An Introduction to the Uniform Design for Industrial Experiments with Model Unknown

Kai-Tai Fang, Chair Prof.

Hong Kong Baptist University

Email:ktfang@hkbu.edu.hk

Hong Kong Society for Quality

August 29, 2003

Ford Motor Company

21500 Oakwood Blvd.
POEE - FK143 Zone E-3, MD65
Dearborn, MI 48121

July 12, 2002

To: Prof. Kai-Tai Fang
Department of Mathematics
Hong Kong Baptist University
Kowloon Tong, Hong Kong

Dear Prof. Fang,

I would like to invite you to visit Ford Motor Company from July 30-August 1, 2002 to provide a seminar on "Uniform Design for Computer Experiments And Industrial Experiments."

In the past few years, we have tremendous successes in using Uniform Design for computer experiments. The technique has become a critical enabler for us to execute "Design for Six Sigma" to support new product development, in particular, automotive engine design. Today, computer experiments using uniform design have become standard practices at Ford Motor Company to support early stage of product design before hardware is available.

We would like to share with you our successful real world industrial experiences in applying the methodology that you developed. Additionally, your visit will be very valuable for us to gain more insight about the methodology as well as to learn the latest development in the area.

Ford Motor Company will provide all necessary travel accommodation during this trip.

Sincerely,

A. Sudjianto

Agus Sudjianto, Ph.D. (e-mail: asudjian@ford.com)
Engineering Manager, Analytical Powertrain
Ford Motor Company

Ford Motor Company

Dear Prof. Fang,

I would like to invite you to visit Ford Motor Company form July 30 – August 1, 2002 to provide a seminar on "Uniform Design for Computer Experiments And Industrial Experiments".

Ford Motor Company

In the past few years, we have tremendous in using Uniform Design for computer experiments. The technique has become a critical enabler for us to execute "Design for Six Sigma" to support new product development, in particular, automotive engine design. Today, computer experiments using uniform design have become standard practices at Ford Motor Company to support early stage of product design before hardware is available.

Ford Motor Company

We would like to share with you our successful real world industrial experiences in applying the methodology that you developed. Additionally, your visit will be very valuable for us to gain more insight about the methodology as well as to learn the latest development in the area.

Agus Sudjianto, Engineering Manager

Kai-Tai Fang at Ford Motor Company

Introduction

Statistical Experimental Designs

Design of experiment is a branch of statistics and has been playing an important role in development of sciences and new techniques, especially, in development of high-tech.

Experiments are performed almost everywhere nowadays, usually for the purpose of discovering something about a particular process or system.

Experimental design

- A good experimental design should minimize the number of experiments to acquire as much information as possible.
- Knowledge discovery

Complexities

Experiments, especially in high-tech experiments, have the following characteristics:

- Multi-factors
- Nonlinearity
- Experimental domain is large
- Underlying model is unknown
- No analytic formula of the response surface

Experiments
-- Underlying model is unknown

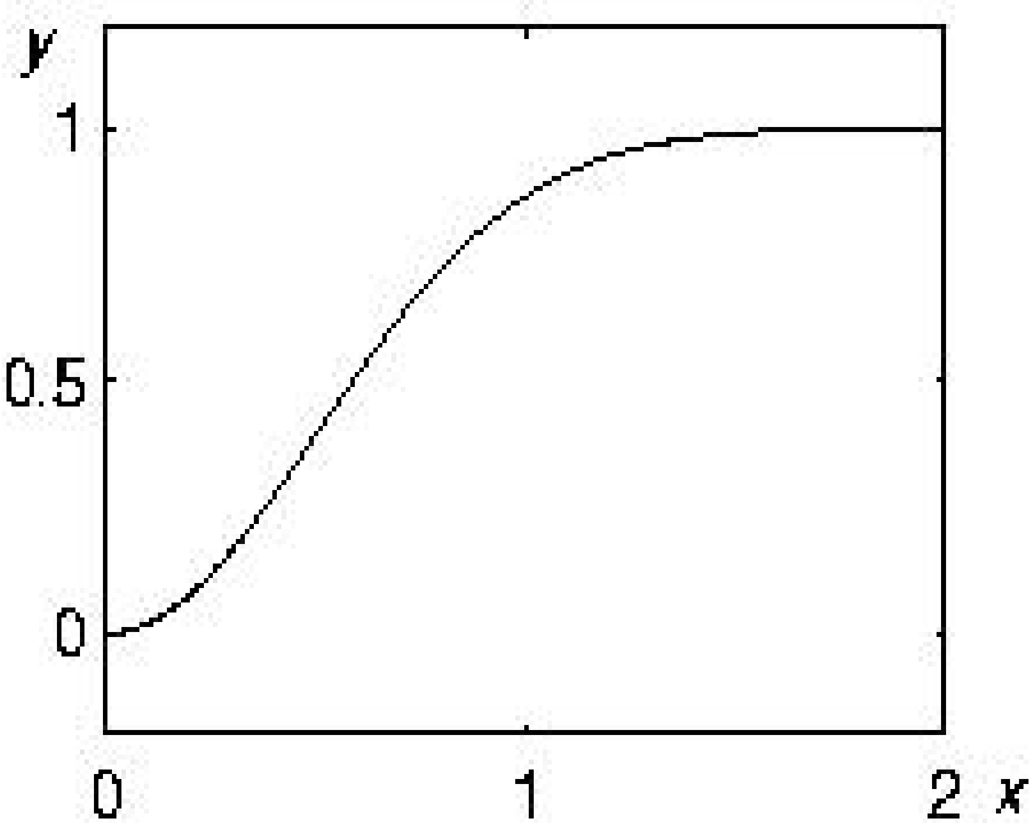
Example 1.

In a biological experiment we wish to explore the relationship between the growth time(x) and the response (y). Assume the underlying model

$$y = y(x) = 1 - e^{-2x^2}, \quad x \in [0, 2], \qquad (2.1)$$

is unknown. There are many ways to design this experiment based on different statistical models.

Weibull Growth Curve Model



1. ANOVA Models

The experimenter observes the response at several growth times, $\mathbf{x}_1, \cdots, \mathbf{x}_q$, that are called *levels*. For each $\mathbf{x}_j$ we repeat experiment $\mathbf{n}_j$ times and related responses are $\mathbf{y}_{1j}, \cdots, \mathbf{y}_{n_j j}$.

A statistical model is

$$y_{ij} = \mu_j + \varepsilon_{ij},\ j = 1, \cdots, q,\ i = 1, \cdots, n_j,$$

where μ_j is the true value $y(x_j)$ and ε_{ij} are random errors that are independently identically distributed according to $N(0, \sigma^2)$.

a. ANOVA Models
-- Factorial Designs

$$y_{ij} = \mu_j + \varepsilon_{ij},\ j=1, \cdots, q,$$

$$= \mu + \alpha_j + \varepsilon_{ij},\ i=1, \cdots, n_j,$$

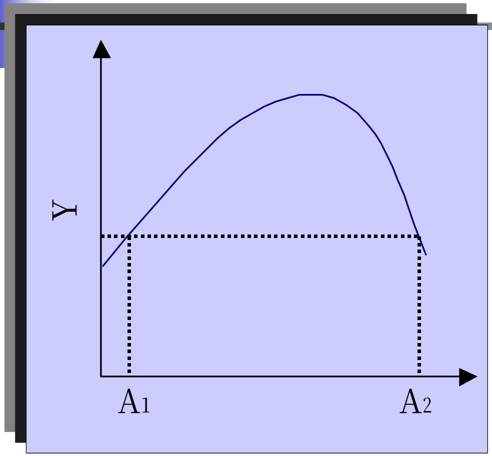
$$\alpha_1 + \cdots + \alpha_q = 0$$

- μ is the overall mean of y over [0,2]
- μ_j is the true value $\boldsymbol{y}(\boldsymbol{x}_j)$
- α_j is the main effect of y at $\boldsymbol{x}_j$
- ε_{ij} are *i.i.d.* random errors $N(0, \sigma^2)$.

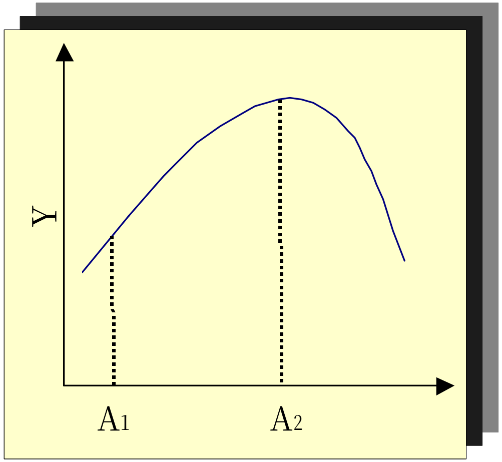
Under the model we need to find a design under which we can efficiently estimate $\{\mu_1, \cdots, \mu_q\}$ or $\{\mu, \alpha_1, \cdots, \alpha_q\}$ and to assess whether $y(x)$ significantly depends on $\boldsymbol{x} = \boldsymbol{x}_1, \cdots, \boldsymbol{x}_q$.

