

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,100

Open access books available

149,000

International authors and editors

185M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter

Designs for Screening Experiments with Quantitative Factors

Nam-Ky Nguyen, Stella Stylianou, Tung-Dinh Pham and Mai Phuong Vuong

Abstract

Most screening experiments in chemometrics and science are quantitative, i.e. continuous factors. These factors should be 3-level and the designs for these experiments should also be 3-level. However, popular designs for screening experiments are still Plackett-Burman designs (PBDs) and 2-level fractional factorial designs (FFDs) such as resolution III and resolution IV FFDs. This chapter introduces the conference matrices as an alternative to PBDs and resolution III FFDs and definitive screening designs, a conference matrix-based class of designs, as an alternative to resolution IV FFDs. A table of conference matrices of up to order 32 and examples are also provided for illustration.

Keywords: conference matrices, definitive screening designs, fractional factorial designs, Plackett-Burman designs, response surface designs, screening designs

1. Introduction

Screening experiments are used at the initial stage of experimentation and aim at identifying the dominant main effects out of a large set of potentially active factors. The benefit of the screening approach is the use of a cost-effective design and process to separate the influential variables from the non-influential ones. By using the active effects that have been identified by the screening process, the research can run additional follow-up experiments to fit higher-order effects and build a better and more complex model. Screening is traditionally performed by applying a linear model using a 2-level FFD. When the screening process involves quantitative factors, other designs with better properties are also available in the recent literature.

Consider a 2^{15-11} experiment conducted by Poorna & Kulkarnin [1] (hereafter abbreviated as PK) to investigate 15 2-level factors, which might affect inulinase production. These factors come from four carbon sources: **A** Inulin (%), **B** Fructose (%), **C** Glucose (%), **D** Sucrose (%); four organic nitrogen sources: **E** Corn steep liquor (%), **F** Peptone (%), **G** Urea (%), **H** Yeast extract (%); four inorganic nitrogen sources: **J** Corn steep liquor (%), **K** Peptone (%), **L** Urea (%), **M** Yeast extract (%); and three other parameters: **N** Trace element solutions (mL), **O** Inoculum level (10^6 spores/mL), **P** (pH). The two responses are inulinase activity (units/mL) at

60 hours and dry weight biomass (mg/mL). This experiment is also summarized in Example 6.4 of [2].

The design for the PK experiment is a 2_{III}^{15-11} (a resolution III FFD for 15 factors in 16 runs) given in **Table 1(a)**. The 11 design generators for this experiment are: **E** = -**ABCD**, **F** = **BCD**, **G** = **ABC**, **H** = -**CD**, **I** = -**BD**, **J** = **ABD**, **K** = **ACD**, **L** = -**AC**, **M** = -**AD**, **N** = -**AB**, and **O** = -**BC**. For a resolution III design, no main effects (MEs) are aliased with any other MEs, but MEs are aliased with 2-factor interactions (2FIs). For this design, each ME is aliased with seven 2FIs. For example -**A** = **BN** = **CL** = **DM** = **EF** = **GO** = **HK** = **IJ** and -**D** = **AM** = **BI** = **CH** = **EG** = **FO** = **JN** = **KL**. An alternative design with similar number of runs, whose MEs are pairwise orthogonal and are not fully aliased with 2FIs, will be presented in this chapter.

Let us examine another experiment on the human blood formation that originates in hematopoietic stem cells (HSC) and hematopoietic progenitor cells (HPC). Due to the complexity of clinical use of cells in serum medium, Yao et.al. [3] set up a 2^{8-4} experiment to screen out the most important serum substitutes affecting the growth of HSC and HPC among eight kinds of compounds: **A** Albumax I (10g/l), **B** BSA (10g/l), **C** TF(0.4g/l), **D** Glutamine (2mM), **E** HC (1mg/l), **F** Peptone (1g/l), **G** 2-ME

