Statistical Design of Experiments Applied to Organic Synthesis

Luis Sanchez
Michigan State University
October 11th, 2006

• Statistical Design of Experiments

DoE

- Methodology developed in 1958 by the British statistician Ronald Fisher
- Strategy
  - Appropriate statistical analysis <u>before</u> any experimental data are obtained
- Objective
  - To get as much information as possible from a minimum number of experiments

Bayne, C. K.; Rubin, I. B., *Practical experimental designs and optimization methods for chemists*. VCH Publishers, USA, **1986**.
Tranter, R., *Design and analysis in chemical research*. Sheffield Academic; CRC Press: Sheffield, England, **2000**.

Experimentation in Organic synthesis

- In any synthetical procedure there are factors

  temperature, time, pressure, reagents, rate of addition, catalyst, solvent, concentration, pH

  that will have an influence on the result

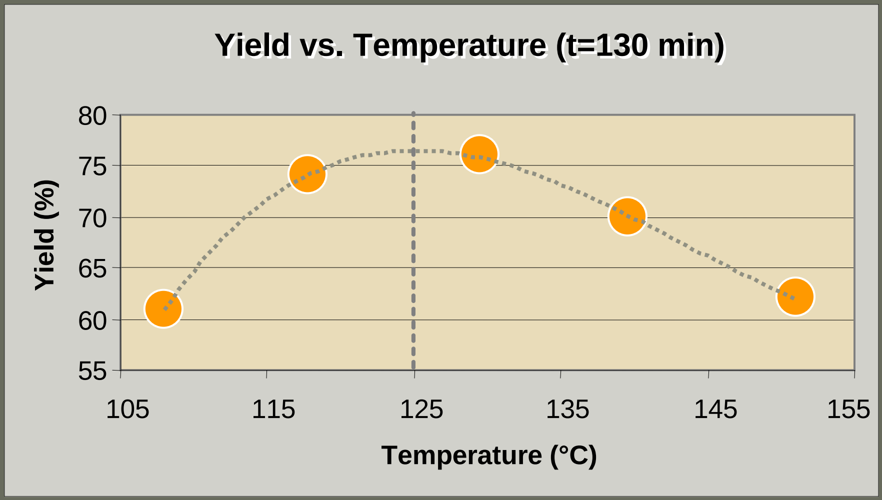
  yield, purity, selectivity

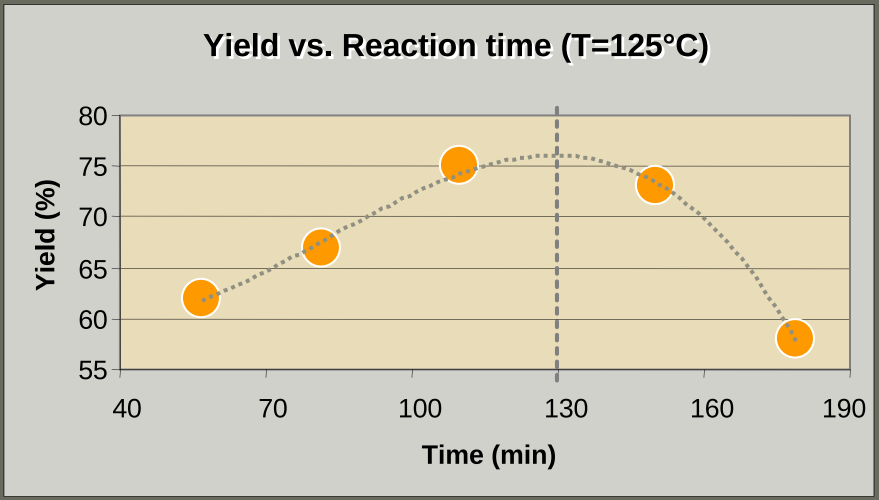
Carlson, R., *Design and optimization in organic synthesis*. Elsevier: Amsterdam ; New York, **1992**.

Conventional approach to optimization

$$X + Y \xrightarrow[t\ \text{minutes}]{T\ ^{\circ}C} Z$$

- Analysis of the reaction conditions that affect the yield:



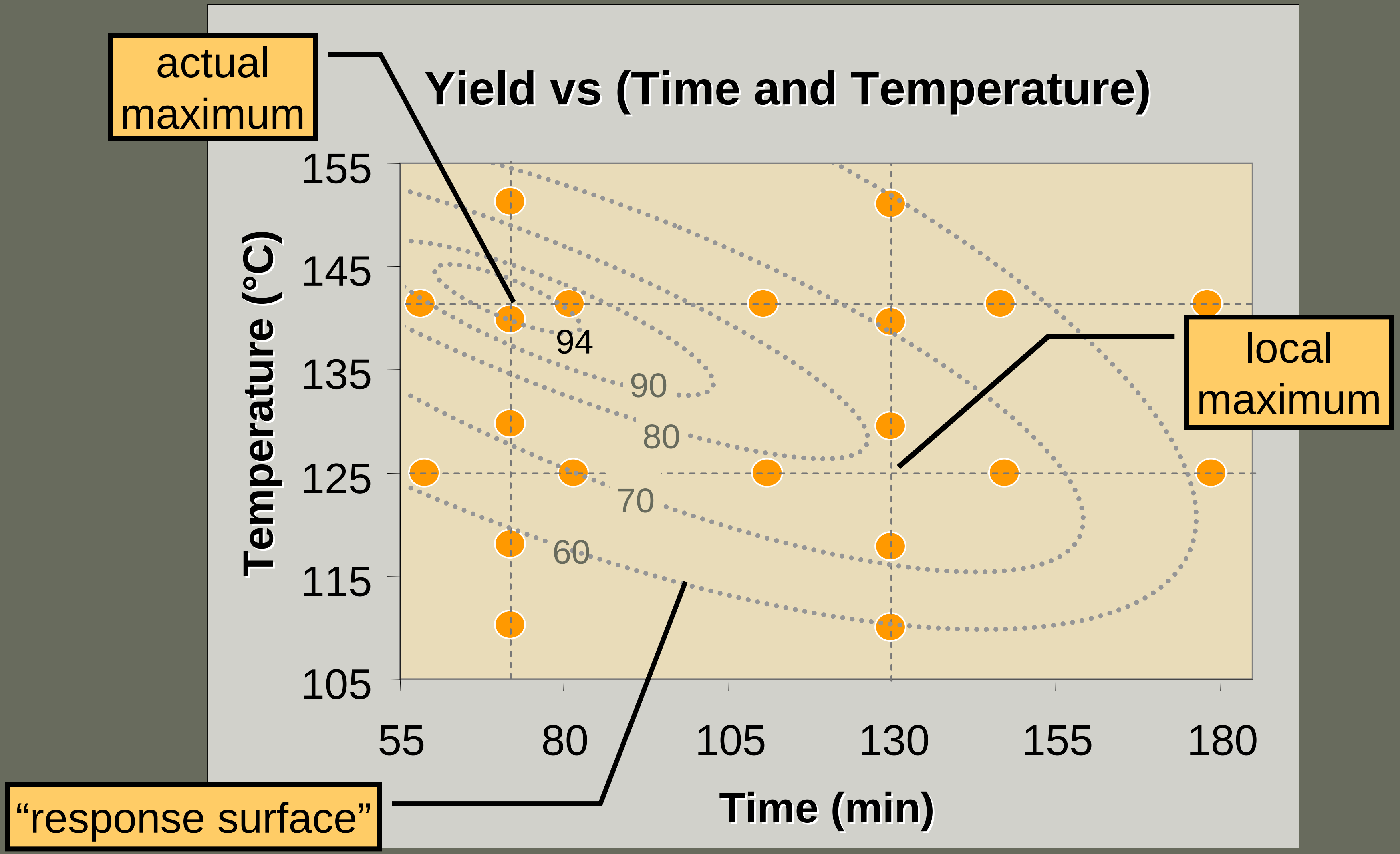


- The maximum yield would be obtained at 125 °C in 130 min?

  Are these really the optimum conditions?

Tranter, R., *Design and analysis in chemical research*. Sheffield Academic; CRC Press: Sheffield, England, **2000**.

- How yield actually behaves

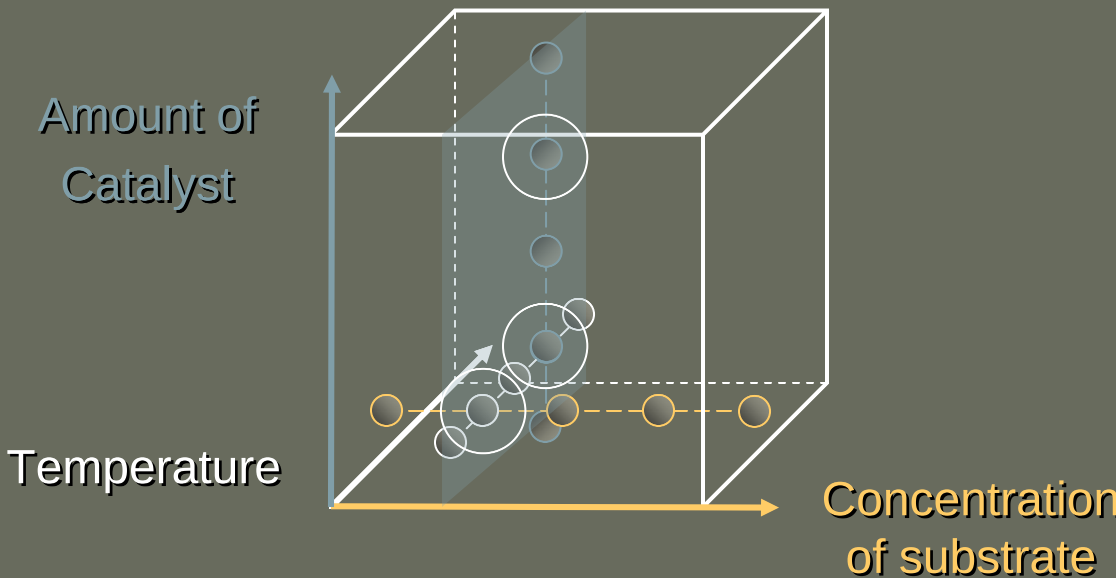


Carlson, R., *Design and optimization in organic synthesis*. Elsevier: Amsterdam ; New York, **1992**.
Tranter, R., *Design and analysis in chemical research*. Sheffield Academic; CRC Press: Sheffield, England, **2000**.

The conventional approach

- Analysis of the effect of one particular reaction condition by keeping all the other ones constant

$$A + B \xrightarrow[\text{T °C} \\ \text{t minutes}]{\text{catalyst}} C$$



The problem:

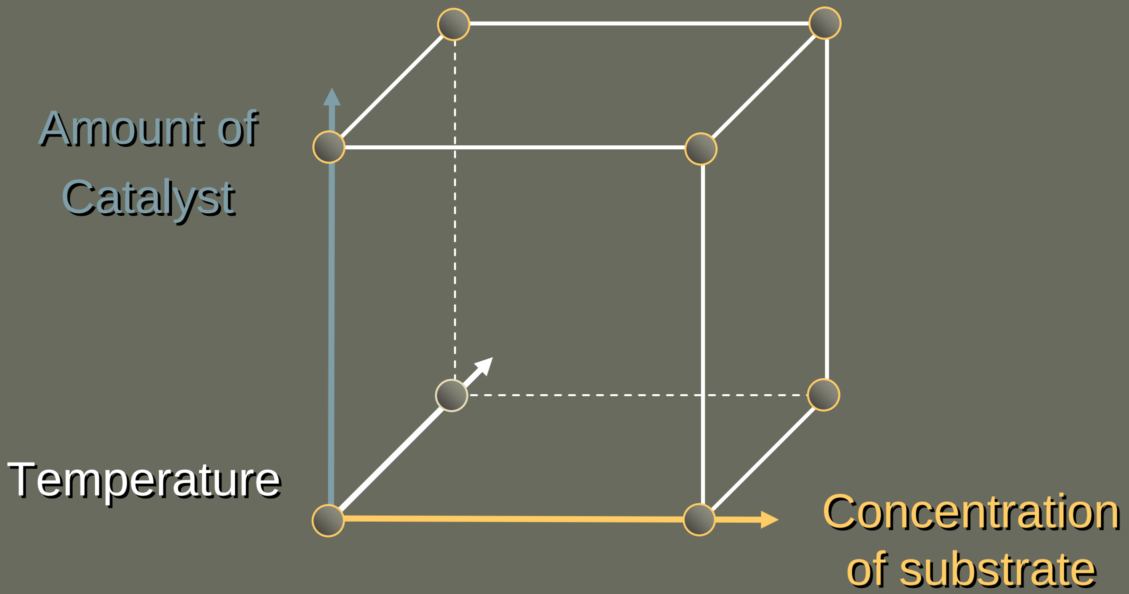
- The optimum conditions obtained depend on the starting point

Owen, M. R.; Luscombe, C.; Lai, L. W.; Godbert, S.; Crookes, D. L.; Emiabata-Smith, D.
Org. Proc. Res. Dev. **2001,** 5, 308-323.