- Two-level factorial designs have been widely used
- Two-level factorial designs are not enough to explore non-linearity, like this example.
- Factorials with more levels are useful

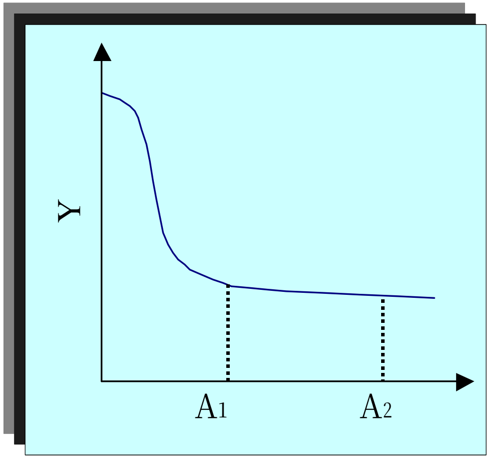
Two-level designs are shown to be insufficient for nonlinear model.



The right experimental range but wrong levels.

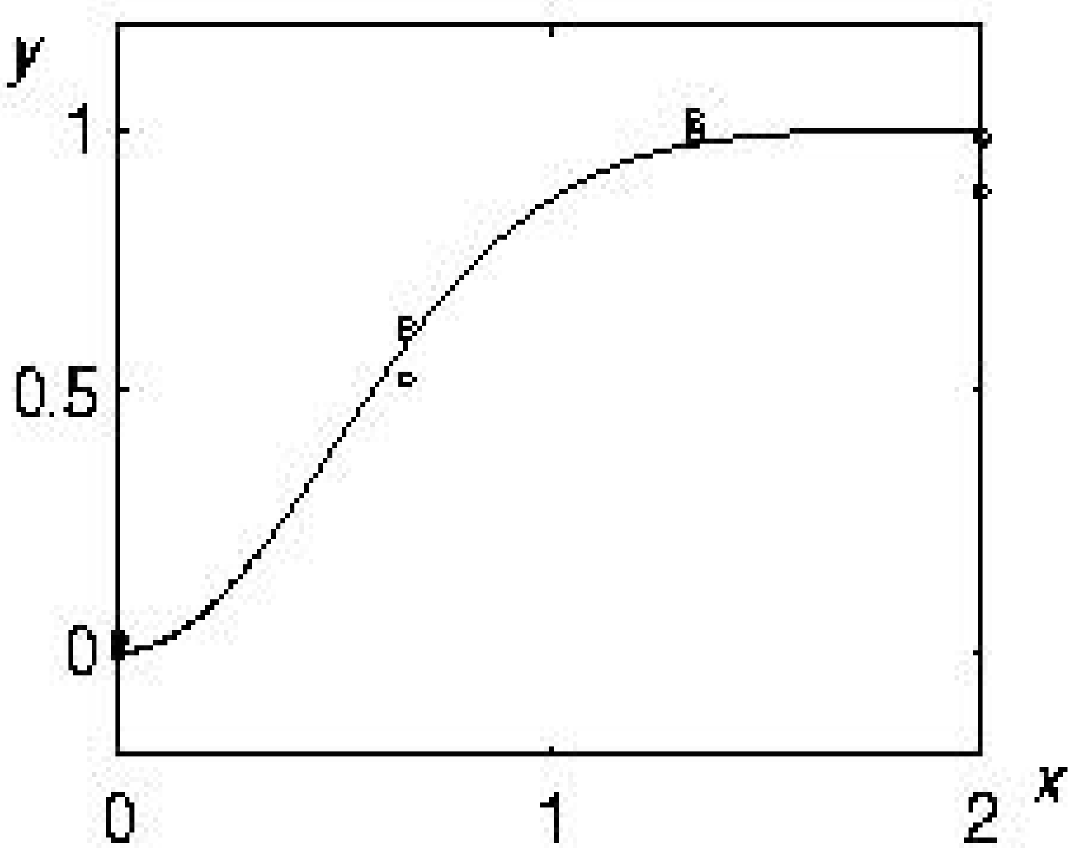


The right experimental range and right levels, but, it cannot explore more detailed relationships between Y and A.



Wrong experimental range.

Factorial design with 4 levels



b. Linear regression models
-- Optimal Designs

From the professional knowledge the experimenter wants to use a regression model to fit relationship between *x* and *y*, for example

[illegible]

or

$$y(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \varepsilon,$$

or more general

$$y(x) = \beta_1 f_1(x) + \cdots + \beta_m f_m(x) + \varepsilon,$$

where functions $f_1, \cdots, f_m$ are known and $\beta_1, \cdots, \beta_m$ unknown.

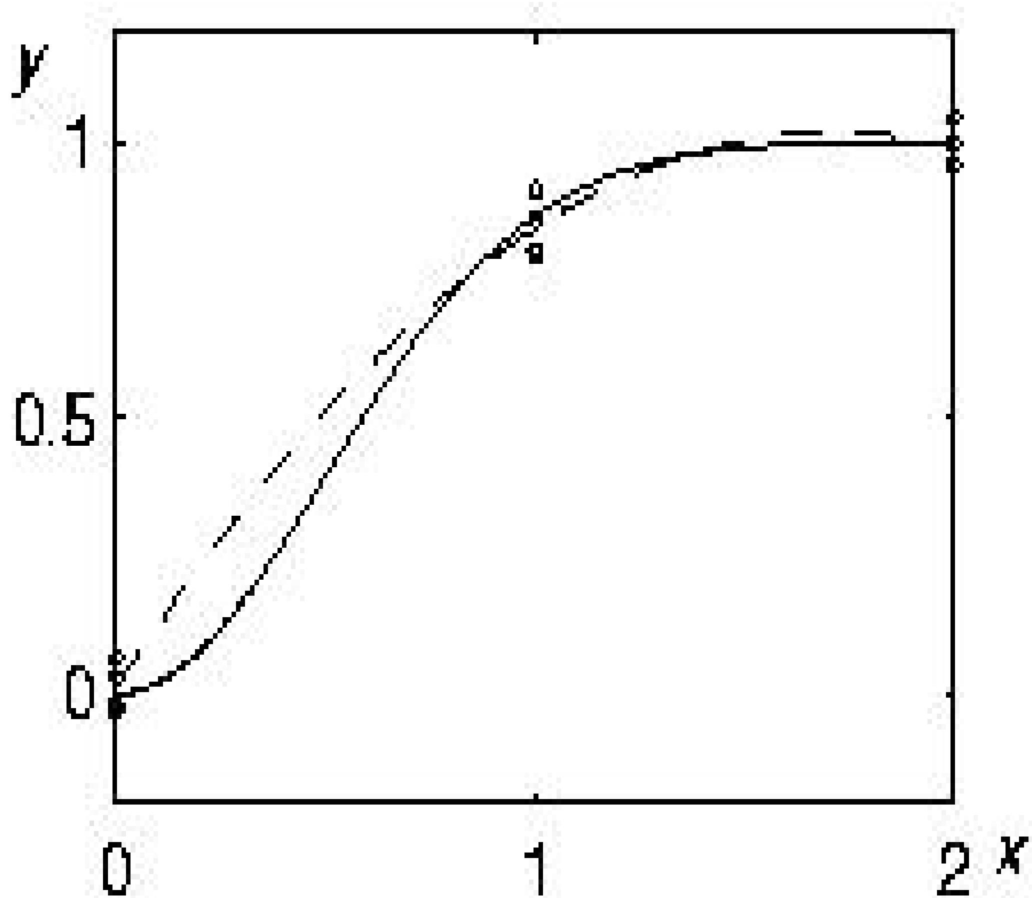
We want to design an experiment with a fixed number of runs such that we can obtain the best estimators of the parameters.

The so-called ***optimal design*** is from this idea. There are several criteria, such as *D*-optimality, *A*-optimality, *E*-optimality, etc, in theory of optimal designs. See Atkinson and Donev (1992) and Pulkelsheim (1993) for the details. When the model is

$$y(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon,$$

the corresponding *D*-optimal design is presented

D-optimal design for second-order polynomial model



c. Nonparametric regression models -- Uniform Designs

When the experimenter do not have any prior knowledge about the underlying model, a nonparametric regression model

$$y = g(x) + \varepsilon,$$

where function g is unknown, can be employed.

c. Nonparametric regression models -- Uniform Designs

We want to estimate *y(x)* at each *x*, i.e. to find an **approximate model**

$$\widehat{y} = \hat{g}(x)$$

A natural idea is to observe *y* at $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$, that are uniformly scattered in the domain, i.e., a **space filling design**, the **uniform design** is one of space filling designs.

Many smooth techniques, such as

- polynomial regression model
- kernel estimator
- Kriging models
- wavelets
- spline, B-spline
- Artificial neural networks

can be used for estimation of the function *g*.

