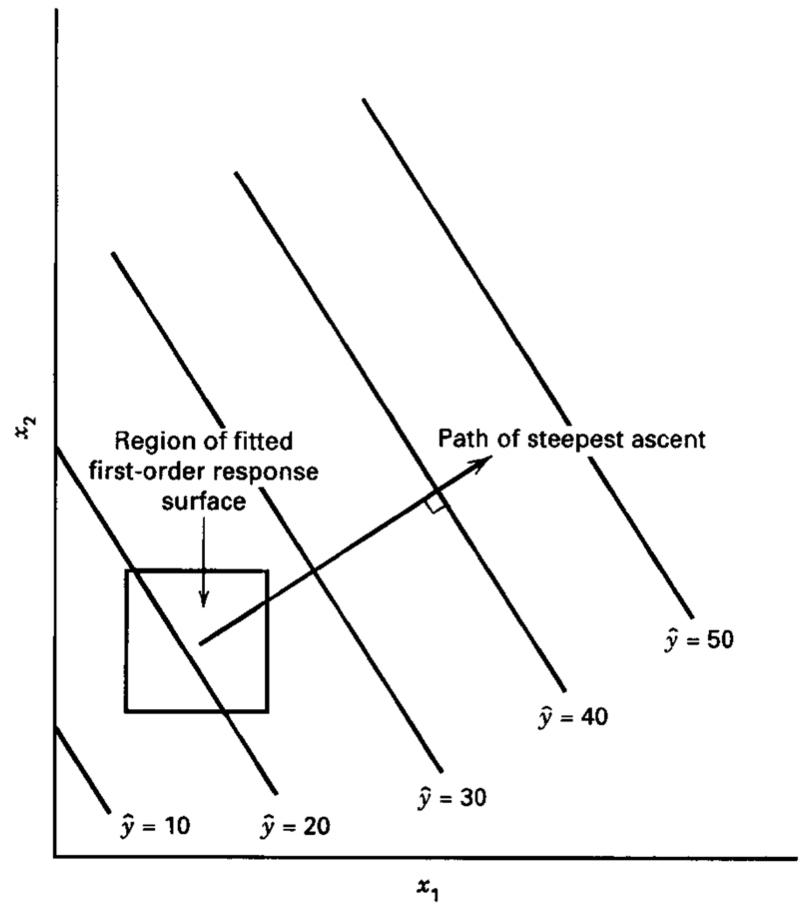
The Method of Steepest Ascent

A procedure for moving sequentially from an initial "guess" towards to region of the optimum

Based on the fitted first-order model

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$$

Steepest ascent is a gradient procedure



First-order response surface and path of steepest ascent.

An Example of Steepest Ascent

- Objective : Determine the operating conditions that maximize yield.
- Variables : Reaction Time & Temperature
- Current Settings: 35 minutes, and 155 F
  - Resultant yield = 40%
- Parameter settings:
  - Time (30, 40) minutes
  - Temp (150, 160) minutes
- Design: Full Factorial with 5 center points

An Example of Steepest Ascent
Initial Design

- Design: Full Factorial with 5 center points
  - Allows us to estimate error
  - Can check for interactions in the model
  - Can check for curvature
  - Economical – if more factors, what could we do?

An Example of Steepest Ascent
Experimental Data & Regression Model

			A:Time	B:Temp	yield
Std	Run	Block	minutes	degC	percent
1	7	{ 1 }	-1	-1	39.3
2	6	{ 1 }	1	-1	40.9
3	5	{ 1 }	-1	1	40
4	2	{ 1 }	1	1	41.5
5	9	{ 1 }	0	0	40.3
6	4	{ 1 }	0	0	40.5
7	1	{ 1 }	0	0	40.7
8	3	{ 1 }	0	0	40.2