• The DoE approach

- To rationally choose points throughout the cube to fully represent the entire space.

$$A + B \xrightarrow[\text{T °C, t minutes}]{\text{catalyst}} C$$



Owen, M. R.; Luscombe, C.; Lai, L. W.; Godbert, S.; Crookes, D. L.; Emiabata-Smith, D.
Org. Proc. Res. Dev. **2001,** 5, 308-323.

Outline

- Determining important reaction conditions
  - Fractional factorial design
- Analysis of reaction condition effects
  - Factorial design
- Estimation of the optimum conditions
  - Response surface analysis

Factorial designs

- Two types of reaction conditions:
  - Numeric

    temperature, pH, rate of addition, concentration
  - Categoric

    solvent, inert atmosphere, presence of molecular sieves, use of a particular reagent

- Each reaction condition will be screened over a defined set of values (numeric) or options (categoric)

- Experiments are run using all the possible combinations

• m^n Factorial designs

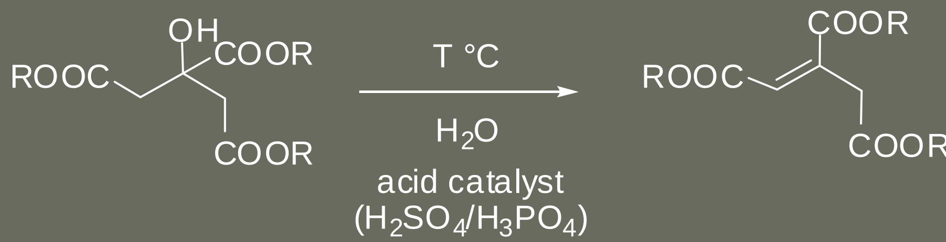
number of values for each reaction condition

$$m^n$$

number of reaction conditions

- If we analyze 2 values (or options) for 3 reaction conditions, $2^3=8$ experiments need to be run
- A m^n factorial design requires m^n experiments
- The most used method is 2^n design

• 2^3 factorial design



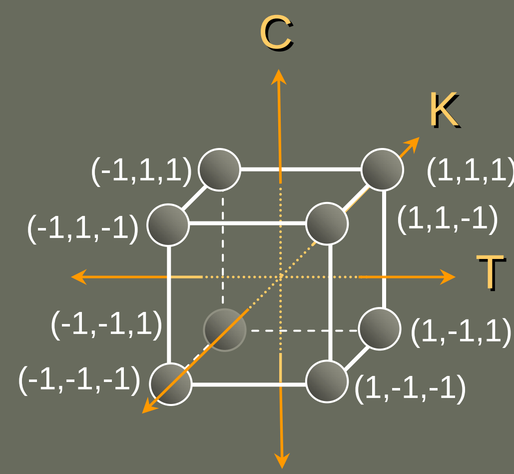
number of values

2^3

number of conditions

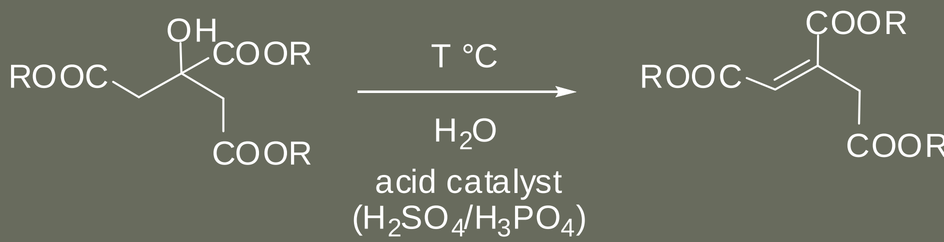
- 2 values (or options) for 3 reaction conditions:

T Temperature (°C)		C Concentration (M)		K Catalyst	
120	160	1.5	2.5	H_3PO_4	H_2SO_4



Box, G. E. P.; Hunter, W. G.; Hunter, J. S., *Statistics for experimenters : an introduction to design, data analysis, and model building*. Wiley: New York, **1978**.

2^3 factorial design



- 8 experimental runs:

run	T	C	K	label	yield (%)
1	-	-	-	1	60
2	+	-	-	t	72
3	-	+	-	c	54
4	+	+	-	tc	68
5	-	-	+	k	52
6	+	-	+	tk	83
7	-	+	+	ck	45
8	+	+	+	tck	80

Box, G. E. P.; Hunter, W. G.; Hunter, J. S., *Statistics for experimenters : an introduction to design, data analysis, and model building*. Wiley: New York, **1978**.

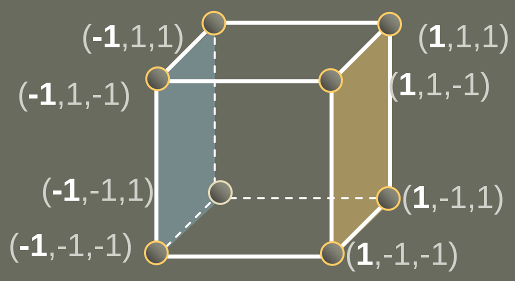
run	T	C	K	label	yield (%)
1	-	-	-	1	60
2	+	-	-	t	72
3	-	+	-	c	54
4	+	+	-	tc	68
5	-	-	+	k	52
6	+	-	+	tk	83
7	-	+	+	ck	45
8	+	+	+	tck	80

Box, G. E. P.; Hunter, W. G.; Hunter, J. S., *Statistics for experimenters : an introduction to design, data analysis, and model building*. Wiley: New York, **1978**.

• Measuring the effect: Temperature

run	T	C	K	label	yield (%)	
1	-	-	-	1	60	12
2	+	-	-	t	72	
3	-	+	-	c	54	14
4	+	+	-	tc	68	
5	-	-	+	k	52	31
6	+	-	+	tk	83	
7	-	+	+	ck	45	35
8	+	+	+	tck	80	

Effect of T



One half of the average of the differences of each pair

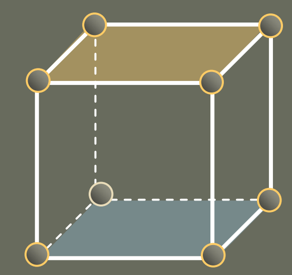
$$= \frac{\left[\frac{(t-1)+(tc-c)+(tk-k)+(tck-ck)}{4}\right]}{2} = \frac{\left[\frac{12+14+31+35}{4}\right]}{2} = 11.5$$

Box, G. E. P.; Hunter, W. G.; Hunter, J. S., *Statistics for experimenters : an introduction to design, data analysis, and model building*. Wiley: New York, **1978**.

Measuring the effect: Concentration

run	T	C	K	label	yield (%)	
1	-	-	-	1	60	-6
2	+	-	-	t	72	
3	-	+	-	c	54	-4
4	+	+	-	tc	68	
5	-	-	+	k	52	-7
6	+	-	+	tk	83	
7	-	+	+	ck	45	-3
8	+	+	+	tck	80	

Effect of C



<u>One half</u> of the average of the differences of each pair

$$= \frac{\left[\dfrac{(c-1)+(tc-t)+(ck-k)+(tck-tk)}{4}\right]}{2} = \frac{\left[\dfrac{(-6)+(-4)+(-7)+(-3)}{4}\right]}{2} = -2.5$$

Box, G. E. P.; Hunter, W. G.; Hunter, J. S., *Statistics for experimenters : an introduction to design, data analysis, and model building*. Wiley: New York, **1978**.

Measuring the effect: Catalyst

run	T	C	K	label	yield (%)
1	-	-	-	1	60
2	+	-	-	t	72
3	-	+	-	c	54
4	+	+	-	tc	68
5	-	-	+	k	52
6	+	-	+	tk	83
7	-	+	+	ck	45
8	+	+	+	tck	80

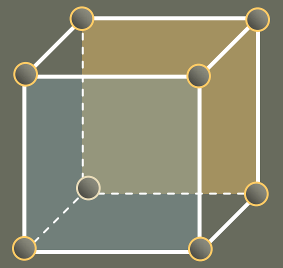
-8

11

-9

12

Effect of K



One half of the average of the differences of each pair

$$= \frac{\left[\frac{(k-1)+(tk-t)+(ck-c)+(tck-tc)}{4}\right]}{2} = \frac{\left[\frac{(-8)+11+(-9)+12}{4}\right]}{2} = 0.75$$

Box, G. E. P.; Hunter, W. G.; Hunter, J. S., *Statistics for experimenters : an introduction to design, data analysis, and model building*. Wiley: New York, **1978**.

Concentration-temperature interaction

run	T	C	K	label	yield (%)
1	-	-	-	1	60
2	+	-	-	t	72
3	-	+	-	c	54
4	+	+	-	tc	68
5	-	-	+	k	52
6	+	-	+	tk	83
7	-	+	+	ck	45
8	+	+	+	tck	80

$12 \rightarrow 6$

$14 \rightarrow 7$

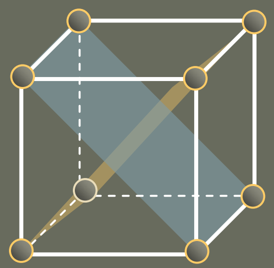
1

$31 \rightarrow 15.5$

$35 \rightarrow 17.5$

2

Effect of C on the effect of T



One half of the average of the differences of each pair of effects

$$\text{on} = \frac{\left\{\left[\frac{(tc-c)}{2}-\frac{(t-1)}{2}\right]+\left[\frac{(tck-ck)}{2}-\frac{(tk-k)}{2}\right]\right\}\Big/2}{2} = \frac{\left\{\left[\frac{14}{2}-\frac{12}{2}\right]+\left[\frac{35}{2}-\frac{31}{2}\right]\right\}\Big/2}{2} = 0.75$$

Box, G. E. P.; Hunter, W. G.; Hunter, J. S., *Statistics for experimenters : an introduction to design, data analysis, and model building*. Wiley: New York, **1978**.

Temperature-concentration interaction

run	T	C	K	label	yield (%)
1	-	-	-	1	60
2	+	-	-	t	72
3	-	+	-	c	54
4	+	+	-	tc	68
5	-	-	+	k	52
6	+	-	+	tk	83
7	-	+	+	ck	45
8	+	+	+	tck	80

-6 → -3

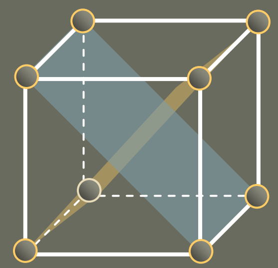
-4 → -2

1

-7 → -3.5

-3 → -1.5

Effect of T on the effect of C



2 One half of the average of the differences of each pair of effects

$$\text{on} = \frac{\left\{\left[\frac{(tc-t)}{2}-\frac{(c-1)}{2}\right]+\left[\frac{(tck-tk)}{2}-\frac{(ck-k)}{2}\right]\right\}\Big/2}{2} = \frac{\left\{\left[\frac{(-4)}{2}-\frac{(-6)}{2}\right]+\left[\frac{(-3)}{2}-\frac{(-7)}{2}\right]\right\}\Big/2}{2} = 0.75$$

Box, G. E. P.; Hunter, W. G.; Hunter, J. S., *Statistics for experimenters : an introduction to design, data analysis, and model building*. Wiley: New York, **1978**.

• Concentration-temperature interaction

run	T	C	K	label	yield (%)
1	-	-	-	1	60
2	+	-	-	t	72
3	-	+	-	c	54
4	+	+	-	tc	68
5	-	-	+	k	52
6	+	-	+	tk	83
7	-	+	+	ck	45
8	+	+	+	tck	80

12 → 6

14 → 7

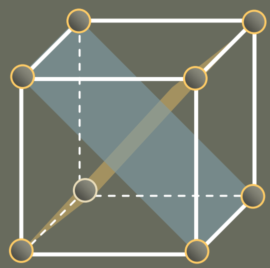
1

31 → 15.5

35 → 17.5

2

Effect of C on the effect of T



<u>One half</u> of the average of the differences of each pair of effects

$$\text{on} = \frac{\left\{\left[\frac{(tc-c)}{2}-\frac{(t-1)}{2}\right]+\left[\frac{(tck-ck)}{2}-\frac{(tk-k)}{2}\right]\right\}\Big/2}{2} = \frac{\left\{\left[\frac{14}{2}-\frac{12}{2}\right]+\left[\frac{35}{2}-\frac{31}{2}\right]\right\}\Big/2}{2} = 0.75$$

Box, G. E. P.; Hunter, W. G.; Hunter, J. S., *Statistics for experimenters : an introduction to design, data analysis, and model building*. Wiley: New York, **1978**.