Experiments can be implemented in

- Industrial factory
- Laboratory
- Computer

The latter is called

- Simulation Experiment, or
- Computer Experiment

Uniform Design

Uniform designs

A demostration example

Example

In an chemical experiment the experimenter chose 4 factors each having 12 levels

Step 1. Choose factors and their levels

Four factors, the amount of formaldehyde (x_1), the reaction temperature (x_2), the reaction time (x_3), and the amount of potassium carbonic acid (x_4), are under consideration. The response variable is designated as the yield (y).

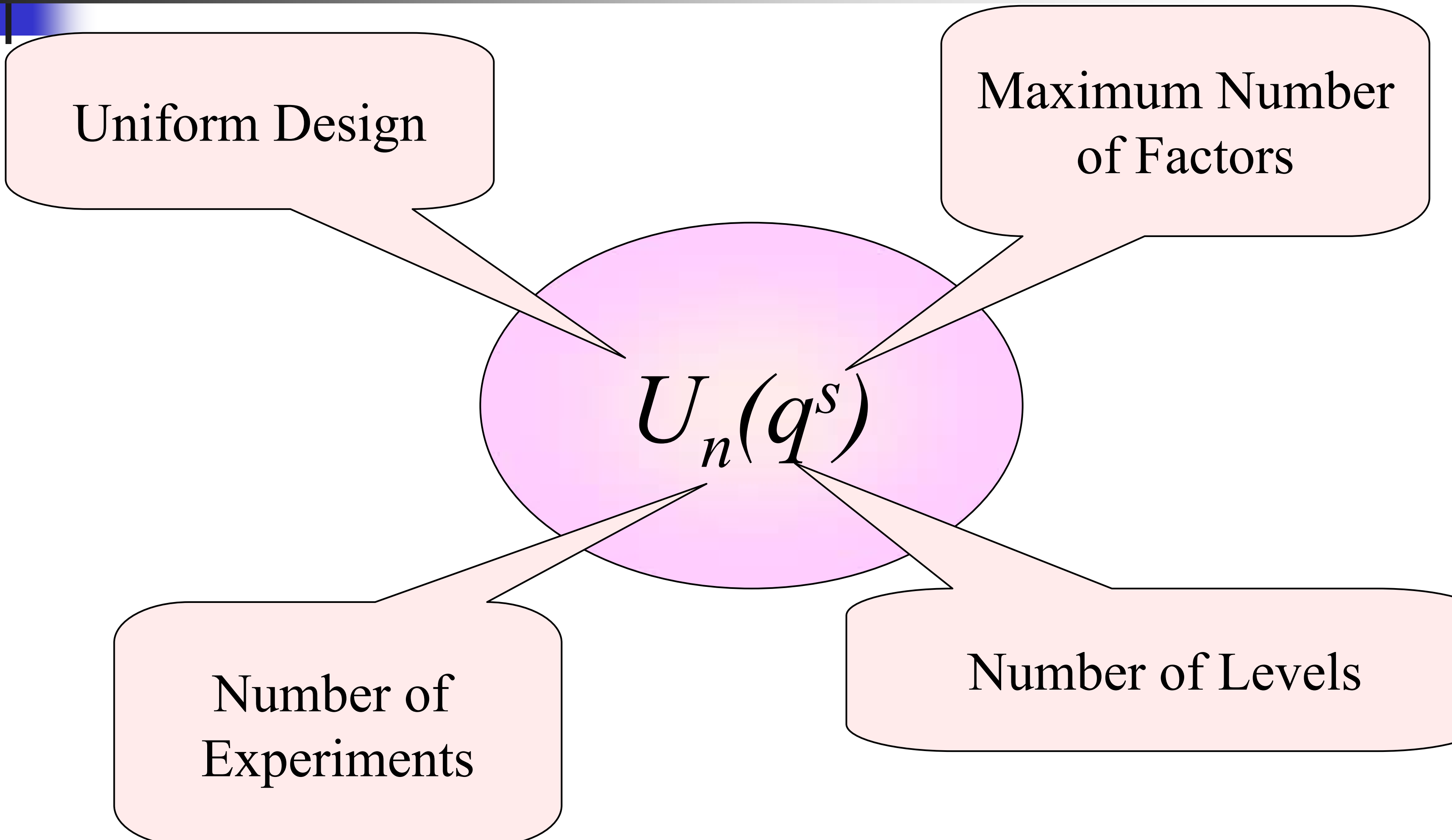
The experimental domain and levels are chosen to be

$$[1, 5.4]\times[5, 60]\times[1, 6.5]\times[15, 70]$$

- x_1: the amount of formaldehyde (*mol/mol*):
  1.0, 1.4, 1.8, 2.2, 2.6, 3.0, 3.4, 3.8, 4.2, 4.6, 5.0, 5.4
- x_2: the reaction temperature (^oC):
  5,10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60
- x_3: the reaction time (*hour*):
  1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 6.5
- x_4: the amount of potassium carbonic acid (*ml*):
  15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70

Step 2. Design

The uniform design, like the orthogonal design, can be tabulated.



http://www.math.hkbu.edu.hk/UniformDesig

$U_{12}(12^4)$

No.	1	2	3	4
1	1	10	4	7
2	2	5	11	3
3	3	1	7	9
4	4	6	5	3
5	5	11	10	11
6	6	9	8	1
7	7	4	5	12
8	8	2	3	2
9	9	7	12	8
10	10	12	6	4
11	11	8	2	10
12	12	3	9	6

$U_6(3^2 \times 2)$

No	1	2	3
1	*1*	*1*	*1*
2	*2*	*1*	*2*
3	*3*	*2*	*1*
4	*1*	*2*	*2*
5	*2*	*3*	*1*
6	*3*	*3*	*2*

- **All the levels of each column appear equally often**
- **Experimental points determined by the table uniformly are scattered on the domain.**

Step 2. Design

$U_{12}(12^4)$ ***Table***

This experiment could be arranged with a UD table of the form $U_{12}(12^4)$, where n is a multiple of 7. It turns out that the experimenter chooses $U_{12}(12^4)$ design.

No.	1	2	3	4	x_1	x_2	x_3	x_4
1	11	8	2	10	5.0	40	1.5	60
2	9	7	12	8	4.2	35	6.5	50
3	8	2	3	2	3.8	10	2.0	20
4	10	12	6	4	4.6	60	3.5	30
5	1	10	4	7	1.0	50	2.5	45
6	2	5	11	3	1.4	25	6.0	25
7	4	6	1	5	2.2	30	1.0	35
8	7	4	3	12	3.4	20	3.0	70
9	6	9	8	1	3.0	45	4.5	15
10	3	1	7	9	1.8	5	4.0	55
11	5	11	10	11	2.6	55	5.5	65
12	12	3	9	6	5.4	15	5.0	40

3. Run experiments

No.	1	2	3	4	x_1	x_2	x_3	x_4	y
1	11	8	2	10	5.0	40	1.5	60	0.1836
2	9	7	12	8	4.2	35	6.5	50	0.1739
3	8	2	3	2	3.8	10	2.0	20	0.0900
4	10	12	6	4	4.6	60	3.5	30	0.1176
5	1	10	4	7	1.0	50	2.5	45	0.0795
6	2	5	11	3	1.4	25	6.0	25	0.0118
7	4	6	1	5	2.2	30	1.0	35	0.0991
8	7	4	3	12	3.4	20	3.0	70	0.1319
9	6	9	8	1	3.0	45	4.5	15	0.0717
10	3	1	7	9	1.8	5	4.0	55	0.0109
11	5	11	10	11	2.6	55	5.5	65	0.1266
12	12	3	9	6	5.4	15	5.0	40	0.1424

Step 4. Modeling

A linear model

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4, \qquad (4.1)$$

by the least square estimation we obtain:

$$E(y) = 0.0533 + 0.0281 x_1 + 0.0010 x_2 + 0.0035 x_3 + 0.0011 x_4$$

ANOVA Table:

Type III Tests					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Stat**	**Prob > F**
x_1	1	0.018	0.018	21.8021	0.0023
x_2	1	0.0033	0.0008	3.9496	0.0872
x_3	1	0.0004	0.0004	0.515	0.4962
x_4	1	0.0046	0.0046	5.6248	0.0495

By the backward elimination

$$y = 0.0107 + 0.0289x_1$$

with $R^2 = 57.68\%$ and $S^2 = 0.0014$.