[A]=2/4*(-39.3+40.9-40+41.5) = 1.55
[B]=2/4*(-39.3-40.9+40+41.5) = 0.65

$$\hat{y} = 40.44 + 0.775x_1 + 0.325x_2$$

An Example of Steepest Ascent
Experimental Data & Regression Model

$$\hat{y} = 40.44 + 0.775x_1 + 0.325x_2$$

$$X_1 = \frac{A - (30+40)/2}{(40-30)/2} = \frac{A-35}{5}$$

$$X_2 = \frac{B - (150+160)/2}{(160-150)/2} = \frac{B-155}{5}$$

$$\hat{y} = 40.44 + 0.775\left(\frac{A-35}{5}\right) + 0.325\left(\frac{B-155}{5}\right)$$

An Example of Steepest Ascent
Checking for Curvature

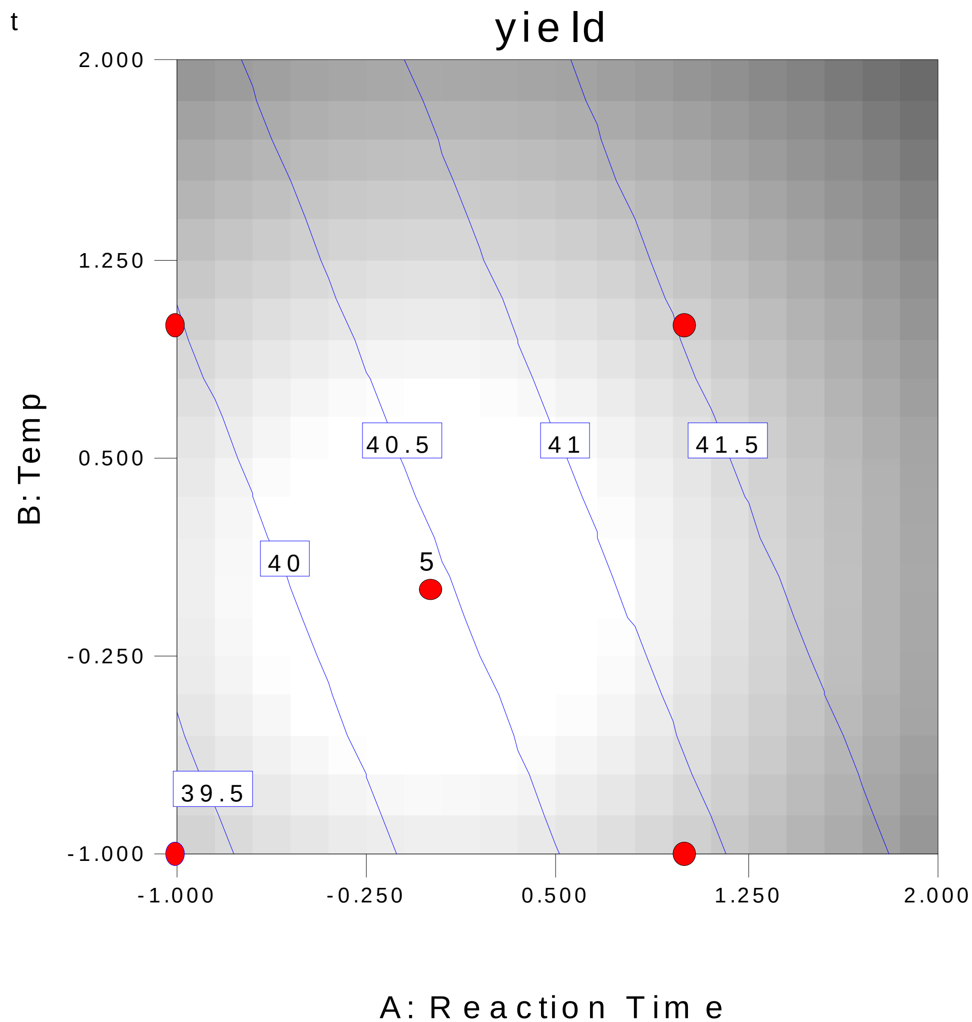
$$SS_{quad} = \frac{n_F n_C (\bar{y}_F - \bar{y}_C)^2}{n_F + n_C} = \frac{(4*5*-0.035)}{9} = 0.0027$$

$$MS_E = \frac{\sum_{centerpoints} (y_i - y)^2}{n_C - 1} = \frac{0.172}{4} = 0.43$$