- Temperature-catalyst interaction

run	T	C	K	label	yield (%)
1	-	-	-	1	60
2	+	-	-	t	72
3	-	+	-	c	54
4	+	+	-	tc	68
5	-	-	+	k	52
6	+	-	+	tk	83
7	-	+	+	ck	45
8	+	+	+	tck	80

12 → 6

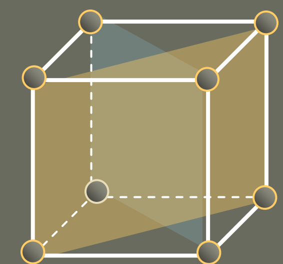
14 → 7

31 → 15.5

35 → 17.5

9.5

10.5



One half of the average of the differences of each pair of effects

$$\text{on} = \frac{\left\{\left[\frac{(tk-k)}{2}-\frac{(t-1)}{2}\right]+\left[\frac{(tck-ck)}{2}-\frac{(tc-c)}{2}\right]\right\}/2}{2} = \frac{\left\{\left[\frac{31}{2}-\frac{12}{2}\right]+\left[\frac{35}{2}-\frac{14}{2}\right]\right\}/2}{2} = 5$$

Box, G. E. P.; Hunter, W. G.; Hunter, J. S., *Statistics for experimenters : an introduction to design, data analysis, and model building*. Wiley: New York, **1978**.

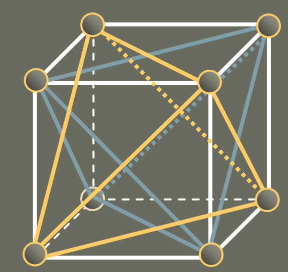
• TCK interaction

run	T	C	K	label	yield (%)
1	-	-	-	1	60
2	+	-	-	t	72
3	-	+	-	c	54
4	+	+	-	tc	68
5	-	-	+	k	52
6	+	-	+	tk	83
7	-	+	+	ck	45
8	+	+	+	tck	80

12 → 6
14 → 7
31 → 15.5
35 → 17.5

9.5 → 4.75
10.5 → 5.25

0.5



on $$= \frac{\left\{\left[\frac{(tck-ck)}{2}-\frac{(tc-c)}{2}\right]-\left[\frac{(tk-k)}{2}-\frac{(t-1)}{2}\right]\right\}/2}{2} = \frac{\left\{\left[\frac{35}{2}-\frac{14}{2}\right]-\left[\frac{31}{2}-\frac{12}{2}\right]\right\}/2}{2} = 0.25$$

Box, G. E. P.; Hunter, W. G.; Hunter, J. S., *Statistics for experimenters : an introduction to design, data analysis, and model building*. Wiley: New York, **1978**.

• Measuring the effect and interactions

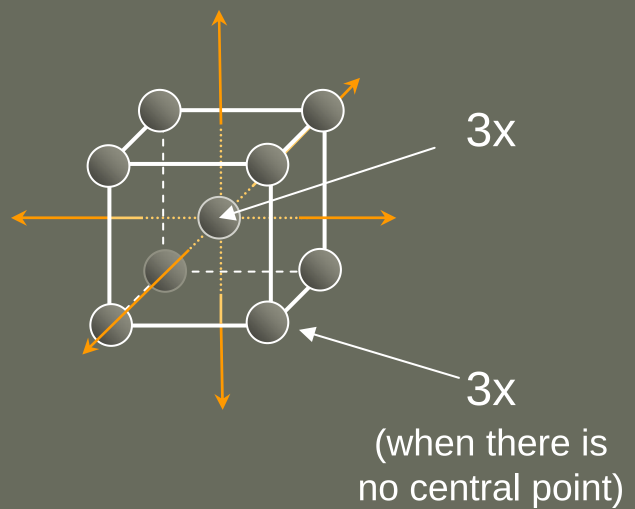
- Yates's algorithm: works for any 2^n factorial design

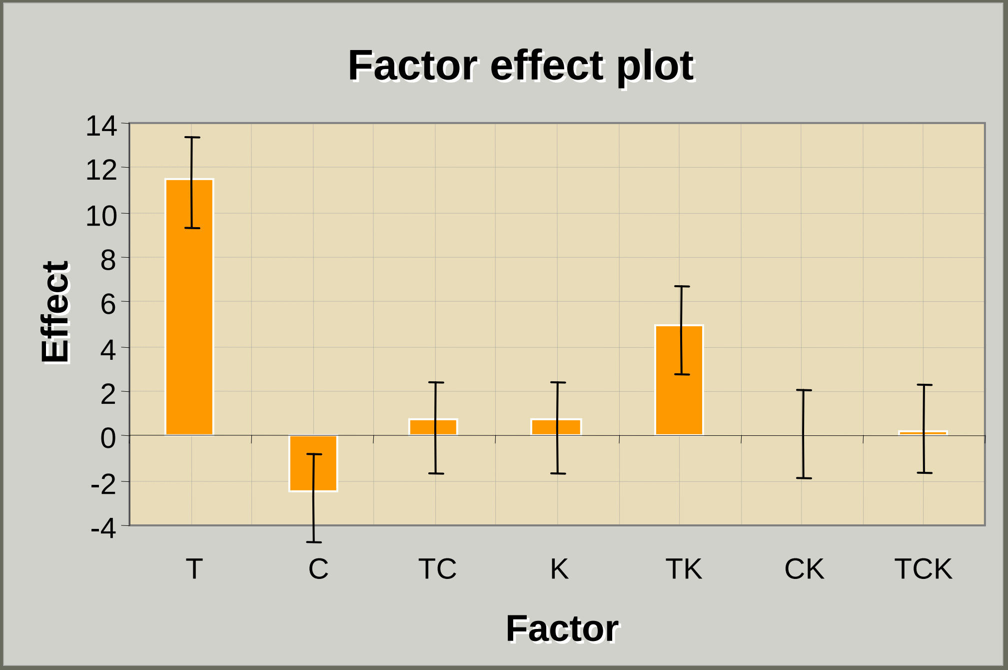
run	T	C	K	label	yield (%)	(1)	(2)	(3)	div	result	
1	-	-	-	1	60	132	254	514	8	64.25	average
2	+	-	-	t	72	122	260	92	8	11.5	*T*
3	-	+	-	c	54	135	26	-20	8	-2.5	*C*
4	+	+	-	tc	68	125	66	6	8	0.75	*TC*
5	-	-	+	k	52	12	-10	6	8	0.75	*K*
6	+	-	+	tk	83	14	-10	40	8	5.0	*TK*
7	-	+	+	ck	45	31	2	0	8	0	*CK*
8	+	+	+	tck	80	35	4	2	8	0.25	*TCK*

Box, G. E. P.; Hunter, W. G.; Hunter, J. S., *Statistics for experimenters : an introduction to design, data analysis, and model building*. Wiley: New York, **1978**.

• What do those numbers mean?

- First we need to evaluate if they are significant





- If the effect of a factor is lower than the standard deviation, it’s likely to be due to experimental error

- # What do those numbers mean?

  - The effects can be used to calculate a function that represents all the experimental runs

	result
average	64.25
T	11.5
C	-2.5
TC	0.75
K	0.75
TK	5.0
CK	0
TCK	0.25

$$\text{yield} = 64.25 + 11.5\text{T} - 2.5\text{C} + 5\text{TK} \pm$$

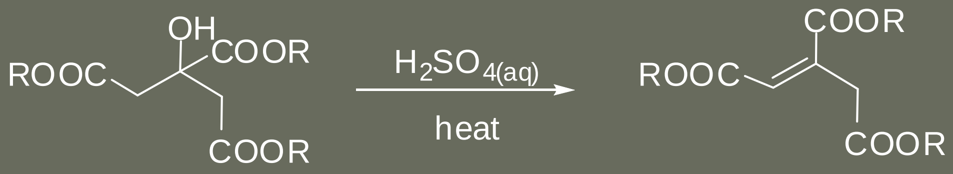
run	T	C	K	label	yield (%)	calculated
1	-	-	-	1	60	60.25 ± 2
2	+	-	-	t	72	73.25 ± 2
3	-	+	-	c	54	55.25 ± 2
4	+	+	-	tc	68	68.25 ± 2
5	-	-	+	k	52	50.25 ± 2
6	+	-	+	tk	83	83.25 ± 2
7	-	+	+	ck	45	45.25 ± 2
8	+	+	+	tck	80	78.25 ± 2

- The meaning of those numbers

$$\text{yield} = 64.25 + 11.5T - 2.5C + 5TK \pm$$

T Temperature (°C)		C Concentration (M)		K Catalyst	
120	160	1.5	2.5	H_3PO_4	H_2SO_4
-1	+1	-1	+1	-1	+1

- Categorical reaction conditions can be optimized



$$\text{yield} = 64.25 + 16.5T - 2.5C \pm$$

Something important

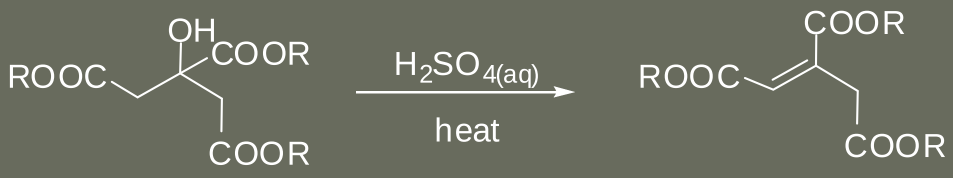
- It was possible to choose one catalyst because the interaction TK was identified

$$\text{yield} = 64.25 + 11.5T - 2.5C + 5TK \pm$$

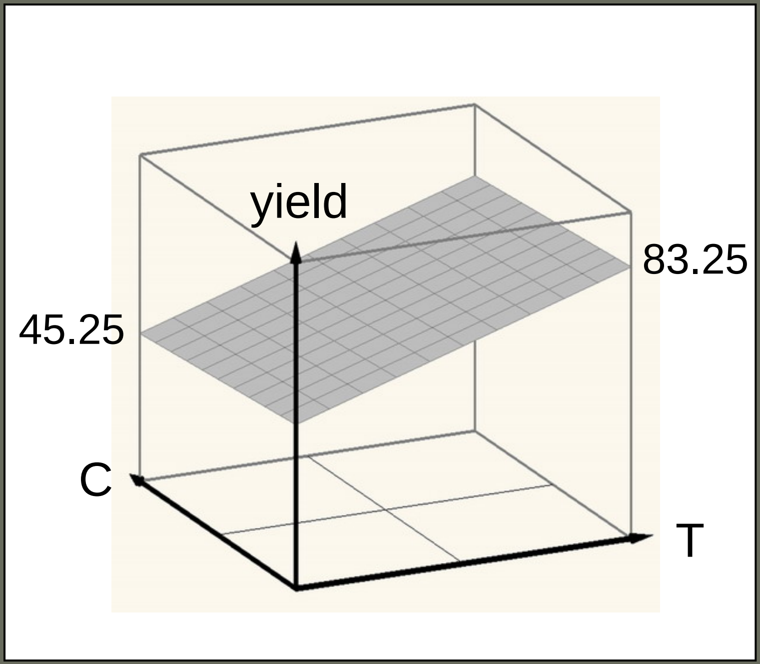
run	T	C	K	yield (%)	
1	-	-	-	60	H_3PO_4
2	+	-	-	72	H_3PO_4
3	-	+	-	54	H_3PO_4
4	+	+	-	68	H_3PO_4
5	-	-	+	52	H_2SO_4
6	+	-	+	83	H_2SO_4
7	-	+	+	45	H_2SO_4
8	+	+	+	80	H_2SO_4

In order to get the maximum yield (maximize the function), the catalyst has to be H_2SO_4

• The meaning of those numbers



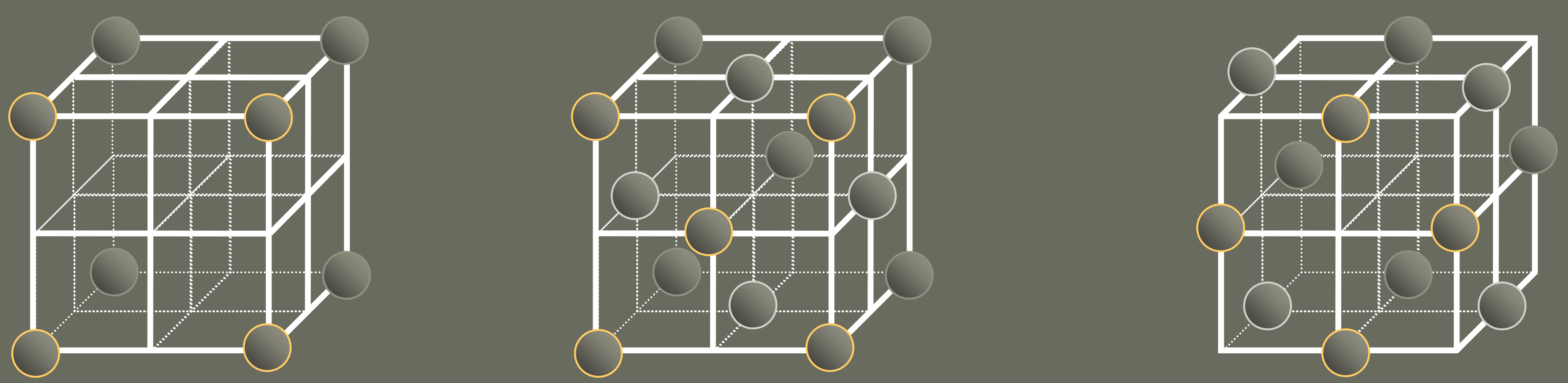
$$\text{yield} = 64.25 + 16.5T - 2.5C \pm$$



- To find the optimum conditions, we need to make sure that this function represents the entire space

Other factorial designs

- Full factorial design
- Central composite
- Box-Benhken



Tye, H. *Drug Discovery Today* **2004,** *9*, 485-491.