Consider the quadratic regression model

$$E(y) = \beta_0 + \sum_{i=1}^{4} \beta_i x_i + \sum_{i \le j} \beta_{ij} x_i x_j$$

with technique of selection of variables MAXR we find

$$\hat{y} = 0.0446 + 0.0029x_2 + 0.0260x_3 + 0.0071x_1x_3 + 0.000036x_2x_4 + 0.000054x_2^2 \quad (1)$$

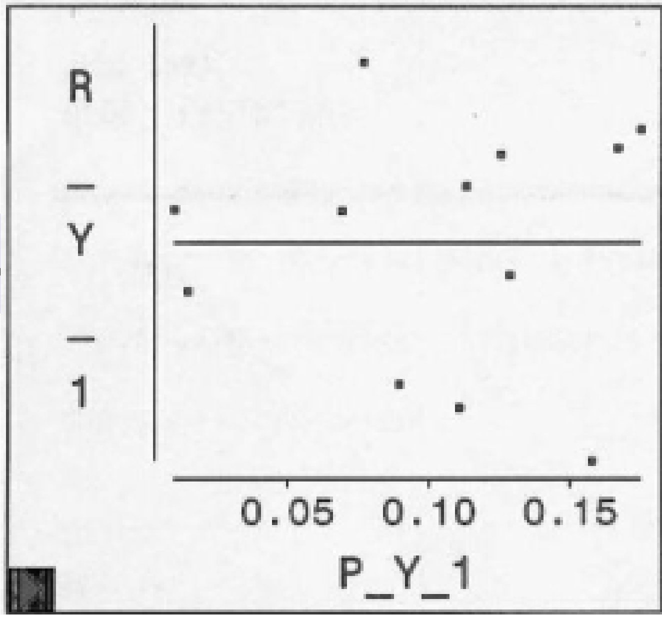
with $R^2 = 97.43\%$ and $S^2 = 0.0001$.

ANOVA Table

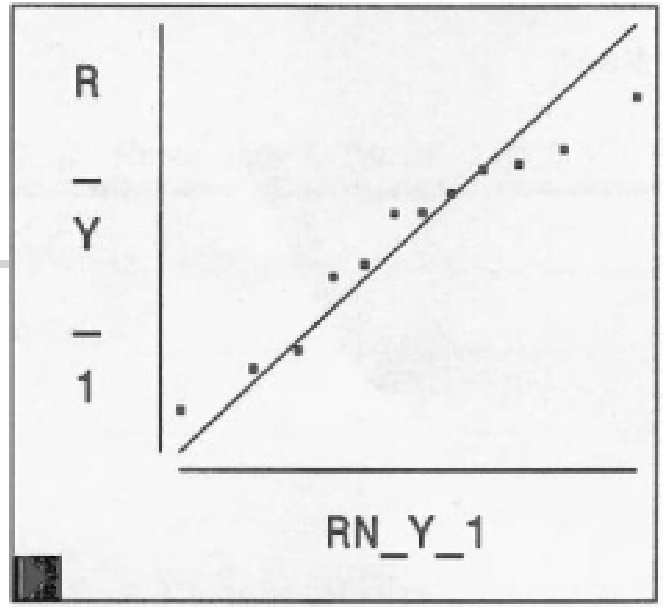
Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Stat**	**Prob > F**
Model	5	0.0323	0.0065	45.5461	0.0001
Error	6	0.0009	0.0001		
C Total	11	0.0332			

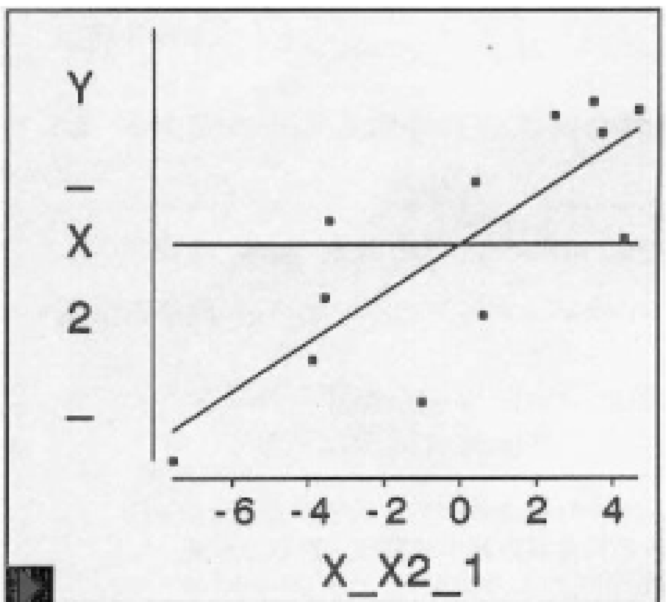
Type III Tests					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Stat**	**Prob > F**
x_2	1	0.0014	0.0014	10.1949	0.0188
x_3	1	0.0125	0.0125	88.0883	0.0001
x_1x_3	1	0.0193	0.0193	135.5636	0.0001
x_2x_4	1	0.0062	0.0062	43.6923	0.0006
x_2x_2	1	0.0024	0.0024	16.8276	0.0063

Residual Plot

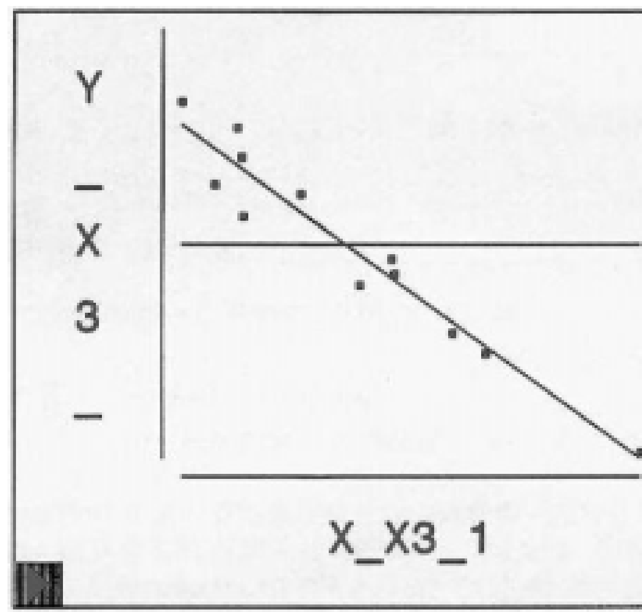


Normal Plot

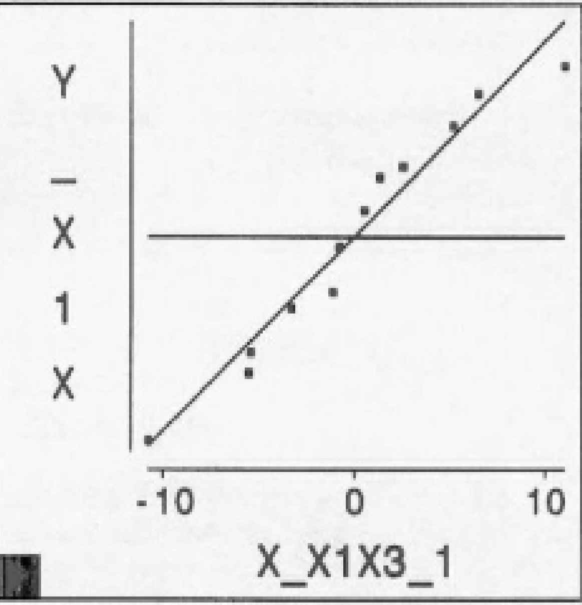


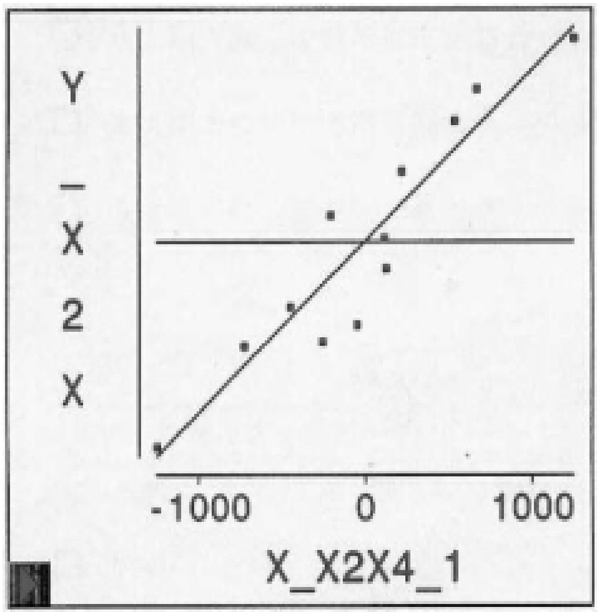


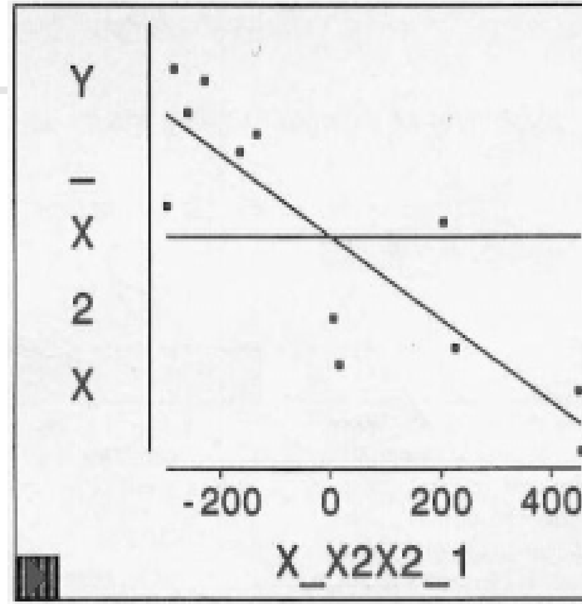
Partial Regression Plot



Partial Regression Plot







Step 5. Prediction and optimization

$$D = \{(\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}, \boldsymbol{x_4}) : 1 \le \boldsymbol{x_1} \le 5.4, 5 \le \boldsymbol{x_2} \le 60\},$$
$$1 \le \boldsymbol{x_3} \le 6.5, 15 \le \boldsymbol{x_4} \le 70\}.$$

that is to find $(\boldsymbol{x_1^*}, \boldsymbol{x_2^*}, \boldsymbol{x_3^*}, \boldsymbol{x_4^*})$ such that

$$\hat{y}(x_1^*, x_2^*, x_3^*, x_4^*) = \max_D \hat{y}(x_1, x_2, x_3, x_4),$$

where $\hat{y}(x_1, x_2, x_3, x_4)$ is given by (1).