No evidence of Curvature…

F would be well below 1 and well below $F_{1,4}$

An Example of Steepest Ascent



An approximate step size and path can be determined graphically

Formal methods can also be used

Types of experiments along the path:

- Single runs
- Replicated runs

Steepest Ascent

$$\hat{y}=40.44+0.775x_1+0.325x_2$$

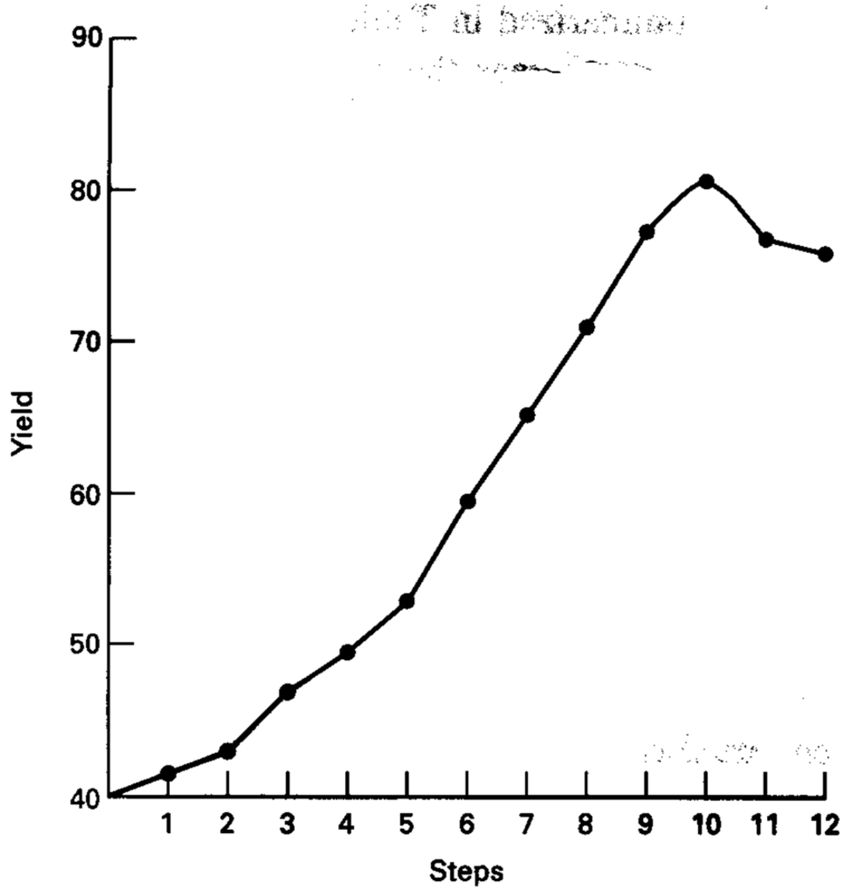
$$\text{Slope}=\frac{.325}{.775}=0.42$$

$$\text{StepSize}=5\text{ minutes }\ (x_1)$$

Steepest Ascent Experiment

Steps	Coded Variables x_1	Coded Variables x_2	Natural Variables ξ_1	Natural Variables ξ_2	Response y
Origin	0	0	35	155	
Δ	1.00	0.42	5	2	
Origin + Δ	1.00	0.42	40	157	41.0
Origin + 2Δ	2.00	0.84	45	159	42.9
Origin + 3Δ	3.00	1.26	50	161	47.1
Origin + 4Δ	4.00	1.68	55	163	49.7
Origin + 5Δ	5.00	2.10	60	165	53.8
Origin + 6Δ	6.00	2.52	65	167	59.9
Origin + 7Δ	7.00	2.94	70	169	65.0
Origin + 8Δ	8.00	3.36	75	171	70.4
Origin + 9Δ	9.00	3.78	80	173	77.6
Origin + 10Δ	10.00	4.20	85	175	80.3
Origin + 11Δ	11.00	4.62	90	179	76.2
Origin + 12Δ	12.00	5.04	95	181	75.1

Results from the Example



Yield versus steps along the path of steepest ascent

Local Optimum found at Step 10.