(a)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	1	1	1	-1	1	1	-1	-1	1	1	-1	-1	-1	-1
1	1	1	-1	1	-1	1	1	1	-1	-1	-1	1	-1	-1
-1	1	1	1	1	1	-1	-1	-1	-1	-1	1	1	1	-1
1	-1	1	-1	-1	1	-1	1	-1	1	-1	-1	1	1	1
-1	1	-1	1	-1	-1	1	1	-1	-1	1	-1	1	1	1
1	-1	1	1	1	-1	-1	-1	1	-1	1	-1	-1	1	1
1	1	-1	-1	-1	1	-1	-1	1	-1	1	1	1	-1	1
-1	1	1	-1	-1	-1	-1	1	1	1	1	1	-1	1	-1
-1	-1	1	1	-1	-1	1	1	-1	1	-1	1	1	-1	1
1	-1	-1	-1	1	-1	1	-1	-1	1	1	1	1	1	-1
-1	1	-1	-1	1	1	1	-1	1	1	-1	-1	-1	1	1
-1	-1	1	-1	1	1	1	1	-1	-1	1	1	-1	-1	1
-1	-1	-1	1	1	1	-1	1	1	1	1	-1	1	-1	-1
1	-1	-1	1	-1	1	1	1	1	-1	-1	1	-1	1	-1
1	1	-1	1	1	-1	-1	1	-1	1	-1	1	-1	-1	1
-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

(b)

A	B	C	D	E	F	G	H
-1	-1	-1	-1	-1	-1	-1	-1
1	-1	-1	-1	-1	1	1	1
-1	1	-1	-1	1	-1	1	1
1	1	-1	-1	1	1	-1	-1
-1	-1	1	-1	1	1	1	-1
1	-1	1	-1	1	-1	-1	1
-1	1	1	-1	-1	1	-1	1
1	1	1	-1	-1	-1	1	-1
-1	-1	-1	1	1	1	-1	1
1	-1	-1	1	1	-1	1	-1
-1	1	-1	1	-1	1	1	-1
1	1	-1	1	-1	-1	-1	1
-1	-1	1	1	-1	-1	1	1
1	-1	1	1	-1	1	-1	-1
-1	1	1	1	1	-1	-1	-1
1	1	1	1	1	1	1	1

Table 1. (a) Poorna & Kulkarni's 2^{15-11} experiment, (b) Yao et al.'s 2^{8-4} experiment.

(55 μ M) and H Insulin (10 μ g/ml) for hematopoietic ex vivo expansion culture. Among them, the first three factors are non-hormonal proteins, Insulin and HC are both hormonal proteins, 2-ME is an antioxidant molecule, and Glutamine is an amino acid.

The design for the experiment in the second example is a 2_{IV}^{8-4} (a resolution IV FFD for eight factors in 16 runs) given in **Table 1(b)** where level -1 means no addition, and 1 indicates a specified concentration of the compound. The four design generators for this experiment are: $E = BCD$, $F = ACD$, $G = ABC$, $H = ABD$. For a resolution IV design, no MEs are aliased with any other ME or 2FIs, but 2FIs are aliased with other 2FIs. For this design, each of 2FIs is aliased with three other 2FIs: $AB = EF = CG = DH$, $AC = DF = BG = EH$, $AD = CF = EF = BH$, $AE = BF = DG = CH$, $AF = CD = BE = GH$, $AG = BC = DE = FH$, and $AH = BD = CE = FG$. Again, in this chapter we will present an alternative design, whose MEs are orthogonal to other MEs and 2FIs and whose 2FIs are not fully aliased with other 2FIs.

Motivated by the above examples, this chapter introduces the use of conference matrices and conference matrix-based designs, including the popular definitive screening designs (DSDs), as an alternative to Plackett-Burman designs or PBDs [4] and resolution III and IV FFDs.