Step 5. Prediction and optimization

By optimization algorithm, it is easily found that and the corresponding response $x_1 = 5.4, x_2 = 50.2, x_3 = 1, x_4 = 70$ is the maximum $\hat{y} = 19.3\%$.

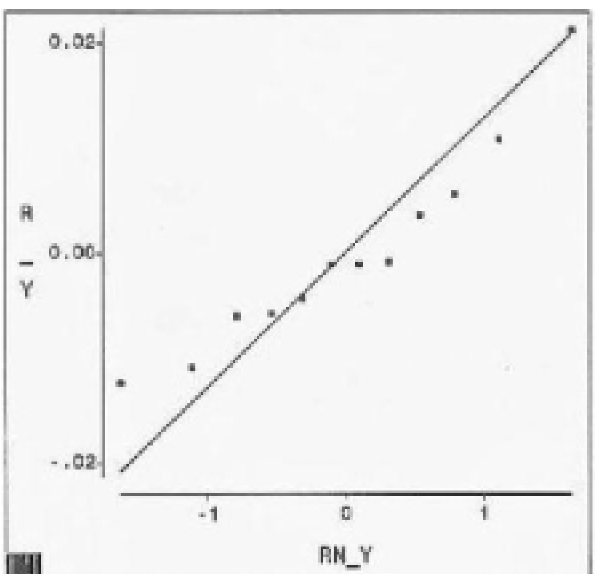
However, this is only a statistical prediction and further verification with confirmation experiments is needed.

Centered model

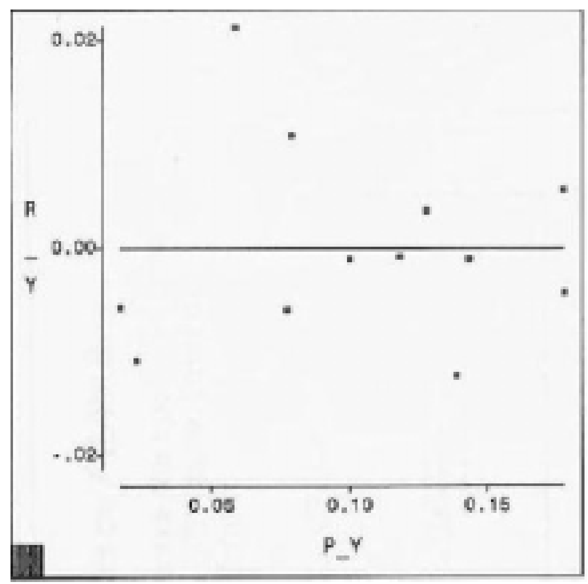
$$\hat{y} = 0.1277 + 0.0281(x_1 - 3.2) + 0.000937(x_3 - 32.5)$$
$$+ 0.00114(x_4 - 42.5) + 0.00058(x_3 - 3.75)(x_4 - 42.5)$$
$$+ 0.000082(x_2 - 32.5)^2, \quad (2)$$

with $R^2 = 97.05\%$ and $S^2 = 0.0002$.

Normal Plot



Residual Plot



Using optimization we find maximum

$$\hat{y} = 26.5\%$$

$$at \;\; \boldsymbol{x_1} = 5.4, \boldsymbol{x_2} = 43.9, \boldsymbol{x_3} = 6.5, \boldsymbol{x_4} = 70.$$

Step 6. Further experiments

The simplest way for additional experiments is to run a few experiments at $x_1 = 5.4, x_2 = 43.9, x_3 = 6.$ and $\boldsymbol{x_4} = 70.$

Many smooth techniques, such as

- polynomial regression model
- kernel estimator
- Kriging models
- wavelets
- spline, B-spline
- Artificial neural networks

can be used for estimation of the function g.

Uniform designs

Historical review

Computer Experiments
-- space filling design

In 1978 three big projects in system engineering raised the same type of problems to me. It needs one day calculation in a computer to obtain the output *y* from the given input under the true model

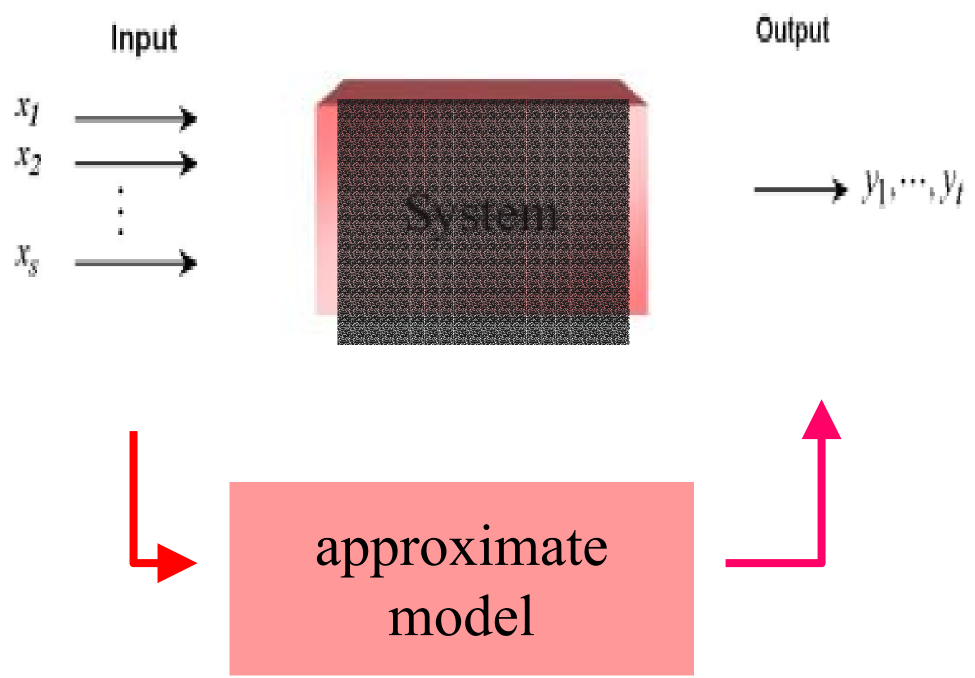
$$y = g(x_1, \cdots, x_s).$$

where the function g has no analytic formula and y is the solution of a set of differential equations.

They wanted to choose a representative set of inputs, $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$, and related output $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n$, to find a good approximate model $\hat{\boldsymbol{g}}$ that is much simpler than the true one.

$$y = \hat{g}(x_1, \cdots, x_s).$$

Computer Experiments



The number of inputs of the interest in these three projects is at least 5 and the number of levels, *q*, of each input (factor), they expected, is at least 18. Since the experiment was expensive and the computational time is long, the number of runs, *n*, wished to be within 50. That is

$$s \geq 5, \ q \geq 18, \ and \ n \leq 50.$$

We (Y. Wang and myself) proposed the uniform design.

- Fang (1980)
  Acta Math. Appl. Sinica
- Wang and Fang (1981)
  Chinese Sci. Bulletin

Applications of approximation models

- Visualization
- estimation
- optimization
- others

A comprehensive review can refer to

- Fang and Wang (1994)
  Number-Theoretic Methods in Statistics

- Fang and Hickernell (1995)
  Invited talk in the ISI 50th Session

- Fang, Lin, Winker and Zhang (2000)
  Technometrics

- Fang and Lin (2003)
  Handbook in Statistics: Statistics in Industry

Uniform designs in computer experiments

Uniform Design in Computer Experiments

Computer models are often used in science and engineering fields to describe complicated physical phenomena which are governed by a set of equation, including linear, nonlinear, ordinary, and partial differential equations. It may take a long time to find the output from the input under the true model

$$y = g(x_1, \cdots, x_s).$$

where the function g has no analytic formula.