Next steps :
Run another 1st Order Model
at new conditions

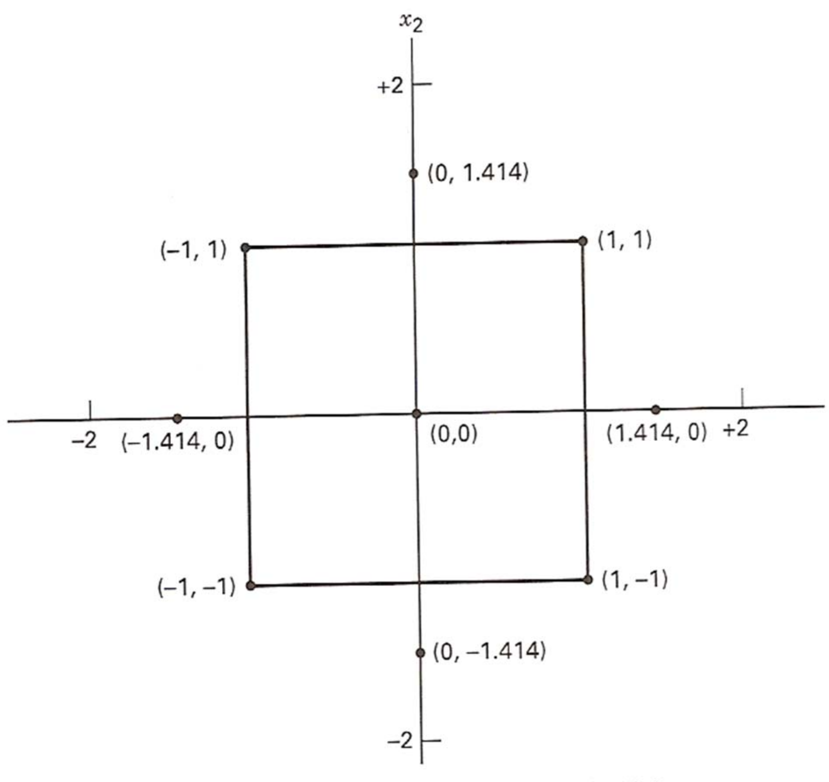
Augment if necessary so that
quadratic terms can be assessed

Analysis of the Second-Order Response Surface Model

			Factor 1	Factor 2	Response 1	Response 2	Response 3
Std	Run	Block	A:time	B:temp	yield	viscosity	Mn
			min	degF			
1	8	Block 1	80	170	76.5	62	2940
2	6	Block 1	90	170	78	66	3680
3	9	Block 1	80	180	77	60	3470
4	11	Block 1	90	180	79.5	59	3890
5	12	Block 1	77.93	175	75.6	71	3020
6	10	Block 1	92.07	175	78.4	68	3360
7	7	Block 1	85	167.93	77	57	3150
8	1	Block 1	85	182.07	78.5	58	3630
9	5	Block 1	85	175	79.9	72	3480
10	3	Block 1	85	175	80.3	69	3200
11	13	Block 1	85	175	80	68	3410
12	2	Block 1	85	175	79.7	70	3290
13	4	Block 1	85	175	79.8	71	3500

This is a central composite design

Analysis of the Second-Order Response Surface Model



central composite design

The Second-Order Response Surface Model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \varepsilon$$

- These models are used widely in practice
- The Taylor series analogy
- Fitting the model is easy, some nice designs are available
- Optimization is easy
- There is a lot of empirical evidence that they work very well