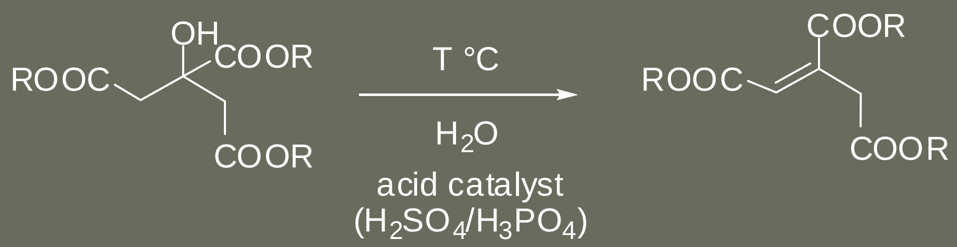
Outline

- Determining important reaction conditions
  - Fractional factorial design
- Analysis of reaction condition effects
  - Factorial design
- Estimation of the optimum conditions
  - Response surface analysis

• Fractional Factorial designs

- Factorial designs work perfectly for determining important factors

  …if you have 3 reaction conditions, as in the example



- If you had to analyze 7 reaction conditions at 2 values each, you would need to run 2^7=128 experiments!
- By virtue of statistics, it is possible to lower that number and get the same information

• m^{n-p} Fractional Factorial designs

number of values for each reaction condition

m^{n-p}

actual number of reaction conditions

number of “ignored” reaction conditions

- A m^{n-p} fractional factorial design requires m^{n-p} experiments

- If we analyze 2 values or options for 4 reaction conditions (as if they were only 3), $2^{4-1}=8$ experiments need to be run

Tranter, R., *Design and analysis in chemical research*. Sheffield Academic; CRC Press: Sheffield, England, **2000**.

Effects vs. interactions

- This is what we got before:

	result
average	64.25
T	11.5
C	-2.5
TC	0.75
K	0.75
TK	5.0
CK	0
TCK	0.25

	Important?
main effects	Very often
2-factor interactions	Often
3-factor interactions	Sometimes
4-factor interactions	Very rarely
more-than-5-factor interactions	If you get to here you have something very unusual!

Tranter, R., *Design and analysis in chemical research*. Sheffield Academic; CRC Press: Sheffield, England, 2000

2^{4-1} Fractional factorial design

- Yates's algorithm:

run	A	B	C	D	yield (%)	(1)	(2)	(3)	div	result	
1	-	-	-	-	#	#	#	#	8	#	av + *ABCD*
2	+	-	-	+	#	#	#	#	8	#	*A* + *BCD*
3	-	+	-	+	#	#	#	#	8	#	*B* + *ACD*
4	+	+	-	-	#	#	#	#	8	#	*AB* + *CD*
5	-	-	+	+	#	#	#	#	8	#	*C* + *ABD*
6	+	-	+	-	#	#	#	#	8	#	*AC* + *BD*
7	-	+	+	-	#	#	#	#	8	#	*BC* + *AD*
8	+	+	+	+	#	#	#	#	8	#	*ABC* + *D*

Box, G. E. P.; Hunter, W. G.; Hunter, J. S., *Statistics for experimenters : an introduction to design, data analysis, and model building*. Wiley: New York, **1978**.

Fractional factorial designs

Number of reaction conditions

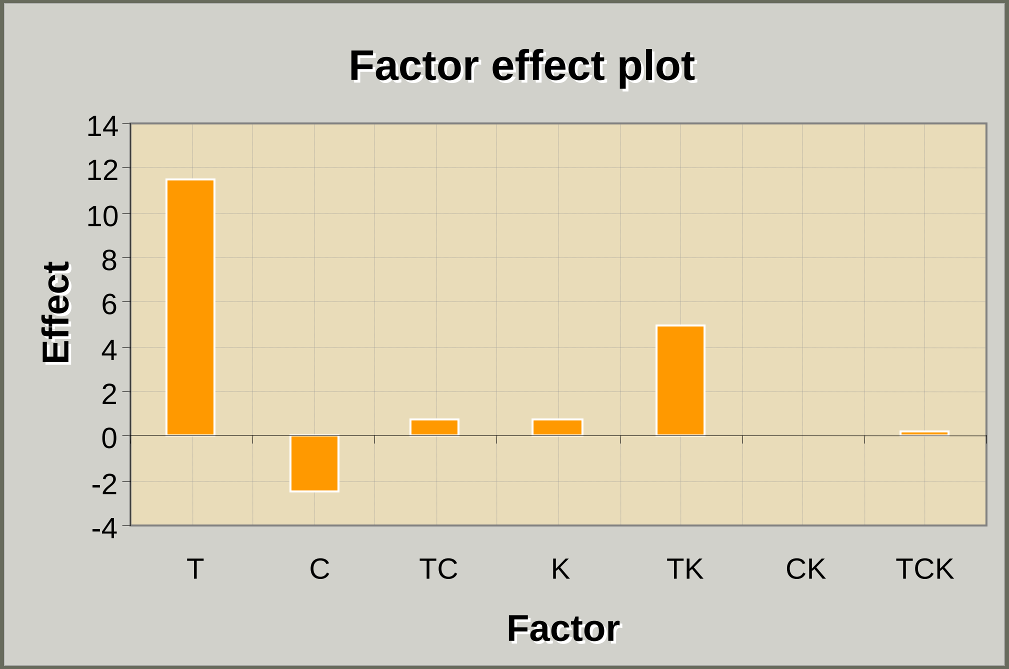
Number of experimental runs

	2	3	4	5	6	7	8	9	10	11	12	13	14	15
4	2^{2}	2^{3-1}												
8		2^{3}	2^{4-1}	2^{5-2}	2^{6-3}	2^{7-4}								
16			2^{4}	2^{5-1}	2^{6-2}	2^{7-3}	2^{8-4}	2^{9-5}	2^{10-6}	2^{11-7}	2^{12-8}	2^{13-9}	2^{14-10}	2^{15-11}
32				2^{5}	2^{6-1}	2^{7-2}	2^{8-3}	2^{9-4}	2^{10-5}	2^{11-6}	2^{12-7}	2^{13-8}	2^{14-9}	2^{15-10}
64					2^{6}	2^{7-1}	2^{8-2}_{V}	2^{9-3}	2^{10-4}	2^{11-5}	2^{12-6}	2^{13-7}	2^{14-8}	2^{15-9}
128						2^{7}	2^{8-1}	2^{9-2}	2^{10-3}	2^{11-4}	2^{12-5}	2^{13-6}	2^{14-7}	2^{15-8}
256							2^{8}	2^{9-1}	2^{10-2}	2^{11-3}	2^{12-4}	2^{13-5}	2^{14-6}	2^{15-7}
512								2^{9}	2^{10-1}	2^{11-2}	2^{12-3}	2^{13-4}	2^{14-5}	2^{15-6}

Design Expert 7.0.3 (Stat-Ease Inc.) (http://www.statease.com)

How to compare the effects?

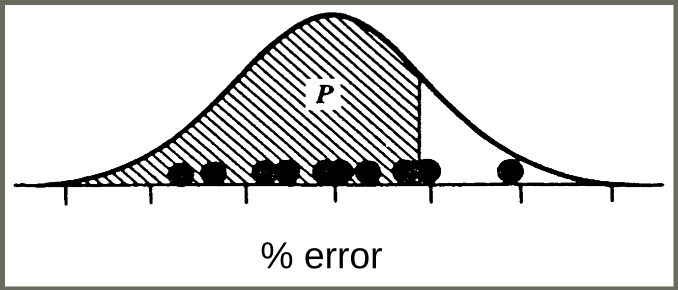
- In the case of 3 reaction conditions, a “Factor effect plot” is enough

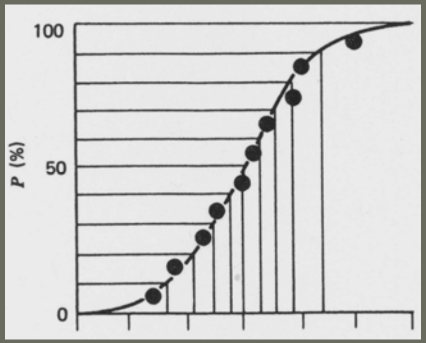


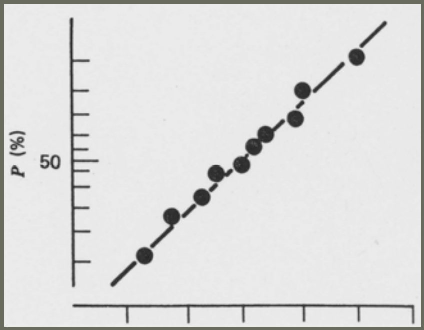
- For a high number of reactions, a normal plot is needed

• Normal plots

- Let's assume that the experimental error follows a normal distribution





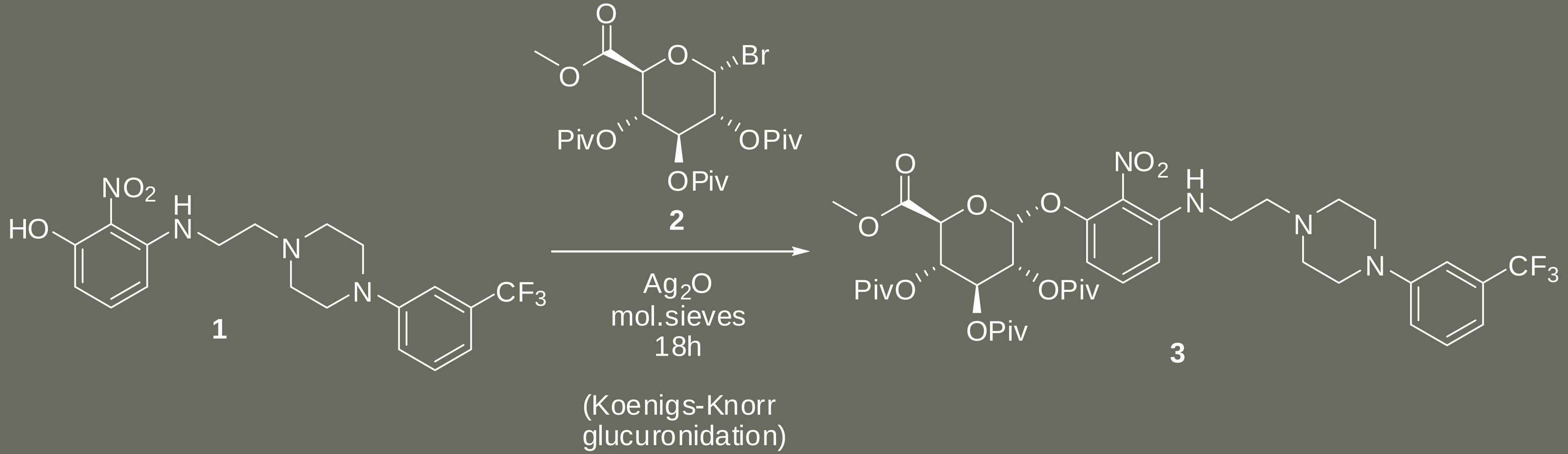


Normal plot

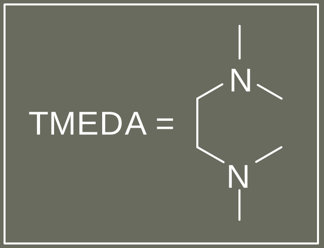
- In a normal plot, reaction condition effects that are due to experimental error will appear forming a straight line

Box, G. E. P.; Hunter, W. G.; Hunter, J. S., *Statistics for experimenters : an introduction to design, data analysis, and model building*. Wiley: New York, **1978**.