A case study of computer experiments

In the study of the flow rate of water from an upper aquifer to a lower aquifer, the aquifers are separated by an impermeable rock layer but there is a borehole through that layer connecting them.

The model formulation is based on assumption of no groundwater gradient, steady-state flow from the upper aquifer into the borehole and from the borehole into the lower aquifer, and laminar, isothermal flow through the borehole.

The response variable *y*, the flow rate through the borehole in m^3/yr, is determined by

$$y = \frac{2\pi T_u [H_u - H_l]}{\log(\frac{r}{r_w})[1 + \frac{2LT_u}{\log(r/r_w) r_w^2 K_w} + \frac{T_u}{T_l}]}.$$

where the 8 input variables are as follows:

$r_w(m)$: radius of borehole

$r(m)$: radius of influence

$T_u(m^2/yr)$: transmissivity of upper aquifer

$T_l(m^2/yr)$: transmissivity of lower aquifer

$H_u(m)$: potentiometric head of upper aquifer

$H_l(m)$: potentiometric head of lower aquifer

$L(m)$: length of borehole

$K_w(m/yr)$: hydraulic conductivity of borehole and the domain is given by

$$r_w \in [0.05,\ 0.15],\ r \in [100,\ 50000],$$

$$T_u \in [63070,\ 115600],\ T_l \in [63.1, 116],$$

$$H_u \in [990,\ 1110],\ H_l \in [700,\ 820],$$

$$L \in [1120,\ 1680],\ K_w \in [9855,\ 12045].$$

The input variables and the corresponding output are denoted by $\boldsymbol{x} = (x_1, \cdots, x_8)$ and $y(\boldsymbol{x})$, respectively. This example has been studied by Worley (1987), An and Owen (2001) and Morris, Mitchell and Ylvisaker (1993).

A. Design of Experiment

From the inference of each variable to the output y, we sort 8 input variables into

$$r_w \geq L \geq H_u \geq H_l \geq k_w \geq T_l \geq T_u \geq r$$

and put them into three groups:

$\{R_w\},\{H_u, H_l, L, K_w\}$ and $\{T_l, T_u, r\}$.

The number of levels of each variable in these three groups is chosen as 16, 8, and 4, respectively.

A. Design of Experiment

A uniform design table $\boldsymbol{U_{32}(32^8)}$ can be found on the UD-web. By the *pseudo-level technique* a $\boldsymbol{U_{32}(16 \times 8^4 \times 4^3)}$ table can be generated and is in fact used for the study. The design and related output are given.

Uniform Design and related output

No	r_w	r	T_u	T_l	H_u	H_l	L	K_w	Y
1	0.05 (1)	33366.67(3)	63070(1)	116.00(4)	1110.00(8)	768.57(5)	1200(2)	11732.14(7)	26.18
2	0.05 (1)	100.00(1)	80580(2)	80.73(2)	1092.86(7)	802.86(7)	1600(7)	10167.86(2)	14.46
3	0.06 (2)	100.00(1)	98090(3)	80.73(2)	1058.57(5)	717.14(2)	1680(8)	11106.43(5)	22.75
4	0.06 (2)	33366.67(3)	98090(3)	98.37(3)	1110.00(8)	734.29(3)	1280(3)	10480.71(3)	30.98
5	0.06 (3)	100.00(1)	115600(4)	80.73(2)	1075.71(6)	751.43(4)	1600(7)	11106.43(5)	28.33
6	0.06 (3)	16733.33(2)	80580(2)	80.73(2)	1058.57(5)	785.71(6)	1680(8)	12045.00(8)	24.60
7	0.07 (4)	33366.67(3)	63070(1)	98.37(3)	1092.86(7)	768.57(5)	1200(2)	11732.14(7)	48.65
8	0.07 (4)	16733.33(2)	115600(4)	116.00(4)	990.00(1)	700.00(1)	1360(4)	10793.57(4)	35.36
9	0.08 (5)	100.00(1)	115600(4)	80.73(2)	1075.71(6)	751.43(4)	1520(6)	10793.57(4)	42.44
10	0.08 (5)	16733.33(2)	80580(2)	80.73(2)	1075.71(6)	802.86(7)	1120(1)	9855.00(1)	44.16
11	0.08 (6)	50000.00(4)	98090(3)	63.10(1)	1041.43(4)	717.14(2)	1600(7)	10793.57(4)	47.49
12	0.08 (6)	50000.00(4)	115600(4)	63.10(1)	1007.14(2)	768.57(5)	1440(5)	11419.29(6)	41.04
13	0.09 (7)	16733.33(2)	63070(1)	116.00(4)	1075.71(6)	751.43(4)	1120(1)	11419.29(6)	83.77
14	0.09 (7)	33366.67(3)	115600(4)	116.00(4)	1007.14(2)	717.14(2)	1360(4)	11106.43(5)	60.05
15	0.10 (8)	50000.00(4)	80580(2)	63.10(1)	1024.29(3)	820.00(8)	1360(4)	9855.00(1)	43.15
16	0.10 (8)	16733.33(2)	80580(2)	98.37(3)	1058.57(5)	700.00(1)	1120(1)	10480.71(3)	97.98
17	0.10 (9)	50000.00(4)	80580(2)	63.10(1)	1024.29(3)	700.00(1)	1520(6)	10480.71(3)	74.44
18	0.10 (9)	16733.33(2)	80580(2)	98.37(3)	1058.57(5)	820.00(8)	1120(1)	10167.86(2)	72.23
19	0.11(10)	50000.00(4)	98090(3)	63.10(1)	1024.29(3)	717.14(2)	1520(6)	10793.57(4)	82.18
20	0.11(10)	100.00(1)	63070(1)	98.37(3)	1041.43(4)	802.86(7)	1600(7)	12045.00(8)	68.06
21	0.12(11)	33366.67(3)	63070(1)	116.00(4)	990.00(1)	785.71(6)	1280(3)	12045.00(8)	81.63
22	0.12(11)	100.00(1)	98090(3)	98.37(3)	1092.86(7)	802.86(7)	1680(8)	9855.00(1)	72.54
23	0.12(12)	16733.33(2)	115600(4)	80.73(2)	1092.86(7)	734.29(3)	1200(2)	11419.29(6)	161.35
24	0.12(12)	16733.33(2)	63070(1)	63.10(1)	1041.43(4)	785.71(6)	1680(8)	12045.00(8)	86.73
25	0.13(13)	33366.67(3)	80580(2)	116.00(4)	1110.00(8)	768.57(5)	1280(3)	11732.14(7)	164.78
26	0.13(13)	100.00(1)	98090(3)	98.37(3)	1110.00(8)	820.00(8)	1280(3)	10167.86(2)	121.76
27	0.14(14)	50000.00(4)	98090(3)	63.10(1)	1007.14(2)	820.00(8)	1440(5)	10167.86(2)	76.51
28	0.14(14)	33366.67(3)	98090(3)	116.00(4)	1024.29(3)	700.00(1)	1200(2)	10480.71(3)	164.75
29	0.14(15)	50000.00(4)	63070(1)	116.00(4)	990.00(1)	785.71(6)	1440(5)	9855.00(1)	89.54
30	0.14(15)	50000.00(4)	115600(4)	63.10(1)	1007.14(2)	734.29(3)	1440(5)	11732.14(7)	141.09
31	0.15(16)	33366.67(3)	63070(1)	98.37(3)	990.00(1)	751.43(4)	1360(4)	11419.29(6)	139.94
32	0.15(16)	100.00(1)	115600(4)	80.73(2)	1041.43(4)	734.29(3)	1520(6)	11106.43(5)	157.59

B. Quadratic regression model

For the modeling, many authors proposed a number of methods. When the function *g* is a periodic, a Fourier regression model is recommended.

The spatial modeling technique of kriging (Koehler and Owen (1996)) is based on a stationary Gaussian stochastic process and the Bayesian approach (Sacks, Welch, Mitchell and Wynn (1989) and Morris, Mitchell and Ylvisaker (1993)) uses the prior information.

$$\widehat{\log(y)} = 4.1560 + 1.9903(\log(r_w) + 2.3544)$$
$$- 0.0007292(L - 1400)$$
$$- 0.003554(H_l - 760)$$
$$+ 0.0035068(H_u - 1050)$$
$$+ 0.000090868(K_w - 10950)$$
$$+ 0.000015325(H_u - 1050)(H_l - 760)$$
$$+ 0.00000026487(L - 1400)^2$$
$$- 0.0000071759(H_l - 760)^2$$
$$- 0.0000068021(H_u - 1050)^2$$
$$- 0.00087286(\log(r) - 8.8914)$$

This model has an *MSE=0.2578156*.