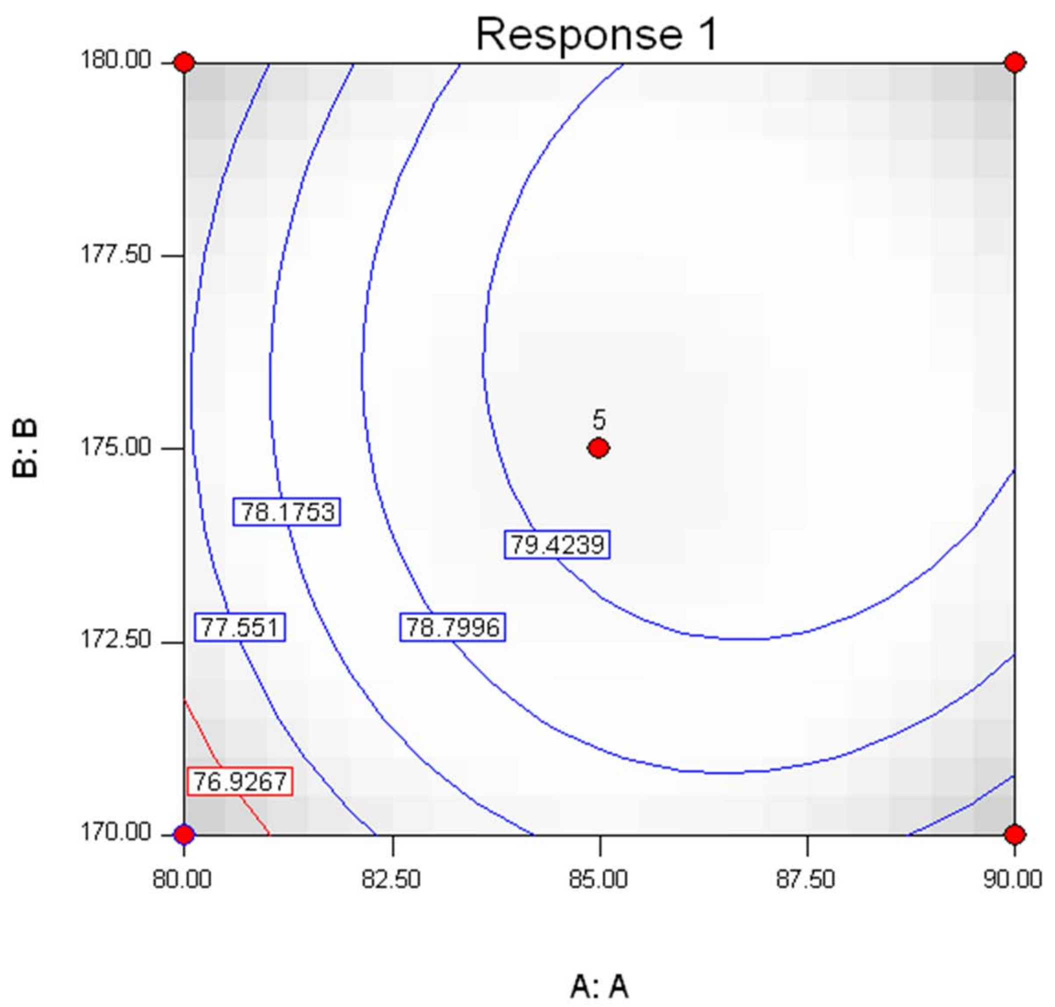
ANOVA

ANOVA for Response Surface Quadratic Model
Analysis of variance table [Partial sum of squares]