2. Conference matrices and its use

A conference matrix C of order m is an $m \times m$ $(0, \pm 1)$ -matrix with zero diagonal satisfying the condition $CC' = (m - 1) I$, where I is the identity matrix. A conference matrix is said to be *normalized* if all entries in its first row and first column are 1 (except the (1,1) entry, which is 0). Removing the first row and the first column of a normalized conference matrix yields its *core*. A conference matrix is said to be *skew-symmetric* if $C = -C'$. It is conjectured that a conference matrix C exists for all $m \equiv 2 \pmod{4}$ as long as $m - 1$ is a sum of two squares. Examples of non-existent conference matrices are the ones of size 2, 34 and 58. More information about the conference matrices can be found in Section 6.1 of [5].

A large number of conference matrices can be constructed by the single cyclic generators. **Table 2** displays two generating vectors for the conference matrices with $m \leq 32$. To generate the conference matrix for $m = 8$, for example, we use the generating vector $(0 + + - + - -)$ to generate its core and then augment it with 0 in the (1,1) entry, -1 in the remaining entries of the first row, and $+1$ in the remaining entries in the first column. Note that if we replace the first element of this generating vector, i.e. 0 by $+1$, we have the generating vector for the PBD with eight runs.

Table 3 displays the conference matrices for $m = 6, 8, 10, 12, 14, 16, 26$ and 28 . Note that, the conference matrices for $m = 10, 16, 26$ and 28 cannot be generated by the cyclic generators. Unlike conference matrices of order $m = 10, 16, 26$ and 28 in [6], the numbers of 1's and -1 's in each column from 2 to m differ by only one. A conference matrix of order $2m$ can be constructed from a conference matrix of order m by the following equation:

$$\begin{pmatrix} C & -(C + I)' \\ C + I & C' \end{pmatrix} \quad (1)$$

Here, C in (1) is a conference matrix of order m , and I is the identity matrix. We use (1) to construct the conference matrix of order 16 in **Table 3** from the conference matrix of order 8 in this table.

<i>m</i>	Generating vectors
4	0 - +
6	0 - + + -
8†	0 + + - + - -
10	See Table 3.
12†	0 + - + + + - - - + -
14	0 + - + + - - - - + + - +
16	See Table 3.
18	0 - - + - + + + - - + + + - + - -
20†	0 + - - + + + + - + + - - - - + + -
22	Not exist.
24†	0 + + + + - + - + + - - + + - - + - - - -
26	See Table 3.
28	See Table 3.
30	0 + - - + + + + - + - - - + - - - - + + + + - - +
32	0 + + - + + - + + + + - - - + - + - + + + - - - - + - - - -

†Obtained from the generating vector for a Plackett-Burman design.
The remaining generating vectors are from Table 1 of [6].

Table 2.
Generating vectors for conference matrices with $m \leq 32$.

One of the most popular use of the conference matrices is to construct *definitive screening designs* (DSDs), which are the 3-level designs introduced in [7] for studying quantitative factors. The design matrix **D** for a DSD can be written as:

$$\begin{pmatrix} \mathbf{C} \\ \mathbf{0} \\ -\mathbf{C} \end{pmatrix} \tag{2}$$

where **C** is a constituent $m \times m$ $(0, \pm 1)$ -matrix with zero diagonal, $-\mathbf{C}$ is the foldover fraction of **C**, and **0** is a row vector of 0's. Note that **0** can contain more than one row vector of 0's.

The model for a 3-level screening design such as a DSD is:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{3}$$

where **y** is the response vector; **X** is the model matrix of size $n \times p$ with $p = 1 + 2m + \binom{m}{2}$; **β**'s is the vector of parameters to be estimated; **ε** is the error vector with components assumed to be independent and identically distributed (iid) $N(0, \sigma^2)$. Let d_{ui} , ($u = 1, \dots, n; i = 1, \dots, m$) be the entry in the u th row and i th column of the design matrix **D**. The u th row of **X** can be written as $(1, d_{u1}, \dots, d_{um}, d_{u1}^2, \dots, d_{um}^2, d_{u1}d_{u2}, \dots, d_{u(m-1)}d_{um})$. The terms in each vector correspond to the intercept, MEs, quadratic effects (QEs) and 2FIs.