ANOVA Tables (SAS Output)

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Stat	Prob > F
Model	9	13.1368	1.4596	108755.062	0.0001
Error	22	0.0003	1.342E-05		
C Total	31	13.1371			

Type III Tests					
Source	DF	Sum of Squares	Mean Square	F Stat	Prob > F
LOG_RW	1	10.3539	10.3539	771441.372	0.0001
L	1	0.4083	0.4083	30420.3025	0.0001
HL	1	0.4714	0.4714	35124.6151	0.0001
HU	1	0.3239	0.3239	24134.5154	0.0001
KW	1	0.0732	0.0732	5457.3163	0.0001
HUHL	1	0.0085	0.0085	630.7976	0.0001
L_2	1	0.0008	0.0008	61.5301	0.0001
HL_2	1	0.0013	0.0013	96.0652	0.0001
HU_2	1	0.0013	0.0013	95.6290	0.0001

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Stat	Prob > F
Model	10	13.1369	1.3137	126225.959	0.0001
Error	21	0.0002	1.041E-05		
C Total	31	13.1371			

Type III Tests					
Source	DF	Sum of Squares	Mean Square	F Stat	Prob > F
LOG_RW	1	10.2746	10.2746	987235.474	0.0001
L	1	0.3090	0.3090	29694.9581	0.0001
HL	1	0.4615	0.4615	44347.0990	0.0001
HU	1	0.2601	0.2601	24990.9369	0.0001
KW	1	0.0727	0.0727	6988.6817	0.0001
HUHL	1	0.0076	0.0076	725.7697	0.0001
L_2	1	0.0009	0.0009	81.9494	0.0001
HL_2	1	0.0013	0.0013	120.5167	0.0001
HU_2	1	0.0012	0.0012	111.9492	0.0001
LOG_R	1	7.671E-05	7.671E-05	7.3711	0.0130

D. Comparisons among different designs and models

We compare the performances of different designs:

- *Latin hypercube design*
- *maximin design*
- *maximin Latin hypercube design*
- *modified maximin design*
- *uniform design*

The latter used the *Latin hypercube design*, *maximin design*, *maximin Latin hypercube design* and *modified maximin design*.

For comparing different models they used the *mean square error* (MSE) as the criterion, i.e.,

$$MSE = \frac{1}{N}\sum_{k=1}^{N}(y(x_k) - \hat{y}(x_k))^2,$$

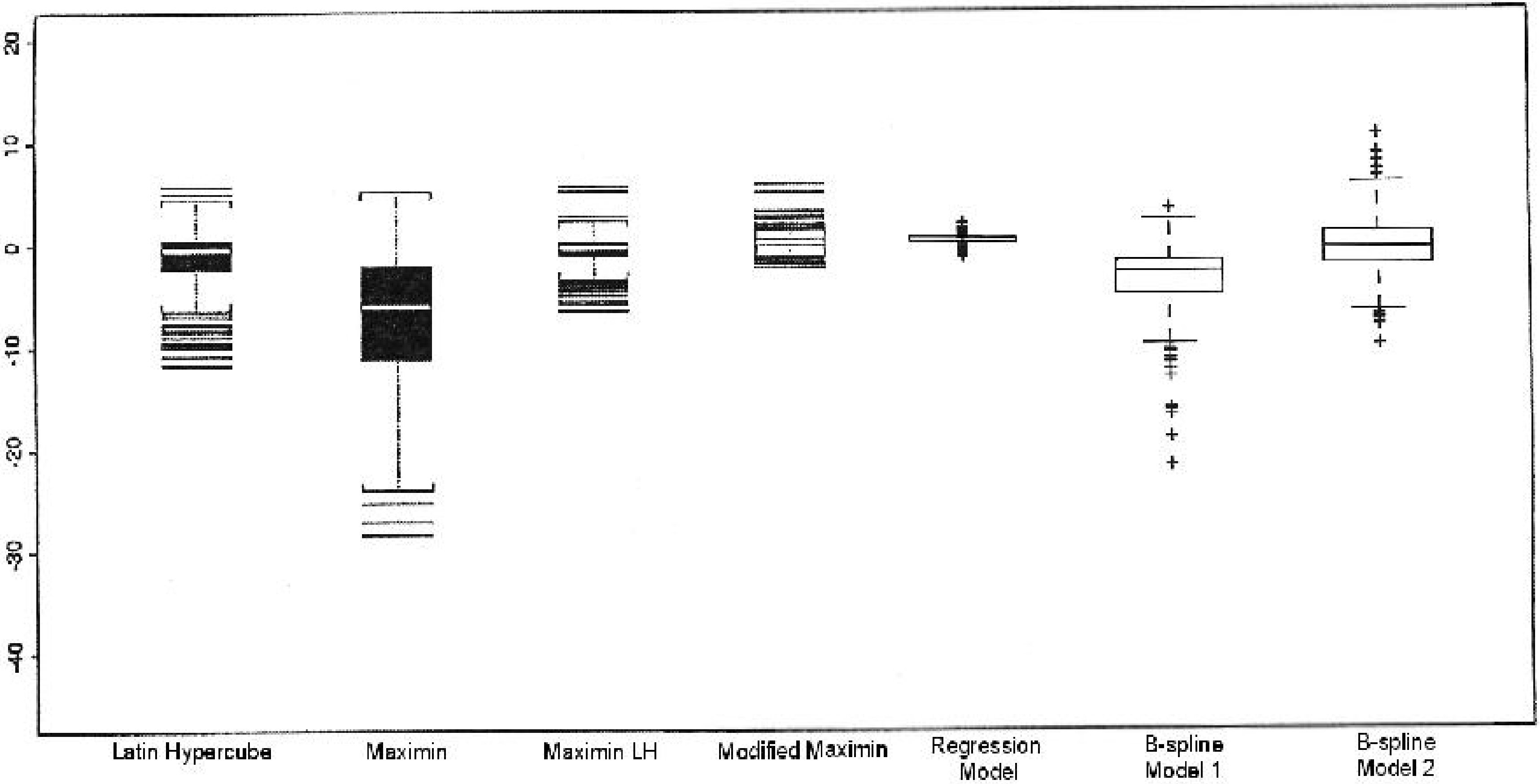
where $\boldsymbol{x}_k, i = 1, \cdots, N$ are randomly chosen from the domain and $\hat{y}(\boldsymbol{x}_k)$ is its predicted value under the underlying model. The value of N is chosen to be greater than 1000.

D. Comparisons among different designs and models

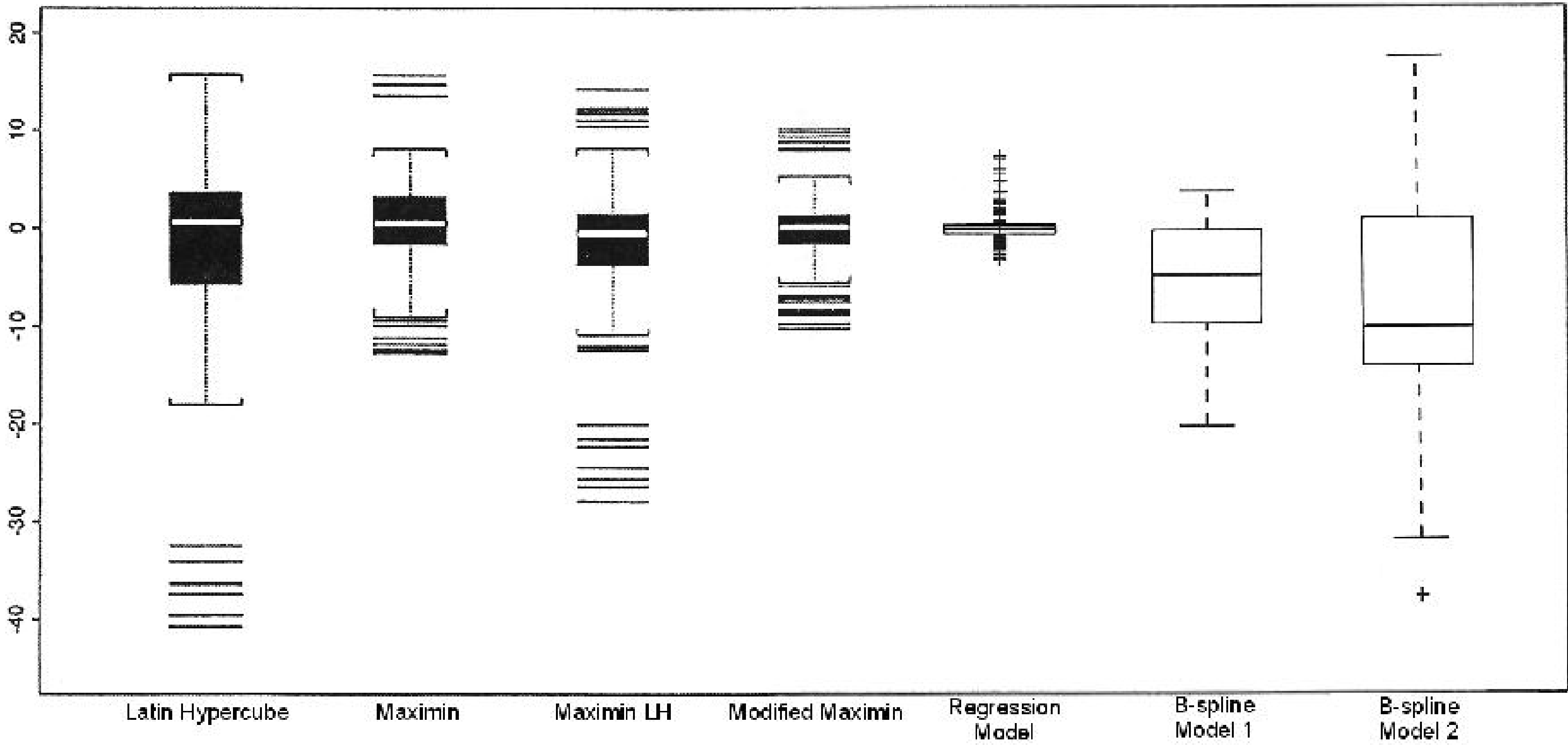
For comparing four different designs, Morris, Mitchell and Ylvisaker (1993) considered prediction errors at 400 random samples in the domain and at the 256 corner points of domain. They plot the prediction errors in two separated figures. Obviously, the *B*-spline model has large errors for the 256 corner points. This bias may be resulted from small number of levels for some input variables.