Application example

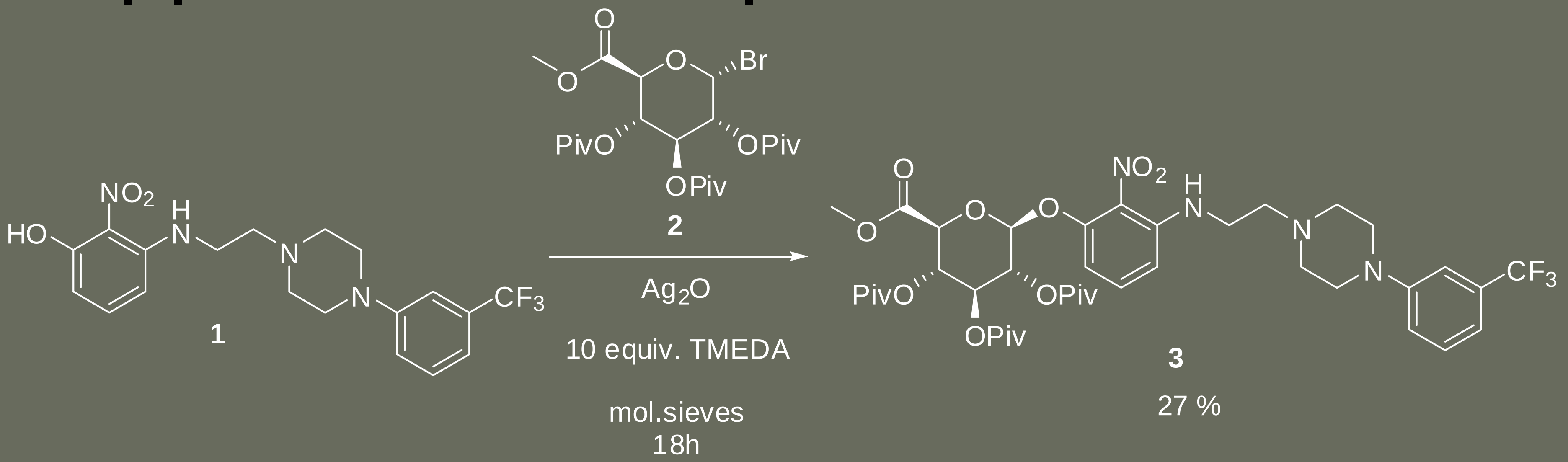


- Chelation was identified as the reason for the bad yield
- Addition of TMEDA (10 equiv.), increased the yield to 27%



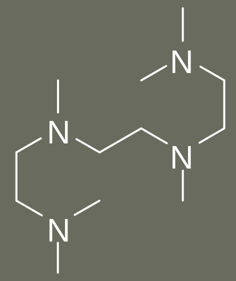
Stazi, F.; Palmisano, G.; Turconi, M.; Clini, S.; Santagostino, M. *J. Org. Chem.* **2004,** 69, 1097-1103.

Application example

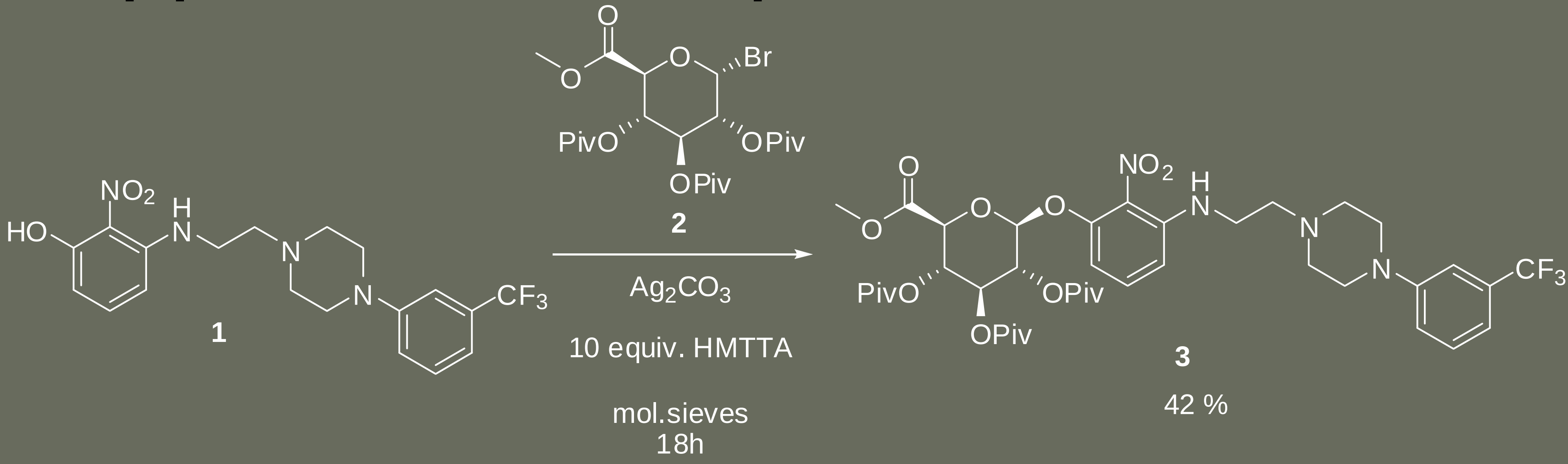


- DoE methods (3^2 factorial design) were applied to screen amine additives and silver sources giving: HMTTA and Ag_2CO_3 as best combination

HMTTA =



Application example



- A 2^{7-4} fractional factorial (8 experiments) design was used:

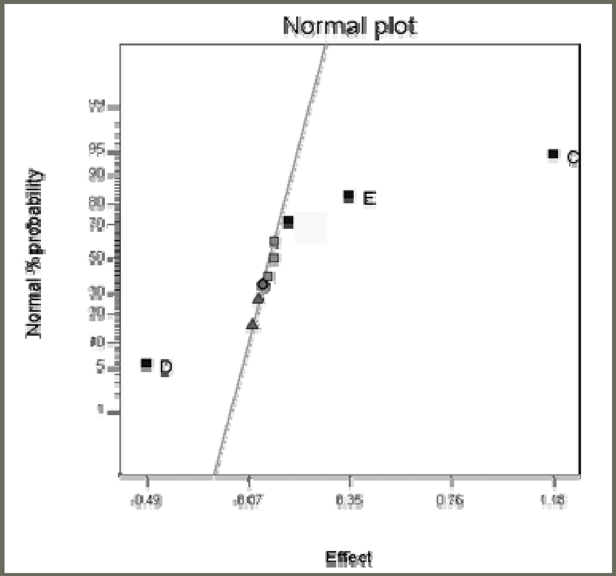
	Reaction condition	-1	+1
A	pre-complex time (min)	0	60
B	reaction time (h)	2	6
C	Ag_2CO_3 (equiv)	1.5	3.8
D	HMTTA (equiv)	1.5	12.6
E	sugar derivative (equiv)	1.5	3
F	4 Å mol sieves (mg)	0	100
G	solvent (mL)	0.5	1.5

Application example

- 2^{7-4} factorial design results:

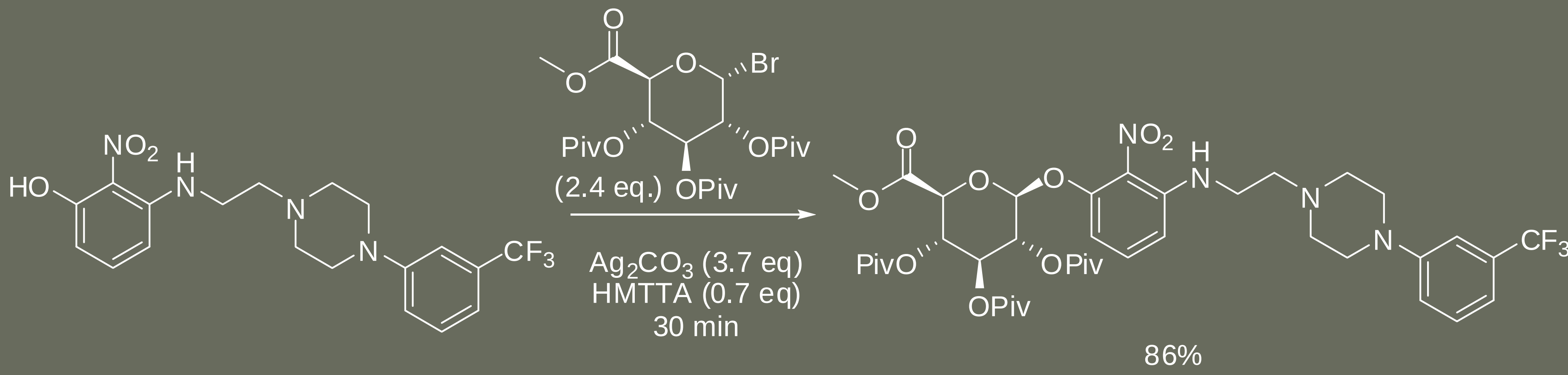
A	pre-complex time (min)
B	reaction time (h)
C	Ag_2CO_3 (equiv)
D	HMTTA (equiv)
E	sugar derivative (equiv)
F	4 Å mol sieves (mg)
G	solvent (mL)

run	A	B	C	D	E	F	G	yield (%)
1	-	-	-	+	+	+	-	14.7
2	+	-	-	-	-	+	+	19.5
3	-	+	-	-	+	-	+	24.4
4	+	+	-	+	-	-	-	11.2
5	-	-	+	+	-	-	+	34.2
6	+	-	+	-	+	-	-	83.2
7	-	+	+	-	-	+	-	56.5
8	+	+	+	+	+	+	+	55.4

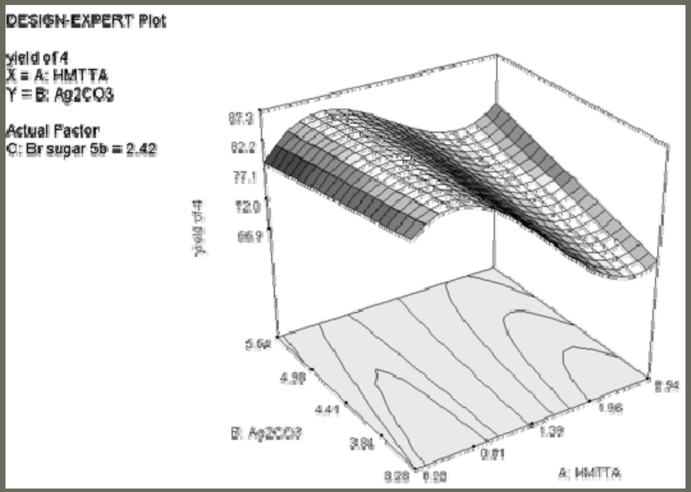


Stazi, F.; Palmisano, G.; Turconi, M.; Clini, S.; Santagostino, M. *J. Org. Chem.* **2004,** 69, 1097-1103.

• Application example



- Finally, a 2^3 factorial design and <u>response surface analysis</u> gave the optimum conditions



Stazi, F.; Palmisano, G.; Turconi, M.; Clini, S.; Santagostino, M. *J. Org. Chem.* **2004,** 69, 1097-1103.