DSDs have the following desirable properties:

- i. The design is mean orthogonal;
- ii. The number of runs is $n = 2m + 1$, i.e. saturated for estimating the intercept, m MEs and m QEs;

3. Discussion

The suggested alternative design for the 2_{III}^{15-11} FFD used in the PK experiment is presented in **Table 4(a)**. This design contains columns 2–16 of the conference matrix of order 16 in **Table 3** and a row of 0's. The 2_{III}^{15-11} FFD has 16 runs, and the suggested alternative has 17 runs. This type of saturated or near-saturated designs is used when the experimental resource is expensive.

To have a complete picture of the aliasing patterns of the PK design and the suggested one, we show the correlation cell plots (CCPs) of the two designs **Figure 1(a)** and **(b)**. These CCPs, proposed by [7], display the magnitude of the correlation (in terms of the absolute values) between m MEs and $\binom{m}{2}$ 2FIs in screening designs. The colour of each cell ranges from white (no correlation) to dark (correlation of 1, which means full aliases). It can be seen that while the cells in CCP for the FFD in **Figure 1(a)** are either white or

(a)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
0	1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1
-1	0	1	1	-1	1	-1	1	-1	-1	1	1	-1	1	-1
-1	-1	0	1	1	-1	1	1	-1	-1	-1	1	1	-1	1
1	-1	-1	0	1	1	-1	1	1	-1	-1	-1	1	1	-1
-1	1	-1	-1	0	1	1	1	-1	1	-1	-1	-1	1	1
1	-1	1	-1	-1	0	1	1	1	-1	1	-1	-1	-1	1
1	1	-1	1	-1	-1	0	1	1	1	-1	1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1	0	1	1	1	1	1	1	1
1	1	1	-1	1	-1	-1	-1	0	-1	-1	1	-1	1	1
-1	1	1	1	-1	1	-1	-1	1	0	-1	-1	1	-1	1
-1	-1	1	1	1	-1	1	-1	1	1	0	-1	-1	1	-1
1	-1	-1	1	1	1	-1	-1	-1	1	1	0	-1	-1	1
-1	1	-1	-1	1	1	1	-1	1	-1	1	1	0	-1	-1
1	-1	1	-1	-1	1	1	-1	-1	1	-1	1	1	0	-1
1	1	-1	1	-1	-1	1	-1	-1	-1	1	-1	1	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

(b)

A	B	C	D	E	F	G	H
0	-1	-1	-1	-1	-1	-1	-1
1	0	1	1	-1	1	-1	-1
1	-1	0	1	1	-1	1	-1
1	-1	-1	0	1	1	-1	1
1	1	-1	-1	0	1	1	-1
1	-1	1	-1	-1	0	1	1
1	1	-1	1	-1	-1	0	1
1	1	1	-1	1	-1	-1	0
0	0	0	0	0	0	0	0
0	1	1	1	1	1	1	1
-1	0	-1	-1	1	-1	1	1
-1	1	0	-1	-1	1	-1	1
-1	1	1	0	-1	-1	1	-1
-1	-1	1	1	1	0	-1	1
-1	1	-1	1	1	0	-1	-1
-1	-1	1	-1	1	1	0	-1
-1	-1	-1	1	-1	1	1	0

Table 4. (a) Suggested design for Poorna & Kulkarni's 2^{15-11} experiment, (b) Suggested design for Yao et al.'s 2^{8-4} experiment.