Ho and Xu (2000) employed the table $U_{30}(30^8)$ to design 30 level-combinations with the *B*-spline model mentioned above for modeling.

Prediction Errors at 400 Random Samples for Seven Design/Model

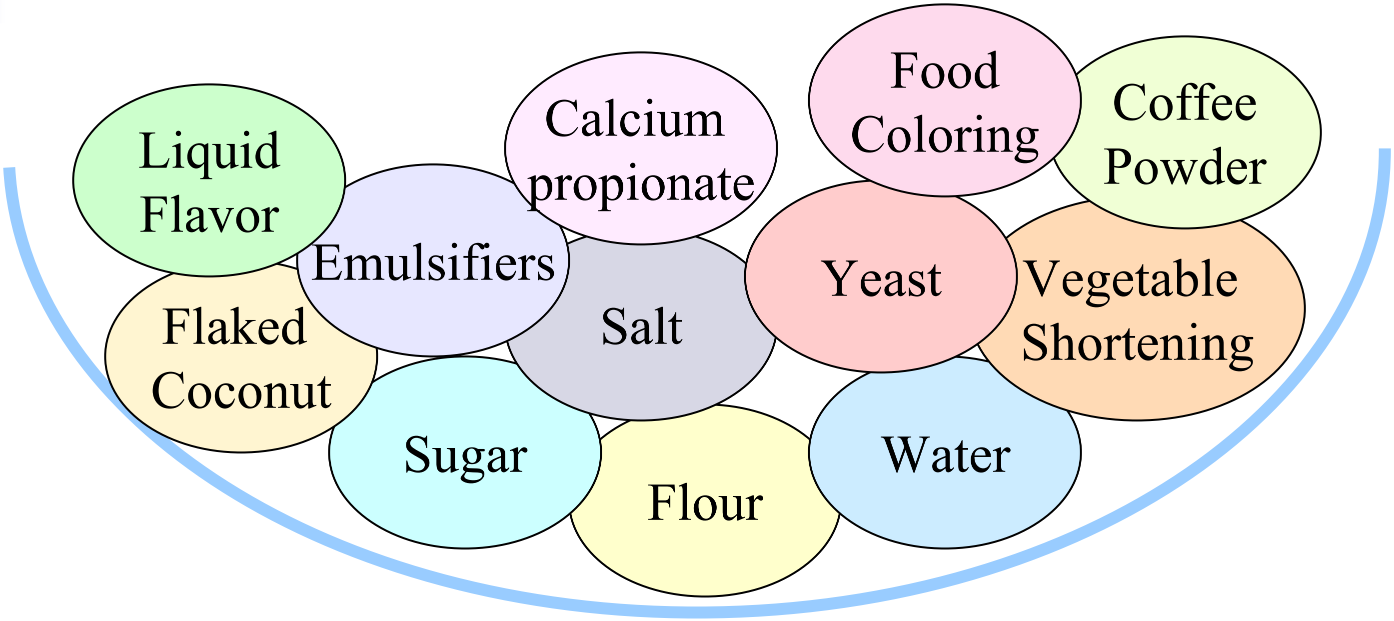


Prediction Errors at 256 Corner Points for Seven Design/Models



Uniform designs with mixtures

Many products are formed by mixing two or more ingredients together.



Coffee Bread

EXPERIMENTS WITH MIXTURES

The UD can be utilized as

- a fractional factorial design
- a design of computer experiments
- a robust design
- a design with mixtures

UD Society

- The UD has been used since 1980.
- Several conference and workshops were held in the past years
- More than 300 case studies of the use of UD during 1994 - 2000
- There is a nationwide society:
  Uniform Design Association of China
  since 1995

The First Conference, Beijing, 1995

Hong Kong Symposium, 1999

Hong Kong Symposium, 1999

Xian Conference, 2001

Comments on uniform design

- Another approach to space-filling design using methods from number theory is briefly described in Exercise 7.7. This approach is reviewed by Fang, Wang and Bentler (1994) and its application in design of experiments discussed in Ch. 5 of Fang and Wang (1994). In the computer science literature the method is often called quasi-Monte Carlo sampling; see Neiderreiter (1992).
  --D.R. Cox and N. Reid (2000), *The Thoery of the Design of Experiments.*

Comments on uniform design

- Another type of space-filling design specifies points in the design space using methods from number theory. The resulting design is called a uniform, or uniformly scattered design.
  --D.R. Cox and N. Reid (2000), *The Thoery of the Design of Experiments.*

Comments on uniform design

- An important class of designs are so-called lattices. These have received considerable attention in number theory under ths heading of low discrepancy sequences A principal text is Niederreiter (1992) and Fang and Wang (1994) (an their earlier work) make a considerable contribution in applications to statistics, including design. -- R.A. Bates, R.J. Buck, E. Riccomagno and H.P. Wynn (1996), *JRSS-B,* **58**, 77-94 (with discussion).

A case study by R.A. Bates, R.J. Buck, E. Riccomagno and H.P. Wynn (1996)

- They considered a two-dimensional exercise for comparing Latin hypercube design, modified Latin hypercube design and lattice designs. They conclude:

- Some conclusions are that the lattice designs do surprisingly well and a good integer lattice is robust against changes of criterion.

Comments on uniform design, by C.F.Jeff Wu and M. Hamada, p.445, “Experiments planing, analysis, and parameter design optimization.

- If some of the noise factors have more than three levels, the run size of the orthogonal array for the noise factors may be too large. An alternative is to employ a smaller plan with uniformly spread point for the noise factors. These plans include Latin hypercube sampling (Koehler and Owen, 1996) and ``uniform’’ designs based on number-theoretic methods (Fang and Wang, 1994). Since the noise array is chosen to represent the noise variation, uniformity may be considered to be a more important required than orthogonality.

The UD entertains several advantages. It can explore relationships between the response and the factors with a reasonable number of runs and is shown to be robust to the underlying model specification.

Wiens D P, (1991) *Stat. & Prob. Letters*.

Hickernell F J (1999) *Stat. & Prob. Letters*.

Xie M Y and Fang K T, (2000) *JSPI*.

Gordon Research Conference, Williams College, MA, USA

July 22-27, 2001

Uniform Design for Simulated Experiments,
By Kai-Tai Fang, 70 minutes

Uniform Design and Its Applications to Chemistry and Chemical Engineering,
By Yizhen Liang, 30 minutes

Discussion, 60 minutes
Chairrd by Dennis Lin

9 topics in every two years

Uniform design

- Flexibility
- Easy to use
- Easy to understand

Merits of the UD method

- **Space filling:** it is capable of producing samples with high representativeness in the experimental domain;
- **Robustness:** it imposes no strong assumption on the model, and is against changes of model in a certain sense;
- **Multiple levels:** it allows the largest possible amount of levels for each factor.

Conclusion remarks

The UD can be utilized as

- a fractional factorial design
- a design of computer experiments
- a robust design
- a design with mixtures

Thank you!

Contact information

Kai-Tai Fang
Hong Kong Baptist University

Tel:3411-7025
Fax:3411-5811

Email:*ktfang@hkbu.edu.hk*
http://www.math.hkbu.edu.hk/UniformDesign

Kai-Tai Fang

- Chair Professor of Mathematics, Hong Kong Baptist University and

  Professor, The Chinese Academia of Sciences.
- Research
  - Multivariate Analysis & Distribution Theory
  - Number-theoretic Methods (Quasi Monte Carlo)
  - Uniform Design
- Fellows: ISI, IMS, ASA
- President, The Uniform Design Assocaition
- Council: ISI, ICSA, HKIS, HKSS
- Visiting Professor
  - Yale, Stanford, UCLA, UNC-Chapel Hill, Swiss Federal Inst of Tech, Ford Motor Company, etc.