Outline

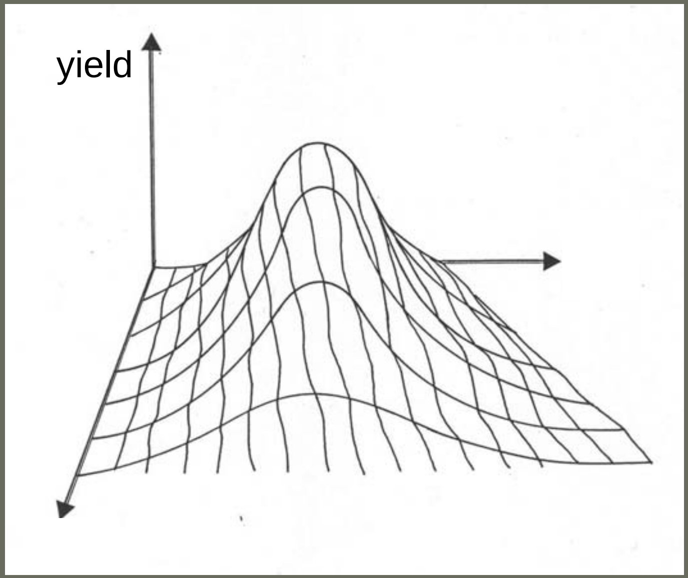
- Determining important reaction conditions
  - Fractional factorial design

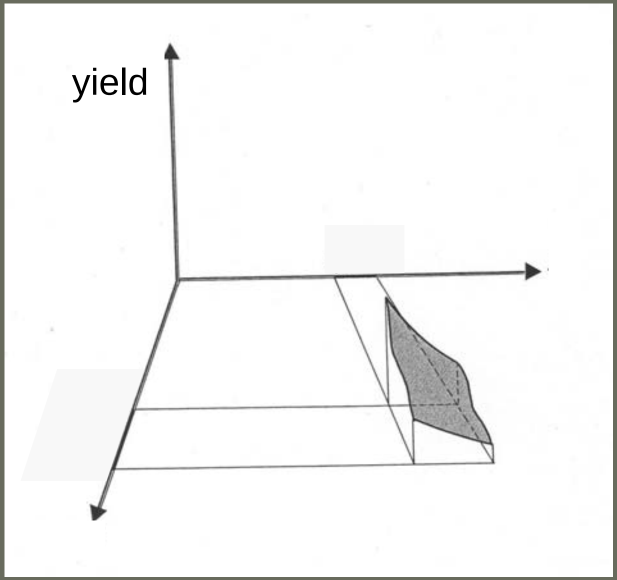
- Analysis of reaction condition effects
  - Factorial design

- Estimation of the optimum conditions
  - Response surface analysis

Response surface analysis

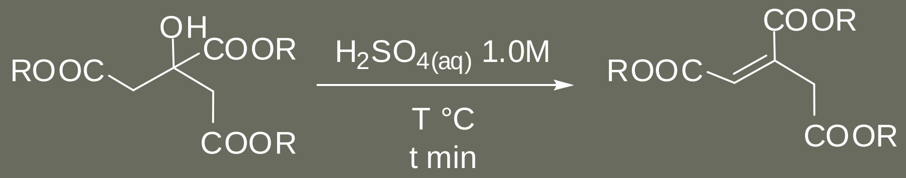
- The problem of optimizing a synthetic reaction corresponds to locate the maximum value of a function from a mathematical point of view





Carlson, R., *Design and optimization in organic synthesis*. Elsevier: Amsterdam; New York, **1992**.

Response surface analysis



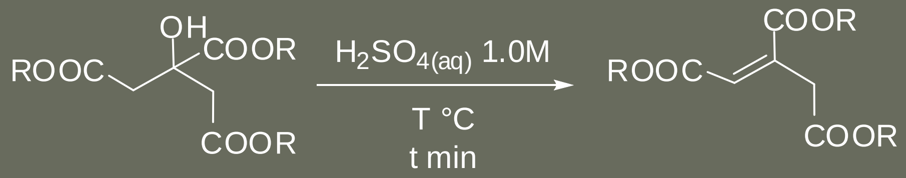
t time (min)		T Temperature (°C)	
70	80	127.5	132.5
-1	+1	-1	+1

run	t	T
1	-	-
2	+	-
3	-	+
4	+	+
5	0	0
6	0	0
7	0	0

Central point: three times to calculate the experimental error

Box, G. E. P.; Hunter, W. G.; Hunter, J. S., *Statistics for experimenters : an introduction to design, data analysis, and model building*. Wiley: New York, **1978**.

• Response surface analysis

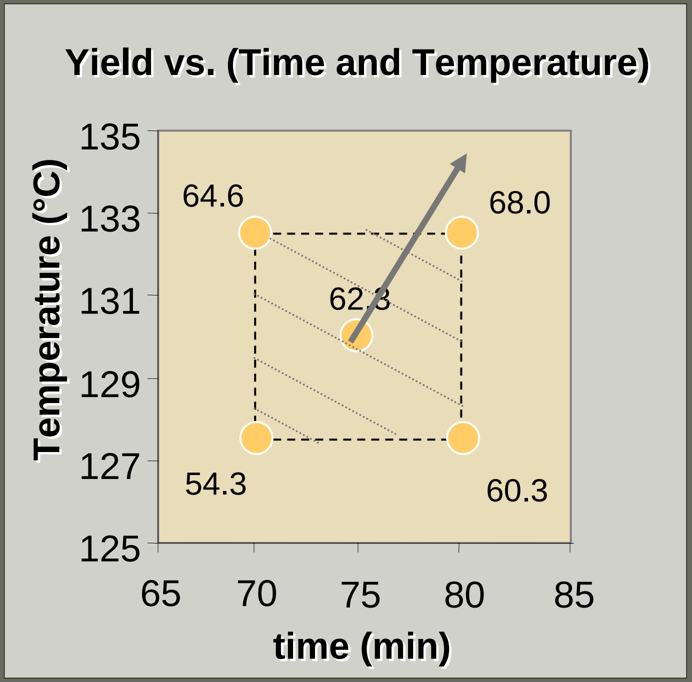


$$\text{yield} = 62.01 + 2.35t + 4.5T \pm$$

run	t	T	yield (%)
1	-	-	54.3
2	+	-	60.3
3	-	+	64.6
4	+	+	68.0
5	0	0	60.3
6	0	0	64.3
8	0	0	62.3

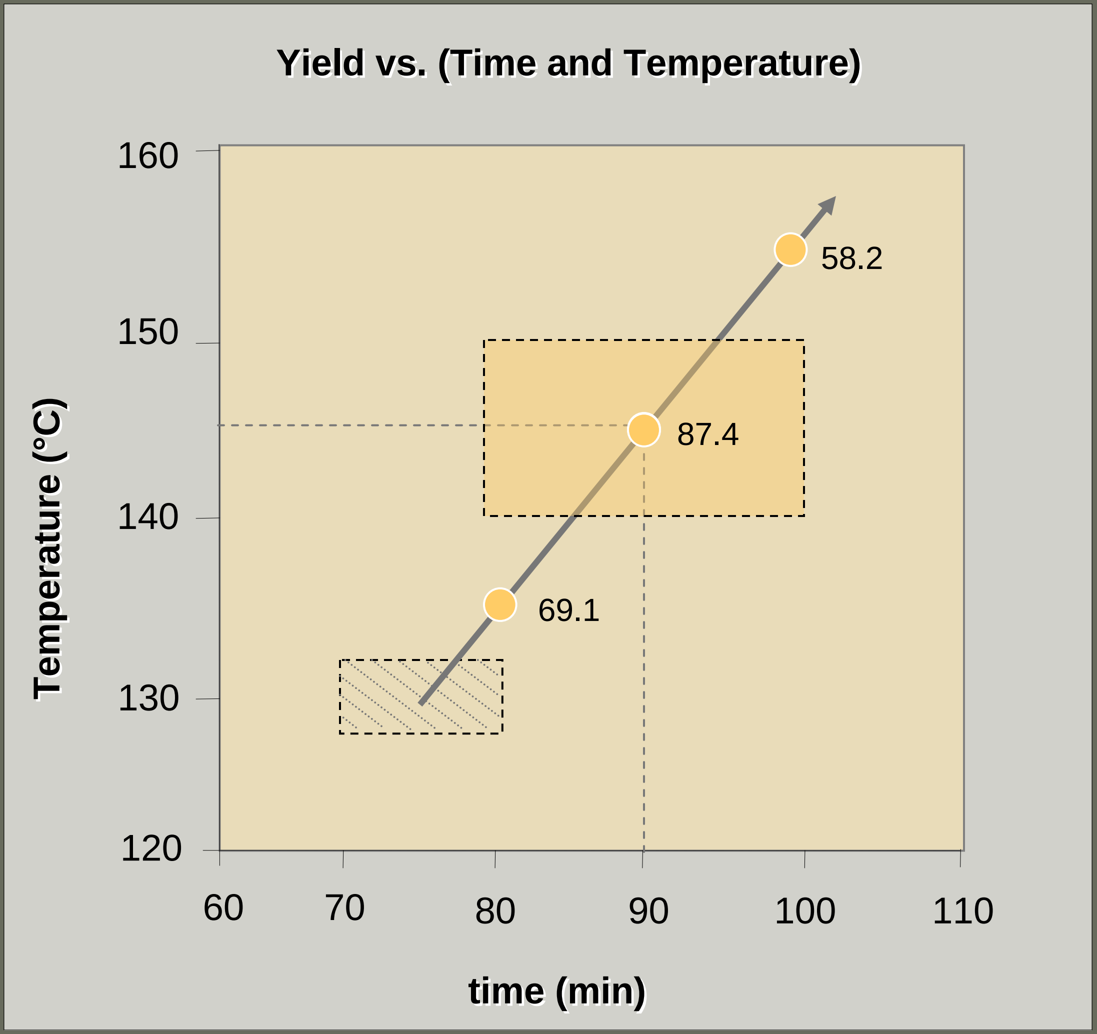
3 central points

$e = 2$



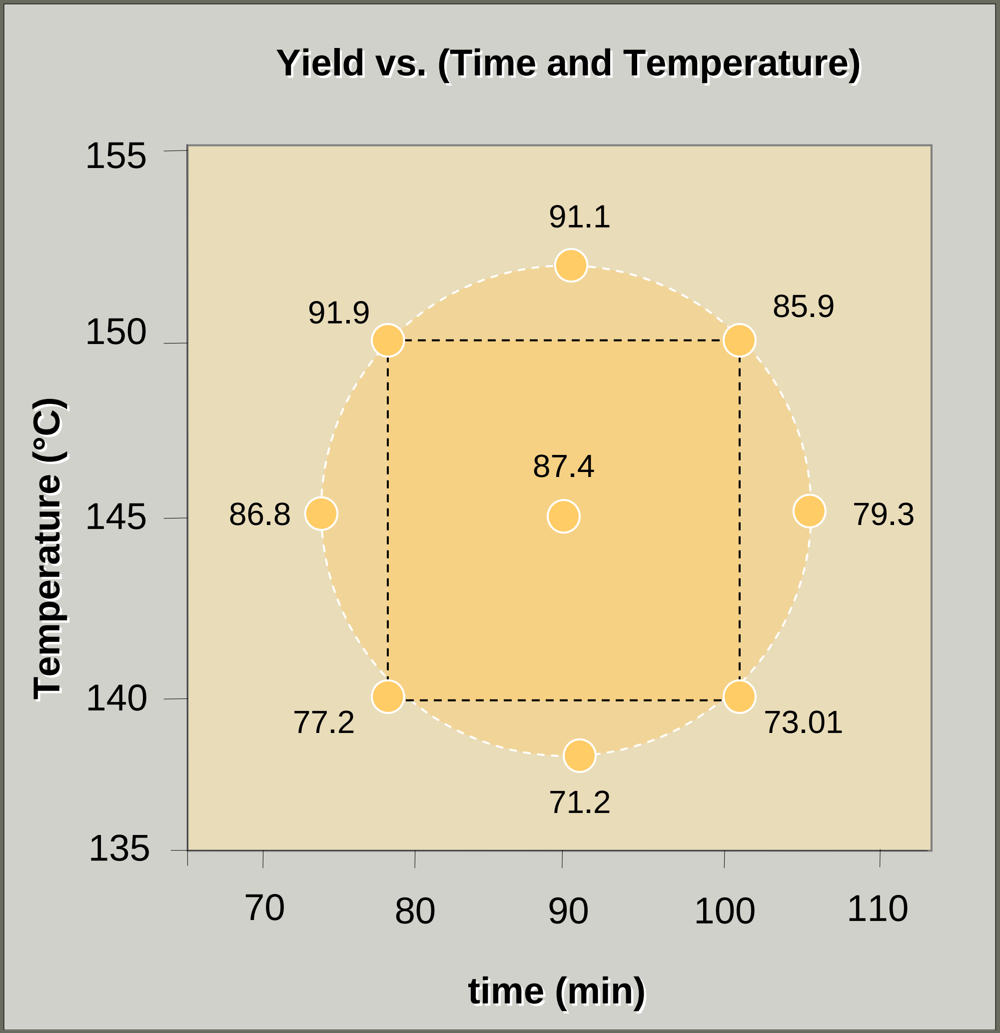
Box, G. E. P.; Hunter, W. G.; Hunter, J. S., *Statistics for experimenters : an introduction to design, data analysis, and model building*. Wiley: New York, **1978**.

• Response surface analysis



Box, G. E. P.; Hunter, W. G.; Hunter, J. S., *Statistics for experimenters : an introduction to design, data analysis, and model building*. Wiley: New York, **1978**.