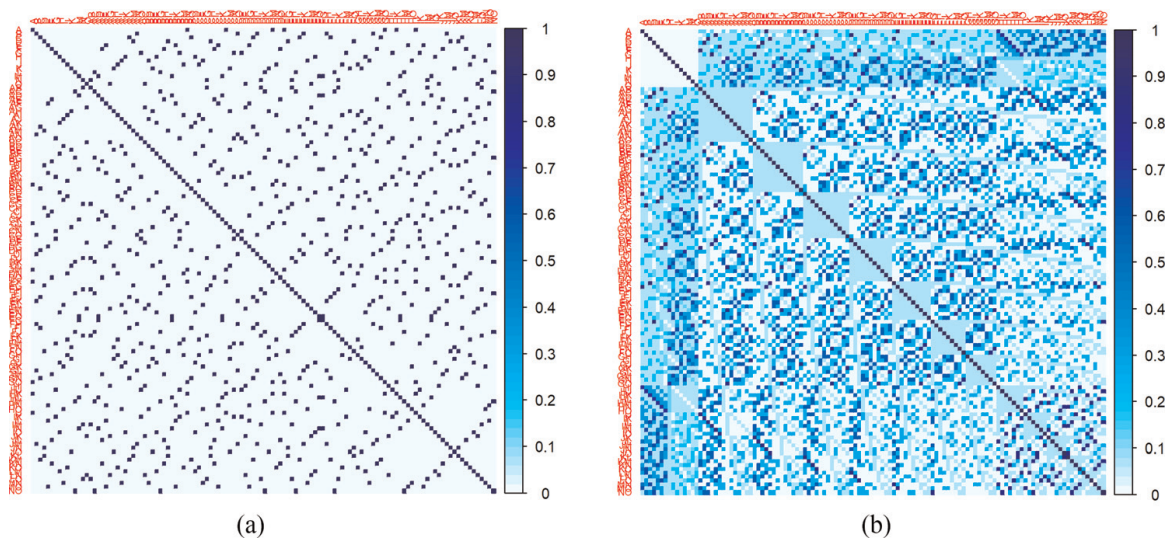


Figure 1. CCPs for (a) a 2^{15-11}_{III} FFD used in the Poorna & Kulkarni's experiment and (b) a suggested design for 15 factors in 17 runs in (Table 4(a)).

dark, there are no dark cells in the suggested design in **Figure 1(b)**, meaning none of the MEs (or 2FIs) are fully aliased with the other 2FIs. Altogether, there are 420 dark cells (which represent full aliases) in the upper/lower diagonal portion of the CCP in **Figure 1(a)**: 105 between the MEs and 2FIs, and 350 among the 2FIs.

Although the first-order D-efficiency of the 2^{15-11}_{III} FFD is higher than that of the suggested design (1 vs. 0.886), and there is a small correlation among the MEs of the suggested design ($|r| = 0.004$), there are at least three reasons for the researchers to choose the latter:

- i. Like the resolution III FFD or the PBD, the half fraction of a 3-level DSD might give researchers conclusions similar to the one obtained when the full DSD is used. Consider the data given in **Table 2** in [8] collected from a DSD for 9 factors in 21 runs (one run is a centre run). Analysing the data using the main-effects model, we found two factors 2 and 3 significant at 10% level and two factors 7 and 8 significant at 1% level with the adjusted $R^2 = 0.6775$. Repeating the analysis with the first half fraction of the design (runs with odd order number) plus the centre run, we found three factors 2, 3, and 4 significant at 10% level and two factors 7 and 8 significant at 5% level with the adjusted $R^2 = 0.9875$.
- ii. The experiment is conducted in stages, and in the first stage, the design is a half fraction of a 3-level DSD and not of the one of a 2-level FFD of resolution IV. In the second stage, the fraction is the foldover of the first half fraction.
- iii. The researchers do not wish to use designs with full aliases between MEs and 2FIs and among 2FIs.

The suggested alternative design for the 2^{8-4}_{IV} FFD is the DSD in **Table 4(b)**. This DSD is constructed by Eq. (2) with C being the conference matrix of order 8 in **Table 3**. The CCPs of the 2^{8-4}_{IV} FFD and the DSD are in **Figure 2(a)** and **(b)**, respectively. As expected, both CCPs show that the MEs are orthogonal to the 2FIs. There are 42 dark cells (which represent full aliases between the 2FIs) in the upper/lower diagonal portion