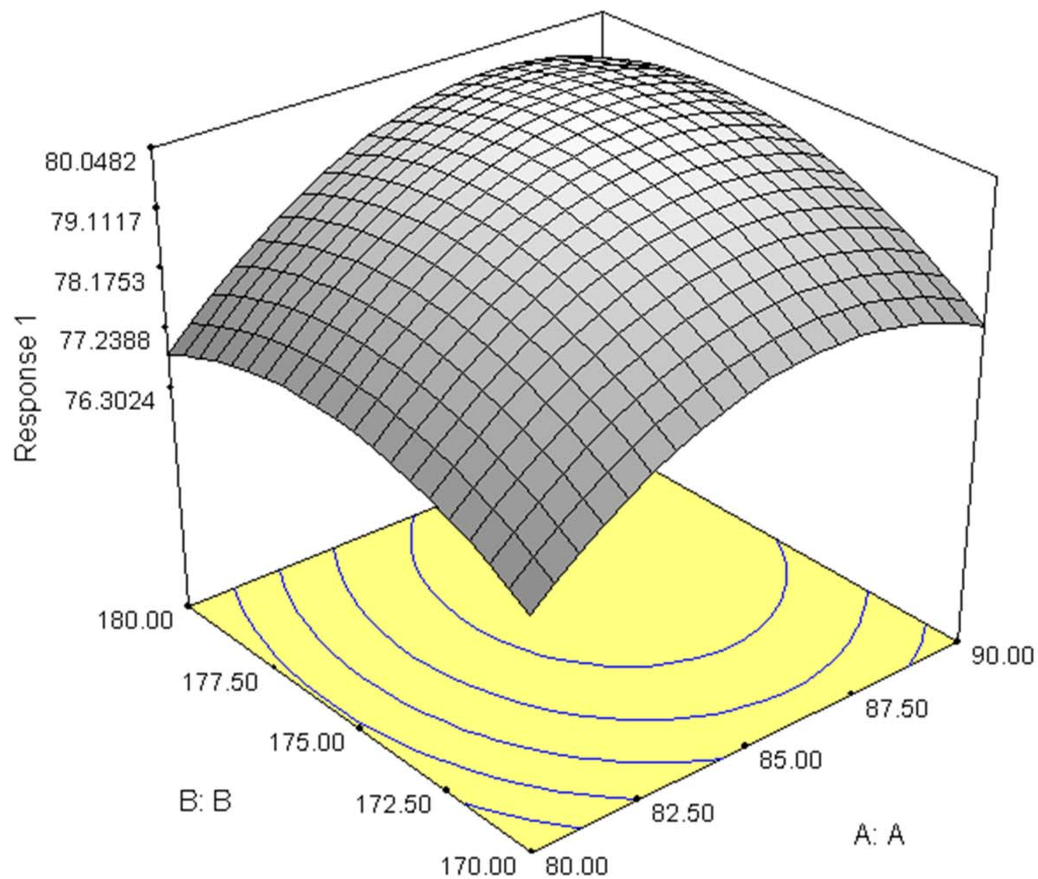
Source	Sum of Squares	DF	Mean Square	F Value	Prob > F
Model	28.25	5	5.65	79.85	< 0.0001
A	*7.92*	*1*	*7.92*	*111.93*	*< 0.0001*
B	*2.12*	*1*	*2.12*	*30.01*	*0.0009*
A^2	*13.18*	*1*	*13.18*	*186.22*	*< 0.0001*
B^2	*6.97*	*1*	*6.97*	*98.56*	*< 0.0001*
AB	*0.25*	*1*	*0.25*	*3.53*	*0.1022*
Residual	0.50	7	0.071		
Lack of Fit	*0.28*	*3*	*0.094*	*1.78*	*0.2897*
Pure Error	*0.21*	*4*	*0.053*		
Cor Total	28.74	12			

$$\hat{y} = 79.94 + 0.99x_1 + 0.52x_2 + 0.25x_1x_2 - 1.38x_1^2 - 1.00x_2^2$$

Contour Plots



Contour Plots



These plots are given in the natural variables

The optimum is at about 87 minutes and 176.5 degrees

Formal optimization methods can also be used (particularly when $k > 2$)

Designs for Fitting Response Surface Models

- For the first-order model, two-level factorials (and fractional factorials) augmented with **center points** are appropriate choices

- The **central composite design** is the most widely used design for fitting the second-order model

- Selection of a second-order design is an interesting problem

- There are numerous excellent second-order designs available