Response surface analysis



- Equation for the 2^2 factorial design:

$$\text{yield} = 82.09 - 2.69t + 6.97T \pm$$

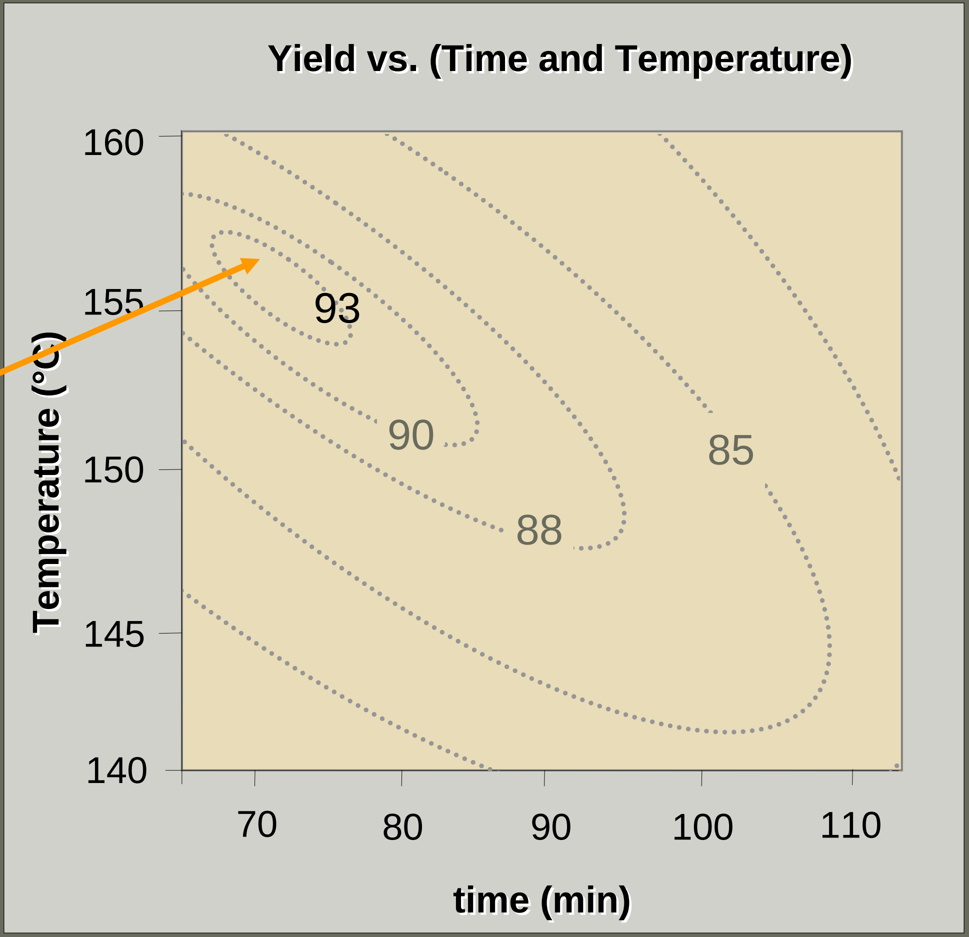
- Calculated equation for the surface:

$$\text{yield} = 87.36 - 2.69t + 6.97T$$
$$- 2.15t^2 - 3.12T^2 - 0.58Tt \pm$$

Box, G. E. P.; Hunter, W. G.; Hunter, J. S., *Statistics for experimenters : an introduction to design, data analysis, and model building*. Wiley: New York, **1978**.

Response surface analysis

$$yield = 87.36 - 2.69t + 6.97T - 2.15t^2 - 3.12T^2 - 0.58Tt \pm$$



Optimum conditions:

T = 157 °C

t = 73 min

yield: 93%

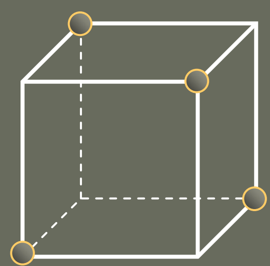
Box, G. E. P.; Hunter, W. G.; Hunter, J. S., *Statistics for experimenters : an introduction to design, data analysis, and model building*. Wiley: New York, **1978**.

- Sequential nature of experimentation

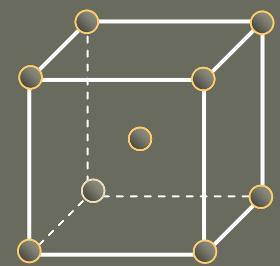
Hypercube design in n dimensions

Design in 2,3,4 dimensions

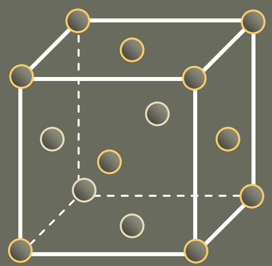
Plan



Fractional factorial design



Full factorial design

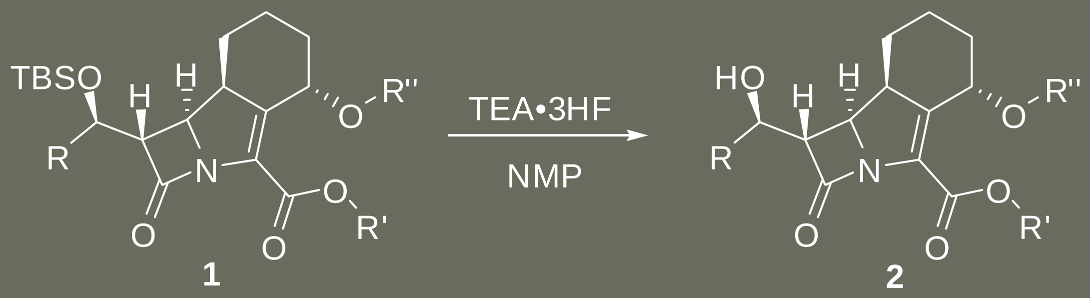


Central composite

Response surface analysis

Tranter, R., *Design and analysis in chemical research*. Sheffield Academic; CRC Press: Sheffield, England, **2000**.

Application of response surface analysis



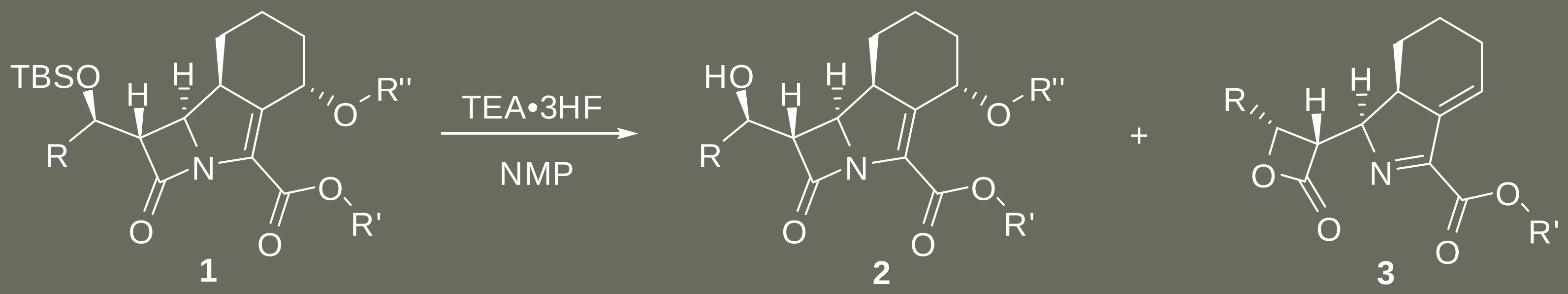
- 2^4 central composite

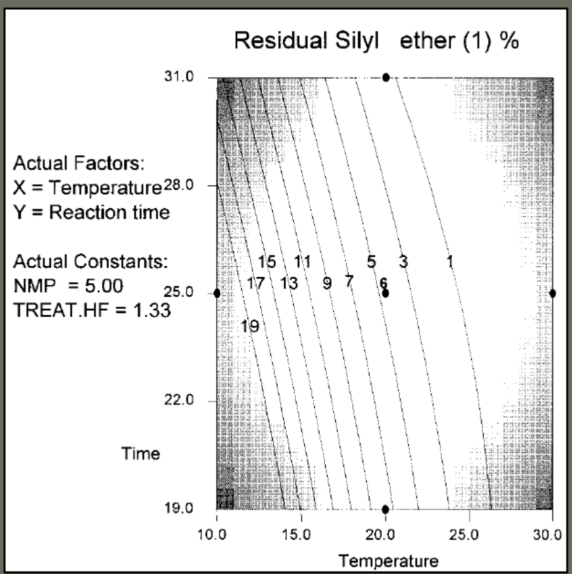
reaction condition	range		units
temperature	10	30	°C
time	19	31	hours
volume of NMP	3	7	mL/g of substrate
equivalents of TEA.3HF	1	1.67	Equivalents

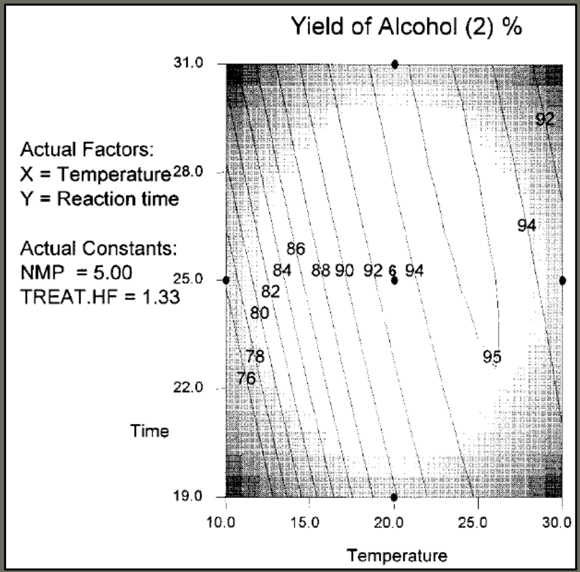
- Monitored results:
  - % yield of alcohol
  - % lactone
  - % remaining silyl ether

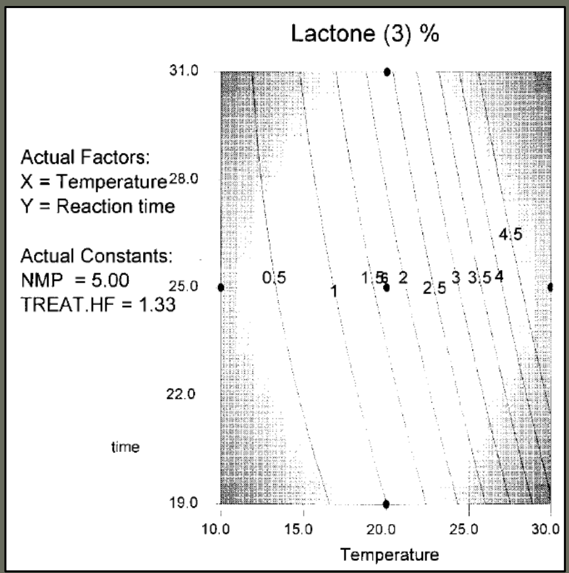
Owen, M. R.; Luscombe, C.; Lai, L. W.; Godbert, S.; Crookes, D. L.; Emiabata-Smith, D. *Org. Proc. Res. Dev.* **2001**, *5*, 308-323.

• Application of response surface analysis



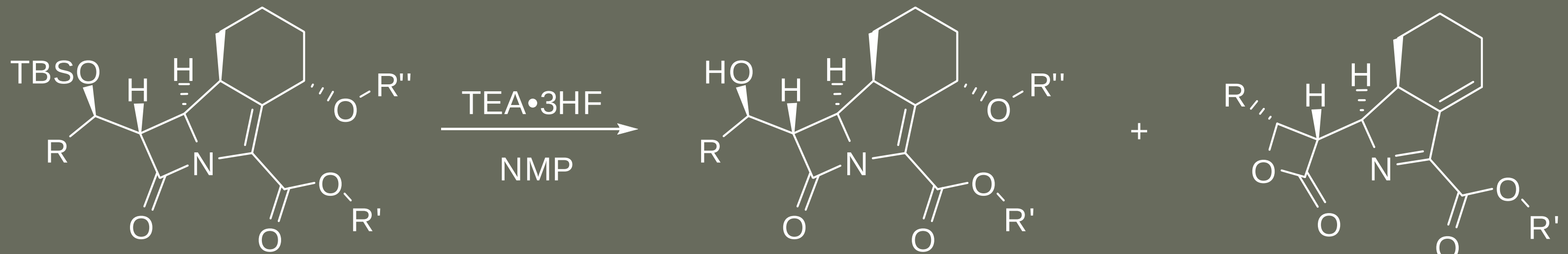






Owen, M. R.; Luscombe, C.; Lai, L. W.; Godbert, S.; Crookes, D. L.; Emiabata-Smith, D. *Org. Proc. Res. Dev.* **2001**, *5*, 308-323.