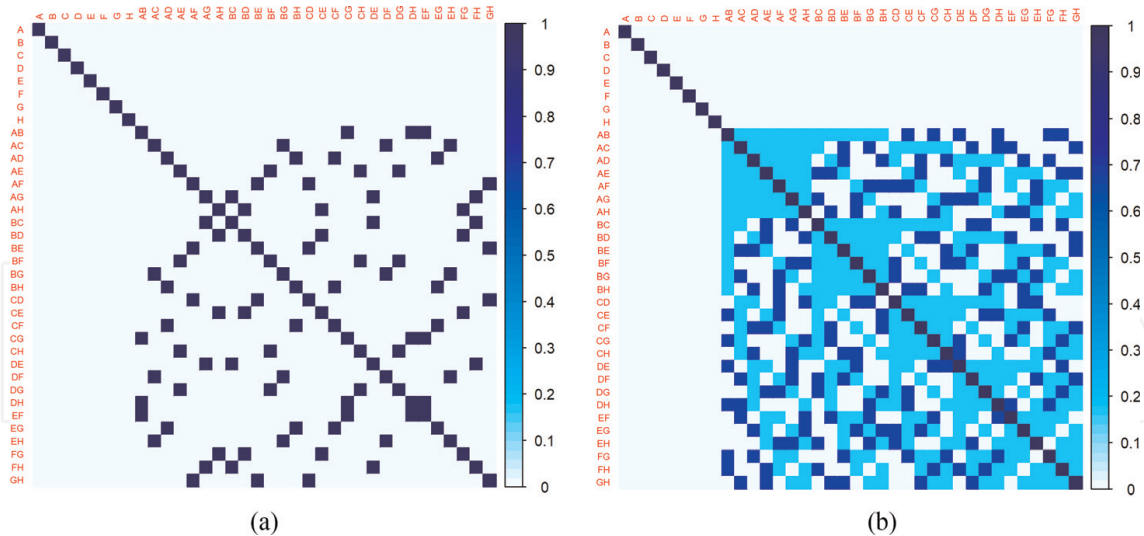


Figure 2. CCPs for (a) a 2^{8-4} FFD used in Yao’s experiment and (b) a DSD for eight factors in 17 runs (Table 4(b)).

of the CCP in **Figure 2(a)**. Unlike the CCP for the 2^{8-4} FFD in **Figure 1(a)**, the one for the DSD in **Figure 2(b)** shows that none of the 2FIs is fully aliased with the other 2FIs.

The reasons for using a DSD instead of a resolution IV FFD are mentioned in the previous section. Unlike resolution IV FFDs, DSDs can also estimate m QEs (in addition to the intercept and m MEs). Also, QEs are orthogonal to MEs and not fully aliased with 2FIs.

Up to this point, we have been discussing conference matrix-based designs when all factors are quantitative. When there are m_3 3-level factors and m_2 2-level factors, i.e. qualitative or categorical factors, we select $m_3 + m_2$ columns from m columns of a conference matrix and then change 0’s to 1’s in the last m_2 columns. The u th row of the model matrix \mathbf{X} in (3) is now written as

$$\left(1, d_{u1}, \dots, d_{u(m_3+m_2)}, d_{u1}^2, \dots, d_{u(m_3+m_2)}^2, d_{u1}d_{u2}, \dots, d_{u(m_3+m_2-1)}d_{u(m_3+m_2)}\right).$$

Let C^* be the matrix formed by these columns, the final design is of the form:

$$\begin{pmatrix} C^* \\ -C^* \end{pmatrix} \quad (4)$$

This simple method of constructing conference matrix-based designs with mixed-level was mentioned in [11]. Another method of constructing this type of design was discussed in [12]. When there are more 2-level factors than 3-level ones, the readers are encouraged to use the Hadamard matrix-based designs discussed in [13]. Note that when there is a need to block a design into two blocks, one of the 2-level factors can be used as a blocking factor. When there is a need to block a design into three or more blocks or when there is more than one blocking factor, readers are encouraged to refer to [14].

4. Conclusions

This chapter advocates the use of conference matrix-based designs for screening experiments when the factors are quantitative. For these experiments, a number of

quantitative factors have to be studied, but only a few of them is expected to be important. In the past, popular designs for this type of experiment are PBDs and 2-level regular FFDs of resolution III and IV. Conference matrix-based designs, unlike the mentioned popular designs, cannot be analysed by hand-calculators. Nowadays, this is not an issue as most data analysis can be done on computers.

Data from experiments using conference matrix-based designs discussed in this paper can be analysed by more advanced statistical methods such as subset or step-wise regression.

Additional use of conference matrices in screening experiments can be found in [15].

The link for the matrices in **Tables 2** and **3** is at <https://designcomputing.net/Cmatrices/>.

Author details

Nam-Ky Nguyen^{1*}, Stella Stylianou², Tung-Dinh Pham³ and Mai Phuong Vuong⁴

1 Vietnam Institute for Advanced Study in Mathematics, Hanoi, Vietnam


2 School of Science, Mathematical Sciences, RMIT University, Melbourne, Australia

3 VNU University of Science, Vietnam National University, Hanoi, Vietnam

4 Hanoi University of Science and Technology, Hanoi, Vietnam

*Address all correspondence to: nknam@viasm.edu.vn

IntechOpen

© 2022 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Poorna V, Kulkarni PR. A study of inulinase production in *aspergillus niger* using fractional factorial design. *Bioresource Technology*. 1995;**54**: 315-320
- [2] Mee RW. *A Comprehensive Guide to Factorial Two-Level Experimentation*. New York: Springer; 2009
- [3] Yao CL, Liu CH, Chu IM, Hsieh TB, Hwang SM. Factorial designs combined with the steepest ascent method to optimize serum-free media for ex vivo expansion of human hematopoietic progenitor cells. *Enzyme and Microbial Technology*. 2003;**33**:343-352
- [4] Plackett RL, Burman JP. The design of optimum multifactorial experiments. *Biometrika*. 1946;**33**:305-325
- [5] Ionin YJ, Kharaghani H. Balanced generalized weighing matrices and conference matrices. In: Colbourn JH, editor. *Handbook of Combinatorial Designs*. 2nd ed. Boca Raton, FL: CRC Press; 2007. pp. 306-313
- [6] Nguyen NK, Stylianou S. Constructing definitive screening designs using cyclic generators. *Journal of Statistics Theory & Practice*. 2013;**7**: 713-724
- [7] Jones B, Nachtsheim CJ. A class of three levels designs for definitive screening in the presence of second-order effects. *Journal of Quality Technology*. 2011;**43**:1-15
- [8] Fidaleo M, Lavecchia R, Petrucci E, Zuorro A. Application of a novel definitive screening design to decolorization of an azo dye on boron-doped diamond electrodes. *International Journal of Environmental Science and Technology*. 2016;**13**:835-842
- [9] Stylianou S. Three-level screening designs applicable to models with second order terms. In: Paper Presented at the International Conference on Design of Experiments (ICODOE-2011). Department of Mathematical Sciences. USA: University of Memphis Memphis; 2011
- [10] Xiao L, Lin DKJ, Bai F. Constructing definitive screening designs using conference matrices. *Journal of Quality Technology*. 2012;**44**:2-8
- [11] Nguyen NK, Kenett RS, Pham TD, Vuong MP. Recent development on D-efficient mixed-level foldover designs for screening experiments. In: Pham H, editor. *Springer Handbook of Engineering Statistics*. 2nd ed. London (accepted): Springer; 2022
- [12] Jones B, Nachtsheim CJ. Definitive screening designs with added two-level categorical factors. *Journal of Quality Technology*. 2013;**45**:121-129
- [13] Nguyen NK, Pham TD, Vuong MP. Constructing D-efficient mixed-level foldover designs using Hadamard matrices. *Technometrics*. 2020;**62**:48-56
- [14] Nguyen NK, Pham TD, Vuong MP. Multiway blocking of designs of experiments. *Statistics and Applications*. 2021;**19**:1-9
- [15] Stylianou S. Foldover conference designs for screening experiments. *Communication in Statistics—Theory and Methods*. 2010;**39**:1776-1784