Application

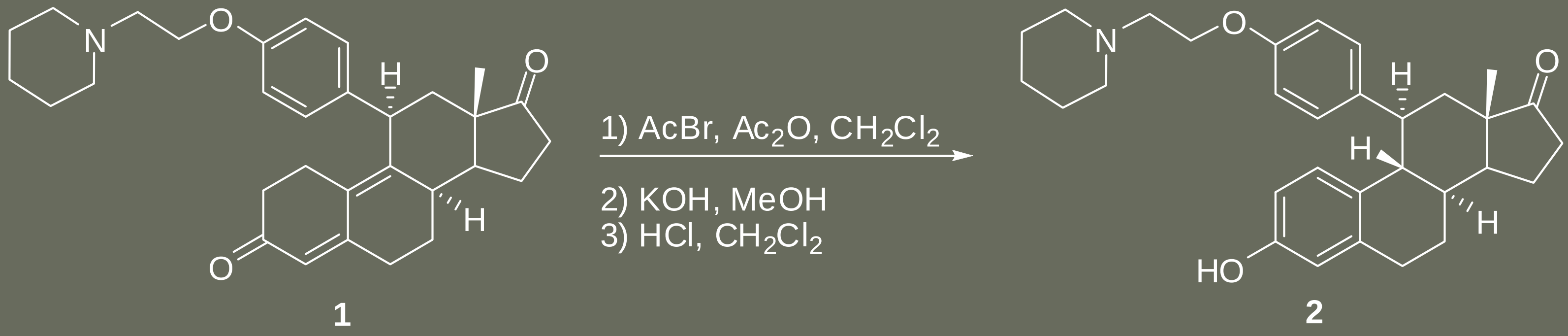


1 2 3

	Predicted conditions				product yield (%)		impurity (%)	
target/constraints	T (°C)	Time (h)	solvent	Et3N·3HF	predicted	actual	predicted	actual
max yield	19	31	3.6	1.42	95.3	95.8	3.3	3.3
lactone < 2%	17	31	4.8	1.50	94.2	94.0	1.9	1.7
lactone < 1.1%	16	29	5.3	1.68	92.4	93.1	1.1	1.1
lactone < 2%, solvent < 3.5 mL/g	14	31	3.45	1.58	93.9	94.2	1.8	2.0
lactone < 2% $Et_3N.3HF$ < 1.18eq.	28	19.5	7	1.17	93.7	93.4	1.9	2.0
lactone < 2%, time < 23 h	24	23	6.3	1.41	94.2	94.2	2.0	1.9

Owen, M. R.; Luscombe, C.; Lai, L. W.; Godbert, S.; Crookes, D. L.; Emiabata-Smith, D.
Org. Proc. Res. Dev. **2001**, *5*, 308-323.

When DoE "fails"



entry	Ac_2O (equiv)	AcBr (equiv)	T (°C)	yield (%) (20 g)	yield(%) (20 kg)	comments
1	3	3.8	23-27	77.3	< 70	original conditions

Conditions: *t* = 4-5h; yield of **2** after crystallization

Larkin, J. P.; Wehrey, C.; Boffelli, P.; Lagraulet, H.; Lemaitre, G.; Nedelec, A.
Org. Proc. Res. Dev. **2002,** *6*, 20-27.

Outline

- Determining important reaction conditions
  - Fractional factorial design
- Analysis of reaction condition effects
  - Factorial design
- Estimation of the optimum conditions
  - Response surface analysis
- Recent advances
  - Software
  - Automation

- “DoE involves a lot of math, it’s rather complicated”

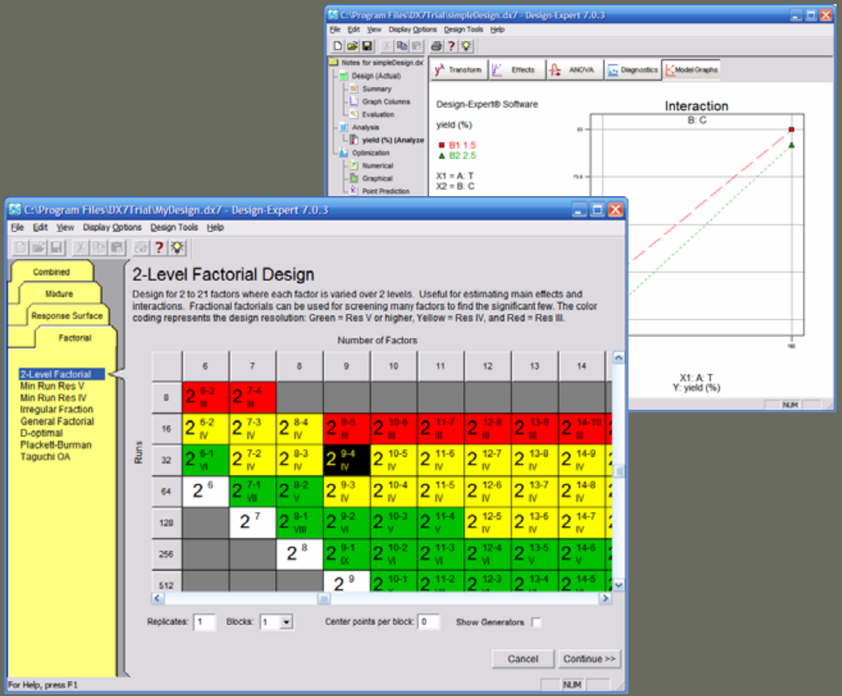
  - People tend not to utilize DoE because of the tedious mathematical manipulations.

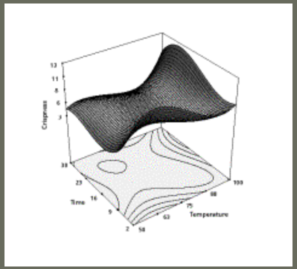
Lendrem, D.; Owen, M.; Godbert, S. *Org. Proc. Res. Dev.* **2001,** *5*, 324-327.

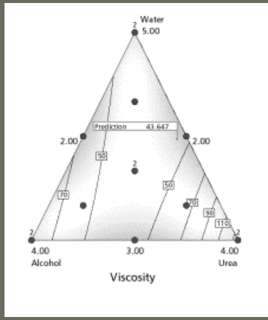
Software

Most commonly used:

- Stat-Ease Design Expert ®
  (http://www.statease.com)
- Umetrics MODDE ®
  (http://www.umetrics.com)
- S-matrix Fusion Pro ®
  (http://www.smatrix.com)







• What if I need to run $>2^4$ experiments?

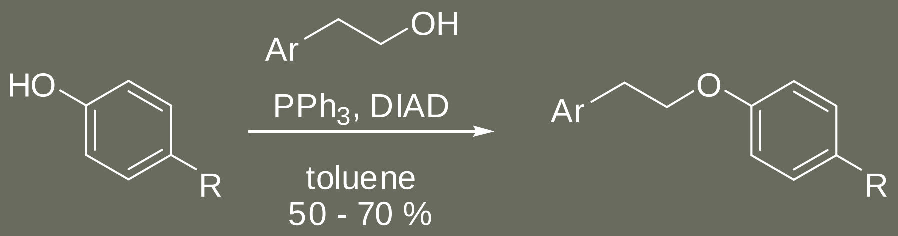
The answer is to use automation

- Some features of automated systems, commercially available:
  - Up to 100 simultaneous reactions
  - Automated liquid handler
  - Vessel volume: 100 μL → 250 mL
  - Temperatures: -100 °C → 350 °C
  - Reflux, N_2 blanketing, automated N_2/vacuum manifold
  - On-line HPLC

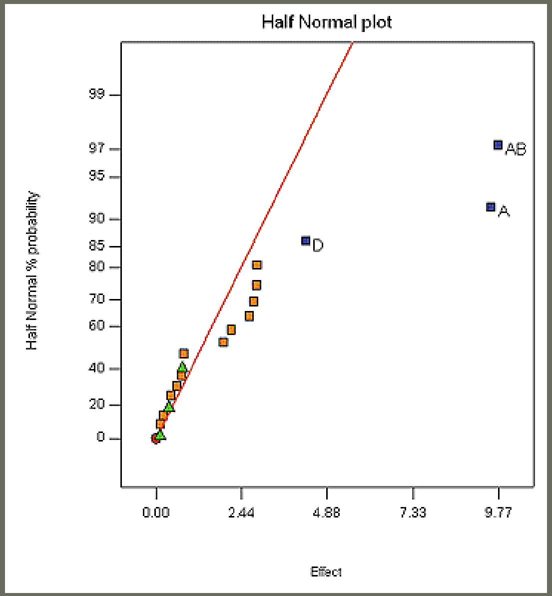
Harre, M.; Tilstam, U.; Weinmann, H. *Org. Proc. Res. Dev.* **1999**, *3*, 304-318.

Example of the use of automation

- System:
  - Automated liquid handler
  - On-line HPLC



- Reaction conditions:
  - A equivalents of alcohol
  - B equivalents of DIAD
  - C volume of toluene
  - D temperature
  - E addition rate of DIAD

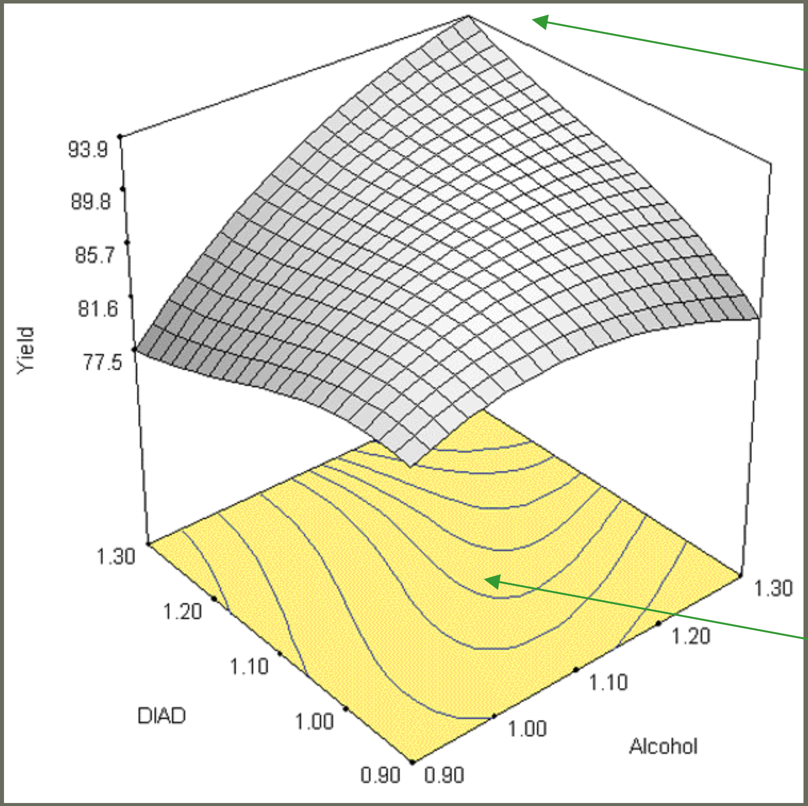


- 20 experimental runs
- Total research time: 5 days

Important factors: ratio DIAD/alcohol, alcohol, temperature

Emiabata-Smith, D. F.; Crookes, D. L.; Owen, M. R. *Org. Proc. Res. Dev.* **1999,** 3, 281-288.

• Why DoE methods are ideal for us



Further exploration would lead us to obtain > 94% yield

1.1 equivalents of DIAD and

1.1 equivalents of alcohol

89% yield, almost pure product after workup

Emiabata-Smith, D. F.; Crookes, D. L.; Owen, M. R. *Org. Proc. Res. Dev.* **1999,** 3, 281-288.

- Some final comments

  - DoE offers powerful mathematical models that are applicable to the behavior of organic reactions

  - DoE methods are a daily practice in industrial chemistry. Current applications and results are not being published

  - DoE is not a substitute for creative chemistry, but it can be a great supplement

DoE is a tool

- A tool… like a hammer
- The only way to know <u>how</u> it works is to use it
- If you don't try it, you will never know that it actually works
- When you get used to the hammer, you wouldn't use a rock again

Lendrem, D.; Owen, M.; Godbert, S. *Org. Proc. Res. Dev.* **2001,** *5*, 324-327.

• Acknowledgements

Prof. Maleczka

Prof. Walker

The Maleczka group

Nicki, Jill, Monica, Feng, Soong-Hyun (Kim),

Il Hwan, Bani, and Kyoungsoo

Aman, Aman, Toyin, Calvin